# ENVIRONMENTS ASSOCIATED WITH DRYLINE CONVECTION IN THE SOUTHERN GREAT PLAINS

A THESIS SUBMITTED TO THE UNIVERSITY OF MANCHESTER
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE FACULTY OF SCIENCE AND ENGINEERING

2021

By
Trevor J. Mitchell
School of Earth and Environmental Sciences

[Blank page]

# Contents

Word Count: 25,657

# List of Figures

# Abstract

Environments Associated with Dryline Convection in
the Southern Great Plains
Trevor J. Mitchell
A thesis submitted to The University of Manchester
for the Degree of Doctor of Philosophy, 2020


This thesis is in the alternative format and comprises two separate journal articles and an intermediate chapter that together form a coherent research project. Synoptic-scale differences between drylines that produce deep, moist convection and those that do not are determined in two papers.

In the first paper, a dataset of drylines within a region of the Southern Great Plains is constructed from surface analyses. Doppler radar, visible and infrared satellite imagery are used to identify convective drylines, where deep, moist convection was deemed to have been associated with the dryline circulation. Composite synoptic analyses of 179 convective and 104 non-convective dryline days reveal previously unidentified differences between convective and non-convective drylines. Convective drylines feature more amplified upper-level flow, associated with a deeper trough in the western US and a stronger downstream ridge than non-convective drylines in the three days preceding a dryline event. As a result of greater poleward low-level moisture transport, significant differences are observed between the composites on the day of a dryline event. The convective composite features greater specific humidity at low to mid levels and higher CAPE than the non-convective composite.

A more objective method of analysis, machine learning, is then investigated as a tool for predicting dryline convection and identifying its sensitivity to numerical weather prediction model output. Gradient boosting is applied to model data obtained for 205 dryline days identified during the climatology. The model has a high probability of detecting convective drylines, but labels too many drylines as convective, resulting in false positives. Analysis of model feature importance reveals variable performance varies spatially. The model attaches high importance to the strength of the upper-tropospheric jet over the Rockies when predicting dryline convection. However, instability and mid-level moisture are important variables in locations immediately east of the dryline. These results are consistent with analysis of the composites, and provide evidence that synoptic-scale processes can help determine whether or not a dryline will produce deep, moist convection.

# Declaration

**The University of Manchester**
*PhD by published work Candidate Declaration*

**Candidate Name:** Trevor J. Mitchell

**Faculty:** Science and Engineering

**Thesis Title:** Environments associated with dryline convection in the Southern Great Plains

**Declaration to be completed by the candidate:**

I declare that no portion of this work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Signed: Date: March 9, 2021

# Copyright

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see `http://http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=24420`), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see `http://www.library.manchester.ac.uk/about/regulations`) and in The University's policy on presentation of Theses

# Acknowledgements

I would like to thank several people who have helped in my four years at the University of Manchester. Firstly, my supervisor David Schultz, who has been patient, understanding and gone above and beyond what could be expected. Without him, this thesis would not have been possible. I would also like to thank Geraint Vaughan for his insightful comments that have improved this research, and for his moral support.

My time in Manchester was made much more enjoyable by my colleagues in our research office, 3.16. In my time here, many have come and gone, but I would like to give a special mention to Callum Thompson and Ty Buckingham. Discussions with them have helped generate and develop research ideas, but just as importantly, they have been vital in helping deal with the stresses of a PhD. I would also like to thank friends outside of my research group. In particular, Simon Ruske, Kathryn Fowler, Sophie Mayne and Matthew Brown who all helped me stay sane! Finally, I would like to thank Julie Samson. Quite simply, the DTP programme is greatly enhanced by her presence.

# Chapter 1

# Introduction

## 1.1 Overview

Drylines are moisture boundaries that separate a moist air mass from a dry air mass, usually identified by horizontal gradients in surface dewpoint. The dewpoint across a dryline can vary by more than 18°C in less than 10 km (Parsons et al. 1991). The moisture gradient is sometimes, but not always associated with a wind shift or pressure trough which aids in identification. Some drylines also exhibit large temperature differences between the moist and dry air. However, drylines are unlike fronts in that the direction of temperature gradient usually reverses at night.

In the USA a dryline is often found in the Great Plains, east of the Rockies. Drylines in the Great Plains separate moist air from the Gulf of Mexico and drier air sourced from elevated regions in the desert south-west of the USA. An example of a Great Plains dryline is shown in Fig. 1.1. The dryline is represented by continuous orange semicircles and arcs from north to south. Drylines most frequently occur in late spring and early summer because the zonal gradient in dewpoint is greatest during these months (Dodd and Dodd 1965; Schaefer 1974). The dryline season coincides with a period of favourable vertical wind shear and instability in the Great Plains, which can lead to the formation of severe thunderstorms.

Convergence in the vicinity of the dryline means they often serve as a focus for convection initiation (Fig. 1.2). Severe thunderstorms that form along the
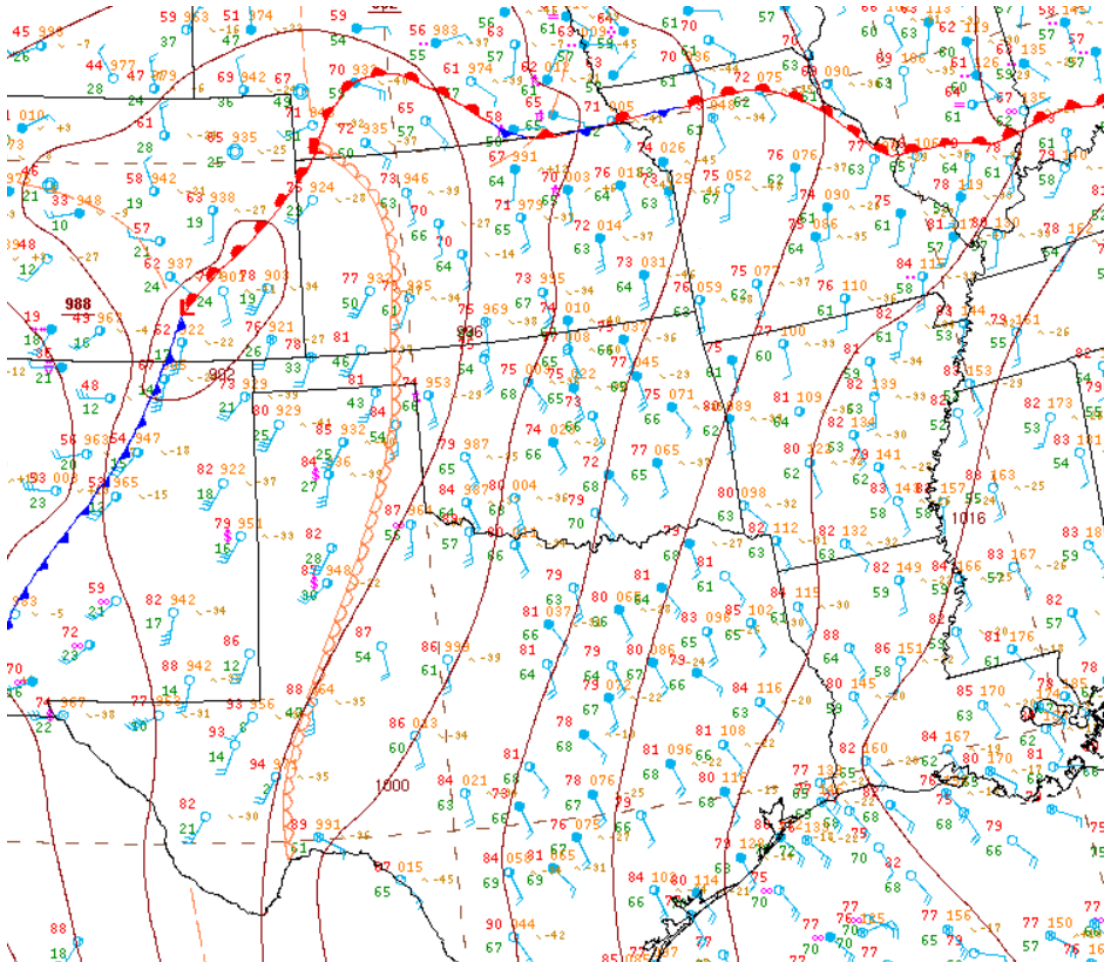
Figure 1.1: An example of a Weather Prediction Center surface analysis depicting a dryline. The dryline is represented by the orange semicircles and stretches from the Kansas-Nebraska border to the USA-Mexico border. Troughs and outflow boundaries are represented by orange dashed lines.

dryline can bring hazards such as strong winds, heavy rain, hail and tornadoes. The environment in the vicinity of the dryline often supports severe thunderstorms if convection can initiate. Therefore, whether or not convection initiation occurs can be the difference between a day with severe thunderstorms and no significant weather at all. Hence, predicting whether a dryline will produce deep, moist convection is of paramount importance.

Deep, moist convection requires moist air to be lifted such that instability can be released and parcels can ascend to a through a large depth of the troposphere. In the vicinity of the dryline, a warm layer of air with steep lapse rates typically
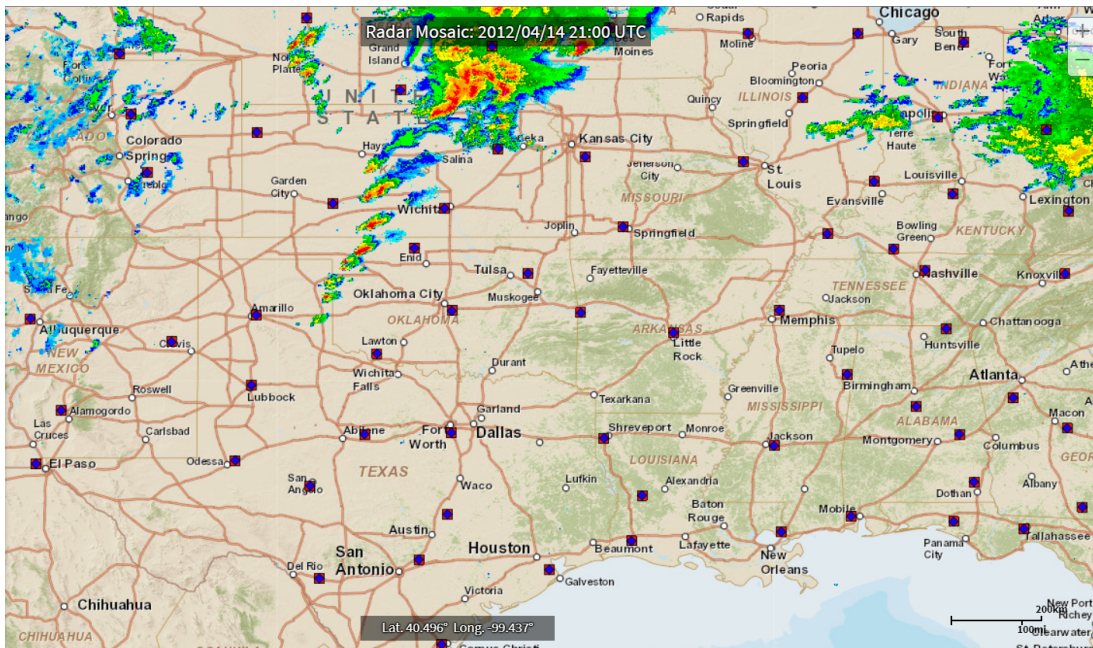
Figure 1.2: A radar reflectivity mosaic obtained using the NCEI/NOAA Interactive Radar Map Tool (NCEI 2016), for the same time as in Fig. 1.1. A string of high reflectivity echoes arc from north-east to south-west. These echoes are thunderstorms that formed along the dryline.

sits above a moist air mass, creating a conditionally unstable environment that creates the potential for this deep convection. The warm layer is usually referred to as an elevated mixed layer (EML), which is generated in conditions of strong heating and mixing in elevated terrain to the west. As the EML is advected into the Great Plains it rides over the top of a low-level moist air mass and creates a capping inversion (Fig. 1.3).

For deep, moist convection to occur, air parcels must either be lifted through the capping inversion, or the inversion must be removed. Synoptic-scale ascent is thought to condition the environment so that either of these processes can occur. Adiabatic cooling associated with large-scale ascent reduces convective inhibition (CIN). However, the magnitude of ascent is not thought to be enough to initiate deep convection due to the large timescale of lifting a parcel from the surface to the level of free convection (LFC); the height at which a moist parcel becomes positively buoyant and can ascend freely. Ascent is thought to bring parcels close to saturation, with mesoscale processes such as the dryline circulation believed to reduce any remaining CIN or provide enough ascent for parcels to reach the
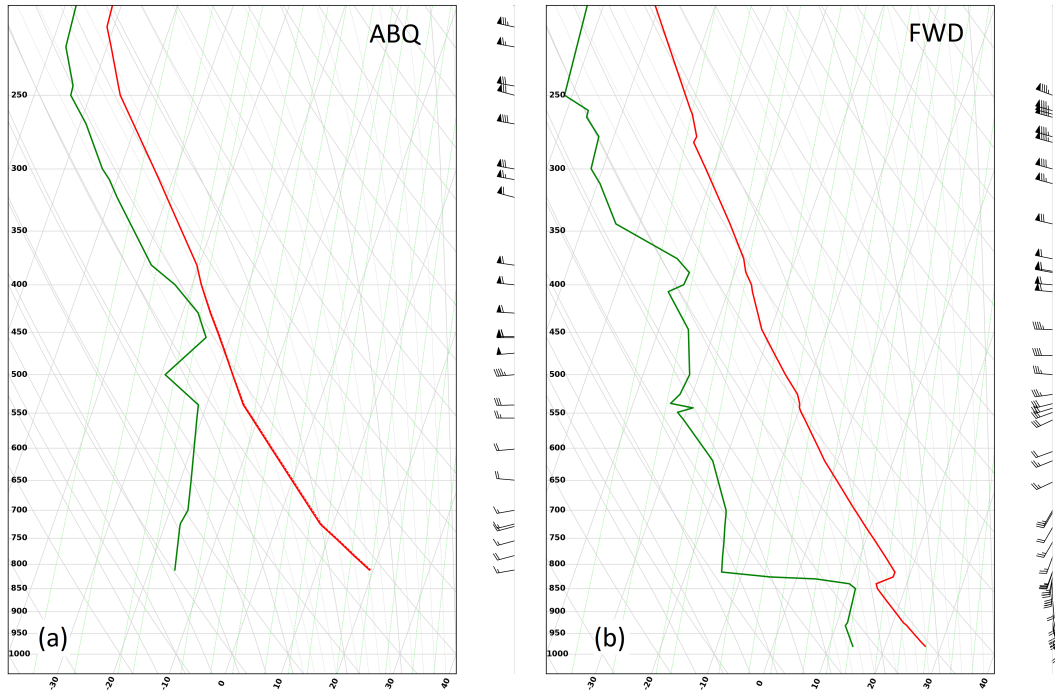
Figure 1.3: Vertical profiles of temperature (red) and dewpoint (green) from (a) Albuquerque, New Mexico and (b) Fort Worth, Texas. The Albuquerque sounding is from 00 UTC 7th April 2008 and the Fort Worth sounding from 24 hours later. At Fort Worth, a moist air mass is in place beneath an inversion at around 840 hPa. Above the inversion, the temperature and moisture profiles are broadly similar to those seen at Albuquerque a day before, suggesting that the well-mixed layer at Albuquerque was advected downstream.

LFC. However, the residence time of an individual updraft in a mesoscale area of ascent is a determining factor in convection initiation. For deep, moist convection, parcels must reach the LFC before they become detrained from a mesoscale updraft by wind shear (Ziegler and Rasmussen 1998).

The necessity of considering these atmospheric processes ranging from the synoptic to the convective-scale makes dryline convection a difficult forecasting problem. Forecasters typically use an ingredients-based approach to forecast deep, moist convection, whether or not the area of interest is along a dryline. Deep, moist convection requires the presence of three key ingredients: moisture, instability and lift (Doswell 1987). These three ingredients are required, but not necessarily sufficient for deep, moist convection, and the dryline is a region in which these ingredients are often present. The environment east of the dryline is

often conditionally unstable as a moist boundary layer exists beneath the EML. Convergence at the dryline can provide the lift necessary to initiate deep, moist convection. However, not all drylines produce deep, moist convection, despite the apparent presence of these ingredients. What are the differences between drylines that produce deep, moist convection and those that do not?

Synoptic-scale processes regulate dryline formation and intensity and are thought to prime the dryline environment for deep, moist convection. However, the precise nature of how these processes lead to convection initiation is not well understood. How does forcing for synoptic-scale ascent determine the initiation of convective storms along the dryline? Stronger drylines are associated with the passage of a short-wave trough, which promotes cyclogenesis, increased confluence and the tightening of the moisture gradient across the dryline (Schultz et al. 2007). Are there similarities between the environments associated with strong drylines and those with dryline convection? More generally, are there large-scale differences between drylines that produce deep, moist convection and those that do not? Or is dryline convection initiation determined only by smaller-scale processes which are harder to predict at short time scales?

This thesis attempts to answer some of these questions via two methods. Firstly, a climatology of dryline convection is constructed from surface analyses, radar, and visible and infrared satellite data. Synoptic composite analyses are created from reanalysis data for two categories: convective and non-convective drylines. Secondly, a machine learning technique is tested using the climatology dataset and reanalysis data to answer two primary questions. Can dryline convection be predicted using machine learning, and if so, which variables are best at predicting whether dryline convection will occur?

The thesis is structured as follows. Section 1.2 defines the dryline, and discusses the importance of dryline convection. The formation and typical characteristics of the Great Plains dryline are presented in section 1.3. Section 1.4 explains current hypotheses of the causes dryline motion. The processes by which dryline convection occurs are synthesized in Section 1.5 Section 1.5 is broken into sections, separated by spatial scale. Section 1.5.1 introduces quasi-geostrophic theory and its applications. Some of the concepts of quasi-geostrophic theory

are applied in section 1.5.2, which discusses the role of synoptic-scale processes in dryline convection. Section 1.5.3 describes how mesoscale processes provide ascent along the dryline. Small-scale convective processes are then discussed in section 1.5.4. Section 1.6 provides an overview of previous dryline climatologies, and discusses the benefits of creating a climatology of dryline convection. Chapter 2 presents a synoptic climatology of dryline convection in the Southern Great Plains. Chapter 3 discusses alternative approaches to examining dryline convection via machine learning. A machine learning technique known as gradient boosting is applied to dryline data in chapter 4. Finally, conclusions of chapters 2 and 4 are presented in chapter 5.

## 1.2  What is a Dryline?

A dryline is an elongated zone of enhanced horizontal moisture gradient, usually at a boundary of air masses of different moisture content. Drylines were first documented after meteorologists noticed that tornadoes were often associated with thunderstorms that formed along instability lines (Beebe 1958). Aircraft traverses of instability lines in western Nebraska were analysed by (Beebe 1958), and (Fujita 1958). Beebe noted a rapid change in 800-hPa moisture which he referred to as a dewpoint "front". Fujita identified a "frontal surface' between dry and moist air at 700-hPa, and referred to the intersection of that surface and the ground as a "dry front". McGuire (1962) may have been the first to refer to these moisture boundaries as drylines when he documented aircraft traverses of moisture discontinuity lines in western Texas. Although these early studies of drylines focused on cases east of the Rockies, drylines are not restricted to the USA. Countries with documented cases of drylines include China, India and Australia (Weston 1972; Golden 1980; Arnup and Reeder 2007).

In the USA a dryline is often found in the Great Plains, east of the Rockies. Drylines in the Great Plains separate moist air from the Gulf of Mexico and drier air from the desert southwest. The elevation of the Rockies and the topography of the Great Plains are significant in the formation of drylines because the Great Plains generally increase in elevation in a westerly direction. The dryline can be considered to be the western edge of the moist air mass from the Gulf of Mexico as it intersects the sloping terrain of the Great Plains.

Gradients in surface dewpoint are frequently used to identify the western edge of the moist air mass. The dewpoint across a dryline can vary by more than 18°C in less than 10 km (Parsons et al. 1991). The moisture gradient typically increases during the late morning and afternoon and is sometimes, but not always, coincident with a wind shift or pressure trough, which can aid in identification. Drylines are most common in late spring and early summer because the zonal gradient in dew point is greatest during these months (Dodd and Dodd 1965; Schaefer 1974). In addition to a large dewpoint gradient, some drylines also have large temperature differences between the moist and dry air. However, unlike fronts, the temperature gradient typically reverses at night. This change in

temperature gradient occurs because the dry air usually has a larger diurnal temperature range than the moist air. Dry air has a lower specific heat capacity than moist air, so it heats more quickly during the day and cools more quickly at night.

Drylines in the USA are the most studied in the world. Convergence in the vicinity of the dryline means they often serve as a focus for convection initiation. Air parcels can often be lifted sufficiently for convective instability to be released and deep, moist convection to occur. In the Great Plains, favourable wind shear and the release of this convective instability can often lead to the formation of severe thunderstorms. These storms can produce hazards such as high winds, lightning, large hail and tornadoes and flash flooding. As a result, dryline convection is closely monitored by forecasters, especially in the spring and early summer when the threat of severe thunderstorms is greatest.

## 1.3   Dryline Formation and Characteristics

The topography of the Great Plains is an important factor in the formation of the dryline because elevation generally increases in a westerly direction (Fig. 1.4). The dryline can be considered as the location where the western edge of a moist Gulf air mass intersects the sloping terrain.Hoch and Markowski (2005) found that the dryline was most commonly located around 101°W at 0000 UTC during spring. They hypothesized that drylines were favoured in this region as it is where the east-west elevation changes at a greater rate. Due to the close proximity of the Gulf of Mexico, drylines are most commonly found in western regions of Texas, Kansas and Oklahoma. However, in conditions of long fetch southerly flow, drylines can also be found further north. For example, Campbell et al. (2014) performed a case study of a dryline in Wyoming. Drylines are not restricted to the Great Plains, they have been identified as far east as Mississippi (Duell and Van Den Broeke 2016).
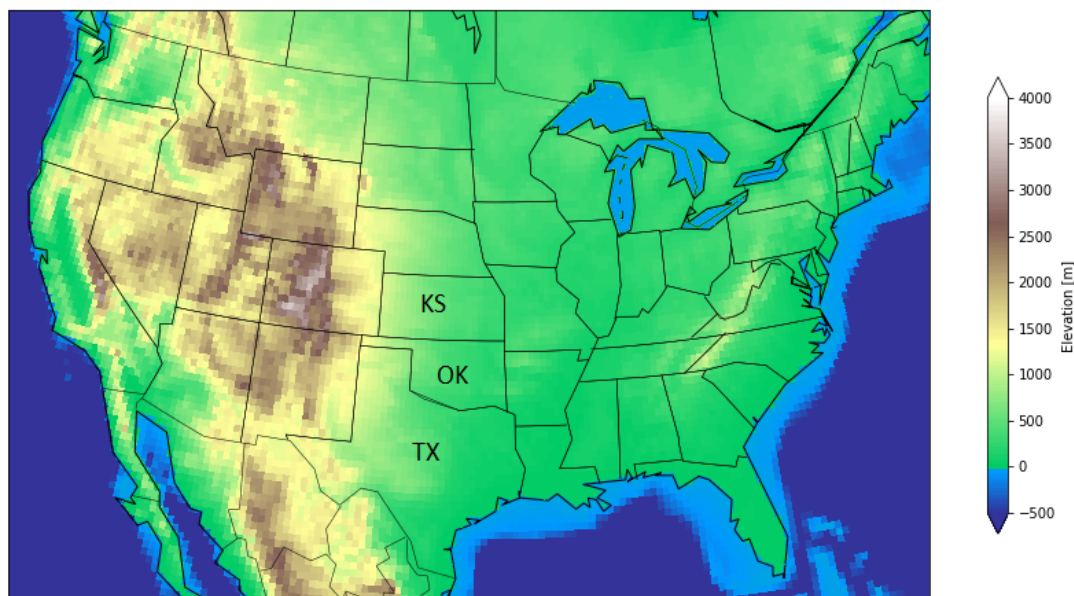


Figure 1.4: The topography of the contiguous United States. Elevation increases from east to west in the Great Plains. Drylines are most commonly found in western parts of Texas (TX), Oklahoma (OK) and Kansas (KS).

Synoptic-scale processes help create the conditions for dryline formation and intensification given the topography of western and central USA. Westerly flow across the Rockies results in lee troughing which can help a west-east moisture

gradient to develop (Fig. 1.5). When there is westerly flow across elevated terrain, column stretching occurs as the air descends in the lee of the mountains. The stretching of a column is associated with increased surface vorticity and convergence, and the formation of a surface low. Convergence and deformation associated with lee cyclones helps create an airstream boundary between the moist Gulf air and the dry air from the south-west (Schultz et al. 2007). Hobbs et al. (1996) proposed that the confluence of the moist Gulf air and warm, dry air from the Rockies creates a strong west-east gradient in moisture. Horizontal convergence associated with lee troughs helps increase the moisture gradient and troughs can acquire dryline-like characteristics.

In addition to aiding dryline formation, synoptic-scale processes are also important in regulating dryline intensity. Schultz et al. (2007) analysed spring days over two years to see if a moisture gradient was present between two Texas locations using West Texas Mesonet data. Composite analyses were produced for strong (upper quartile) and weak (lower quartile) dryline cases sorted by the magnitude of the moisture gradient. The strong cases featured a short wave trough in the westerly flow to the west of the region, and an attendant strong surface low over eastern New Mexico resulting in southerly flow over the southern Great Plains. They concluded that contraction associated with surface cyclones intensifies the large scale west-east moisture gradient.

The west-east moisture gradient is also partly caused by the characteristics of the land surface. Vegetation and soil moisture generally increase in an easterly direction in the Great Plains; therefore sensible heating tends to have an east-west gradient. In a study of the relationship between soil moisture and dryline position and strength, Johnson and Hitchens (2018) found that strong soil moisture gradients may contribute to more intense drylines. They also found that overall soil moisture, not just the magnitude of the horizontal gradient, can influence dryline position. As soil moisture values increased, the dryline tended to form farther west.

Areas farther west tend to be drier, have stronger mixing, and a deeper boundary layer than areas farther east (Lanicci et al. 1987; Ziegler et al. 1995). The deeper boundary layer to the west of the dryline is associated with a mixing down of westerly momentum from aloft, which leads to low-level veering and increased
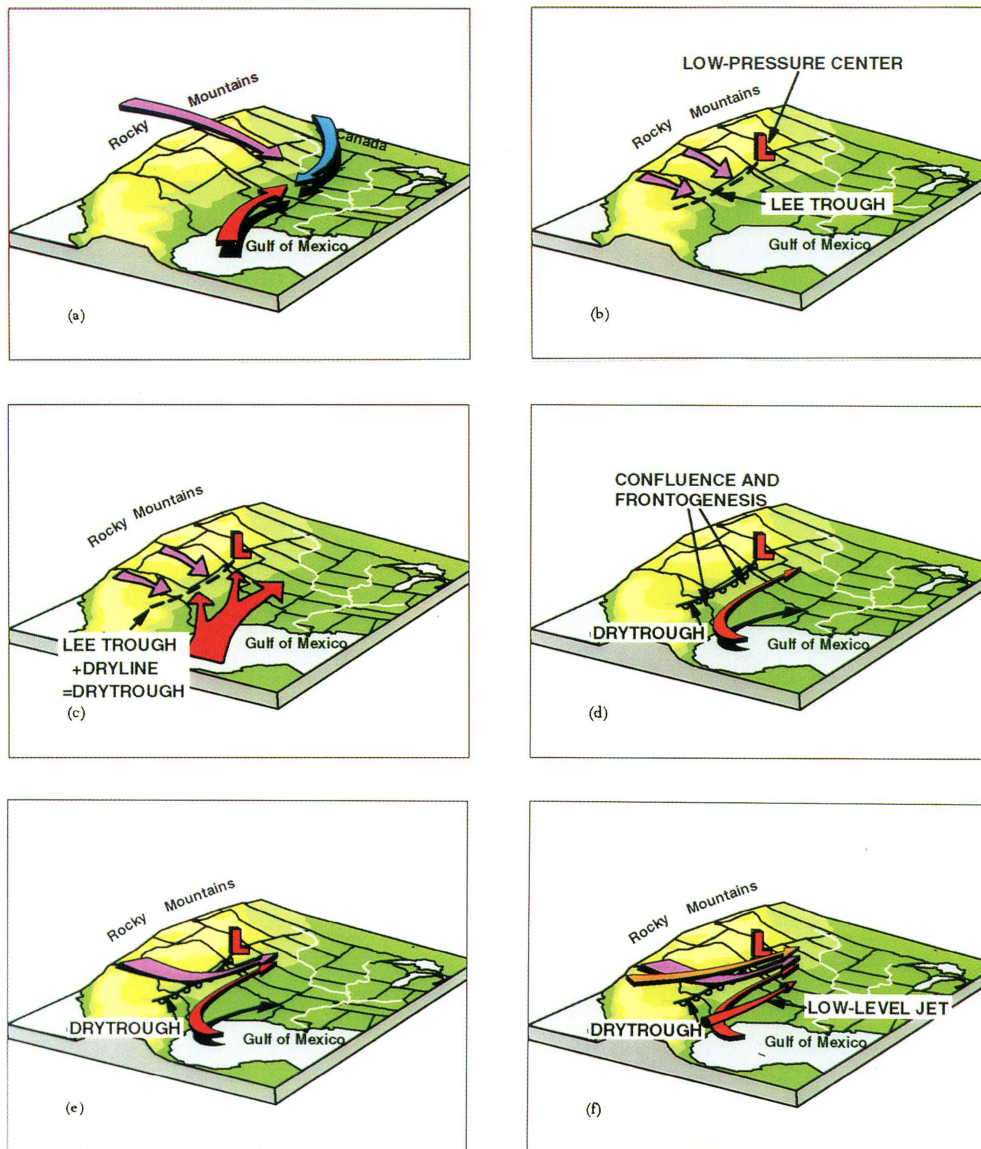
Figure 1.5: Schematics illustrating airflows and the formation of a drytrough as an eastward moving, shortwave trough passes over the Rocky Mountains, (a) The confluence of westerly downslope flow off the Rockies, warm, moist air from the Gulf of Mexico, and cold, dry arctic air. (b) Adiabatic warming of downslope air from the Rockies produces a lee trough, (c) The lee trough acquires the characteristics of a dryline and becomes a drytrough. (d) Confluence and frontogenesis east of the drytrough causes air from the Gulf of Mexico to rise and turn toward the northeast as it approaches the drytrough. (e) Downslope flow of warm, dry air from the Rockies reaches its lowest altitude over the drytrough then rises above the warm, moist air from the Gulf, (f) The elevated mixed layer from the Mexican plateau flows above the other two airstreams.  Also, a low-level jet of warm, moist air flows northward (Hobbs et al. 1996).

convergence (McCarthy and Koch 1982; Atkins et al. 1998).

The differences in boundary layer height are not only caused by horizontal soil moisture gradients. A capping inversion or "lid" is often present above the relatively cool, moist air from the Gulf of Mexico. This lid is an elevated mixed layer, created by strong heating over arid regions such as the Mexican Plateau, then advected into the Great Plains where it overruns the moist air (Lanicci and Warner 1991). During the afternoon, the western edge of the lid roughly corresponds to the surface dryline location (Schaefer 1974).

Vertical mixing is a dominant factor in the intensification of synoptically quiescent drylines; when no major synoptic-scale system is translating from west to east (Schaefer 1986; Hane 2004). Schultz et al. (2007) found that dryline intensification primarily results from a combination of vertical mixing and frontogenetical circulations when upper-level synoptic forcing is lacking.

Solenoidal circulations can increase the moisture gradient at the dryline (Sun and Ogura 1979; Ziegler and Hane 1993), and are frontogenetical in nature. In a modelling study across a dryline in western Oklahoma, Ziegler et al. (1995) found a thermally direct circulation is driven by horizontal gradients in virtual potential temperature. Easterly low-level flow induced by the virtual potential temperature gradient decelerates as it approaches the dryline. The resultant low-level convergent flow at the dryline is frontogenetical in regards both temperature and moisture, intensifying the dryline. The convergence at the dryline is associated with ascent which tilts in an easterly direction over the moist air mass.

Some of the features of a solenoidal circulation can be seen in (Fig. 1.6), which shows cross-sections across a dryline through late afternoon for a dryline observed during the International $H_2O$ Project (Weckwerth et al. 2004; Buban et al. 2007). Near-surface air moves west towards the dryline, rises, then tilts to the east before descending. Also of note is how the westward advance of the dryline appears to coincide with enhancement of the gradient of virtual potential temperature at the boundary.
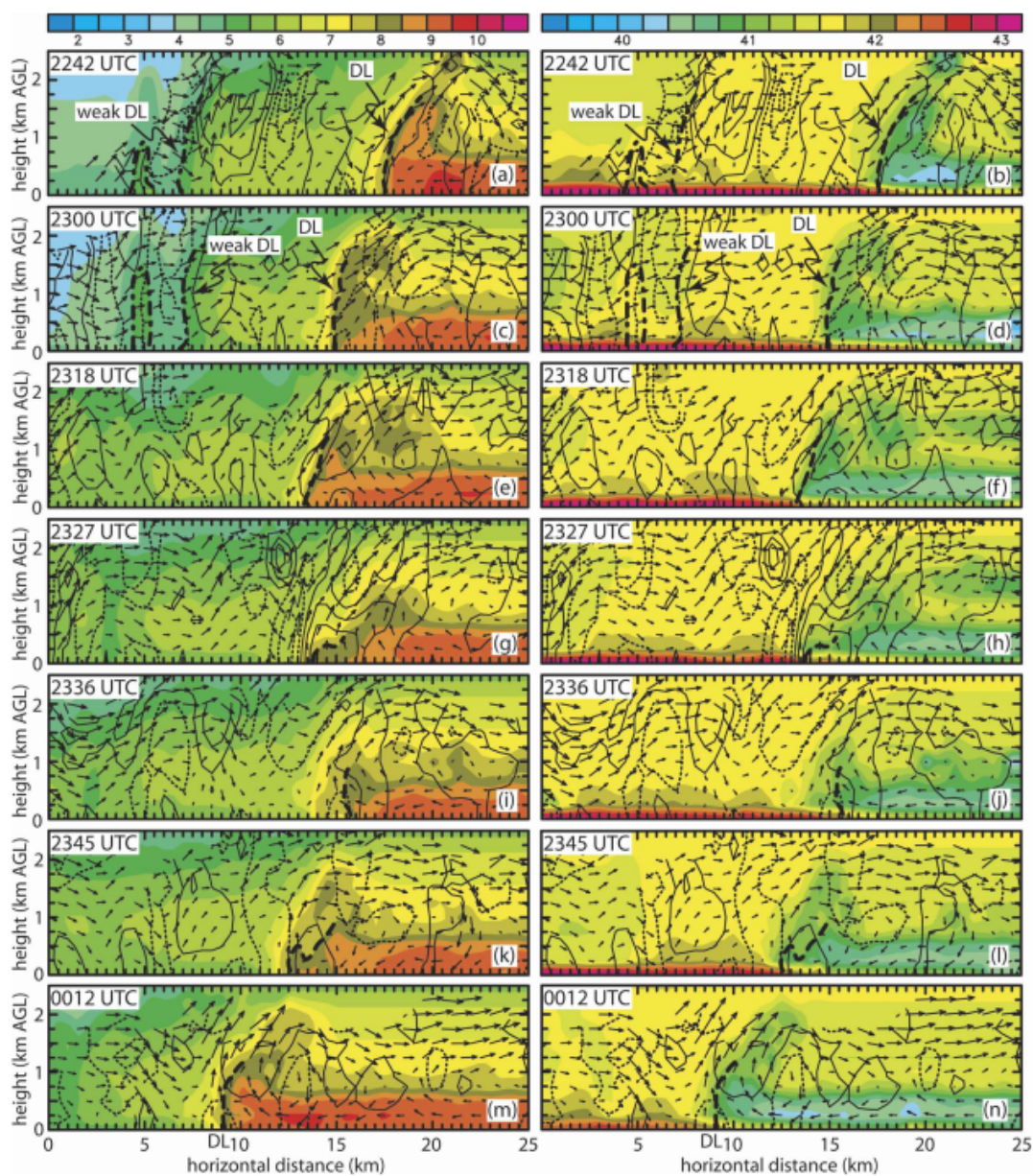
Figure 1.6:  Vertical cross sections of (left) water vapor mixing ratio (g kg$^{-1}$) and (right) virtual potential temperature (K) across a dryline during IHOP. Ground-relative wind vectors (1-km length equal to 10 ms$^{-1}$), and vertical vorticity every 2 x 10$^{-3}$ s$^{-1}$ with positive values (black solid lines) starting at 1 x 10$^{-3}$ s$^{-1}$ and negative values (black dashed line) starting at -1 x 10$^{-3}$ s$^{-1}$ are overlaid (Buban et al. 2007).

Gradients in virtual potential temperature have been observed in several dry-line studies (e.g. Parsons et al. 1991; Sun and Wu 1992; Ziegler and Hane 1993). These gradients can develop if the increase in moisture from west to east is not

large enough to offset the reduction in potential temperature (Ziegler and Hane 1993). In a study of a retrogressing dryline in western Texas, Parsons et al. (1991) found a density gradient in late afternoon and evening. They proposed that the leading edge of the westwardly-moving moisture acts similarly to a density current. Geerts (2008) also observed a density gradient across a dryline in western Texas. He found that the strength of convergence is related to the gradient of virtual potential temperature, with a stronger gradient implying stronger vertical motion. Convergence at the dryline can lead to vertical motion as high as 8–9 ms$^{-1}$ (Weiss et al. 2006).

## 1.4   Dryline Motion

Many drylines migrate eastwards throughout the day. Under synoptically quiescent conditions, mixing is one of the main processes in dryline propagation (Schaefer 1986). The dryline propagates east due to mixing as moist air at the boundary entrains dry air. This eastward advance of the dryline is sensitive to the magnitude of surface heat flux, the strength of the inversion and the depth of the mixed layer (Jones and Bannon 2002). The term propagation is used as drylines do not always exhibit continuous motion (McCarthy and Koch 1982; Schaefer 1986; Crawford and Bluestein 1997). Sometimes a dryline will jump as the inversion is removed over a large area in a short time.

In synoptically active cases, advection is a greater factor in the eastward progression of a dryline. Hane (2004) investigated a north-eastward translating upper-level height minimum with an attendant trough extending southward. He found that the dryline broadly moved northeast with the upper-level wave. Eastward motion and development of a dryline bulge was associated with strong vertical mixing near the surface and strong advection of dry air aloft.

In synoptically quiescent cases, the eastward motion of the dryline is somewhat opposed by south-easterly flow ahead of the lee trough, which typically intensifies due to strong heating over dry plateau regions to the west (Benjamin and Carlson 1986). The presence of the lee trough can lead to the dryline retreating westwards during the evening and overnight. As vertical mixing reduces and a nocturnal inversion starts to form, winds back in response to the lee trough, moisture flows westwards and the dryline retreats.

However, the mass response to a lee trough may not be the sole reason for retrogression of the dryline. Ziegler et al. (1995) speculated that the westward propagation of the density current is opposed by westerly shear at the dryline and the sloping terrain. A reduction of mixing to the west of the dryline and subsequent formation of an inversion would be likely to reduce the magnitude of the westerly shear and slow or even reverse the eastward propagation of the dryline.

# 1.5 Dryline convection

The migration of the dryline is important because deep, moist convection often occurs within the vicinity of it. But how does this convection occur? Doswell (1987) proposed an ingredients-based approach for forecasting deep, moist convection. His approach requires the existence of three main ingredients; moisture, instability and lift. All three ingredients are required for deep moist convection; however, any of them can act as the trigger given the presence of the other two. Furthermore, the processes by which these ingredients are brought together may vary between weather events.

The processes associated with deep, moist convection are often diagnosed using parcel theory. Deep, moist convection requires the release of instability, allowing parcels to ascend through a large vertical depth. Saturated air parcels must be unstable throughout a large depth of the troposphere above the level of free convection (LFC).

Instability is often measured in terms of convective available potential energy (CAPE), which may be considered as the amount of energy available for convection. Once a parcel reaches its LFC, potential energy can be converted to kinetic energy as the parcel accelerates vertically. CAPE is a measure of the instability between the LFC and the equilibrium level; the level at which the parcel is no longer positively buoyant. CAPE is strongly related to both low-level moisture and the environmental lapse rate; the rate at which the atmosphere cools with height. Increases in low-level moisture and lapse rates are both associated with increase in CAPE.

Often, a warmer layer of air exists above the LCL, and air parcels will become negatively buoyant within the inversion. For continued ascent, parcels must obtain enough momentum to rise above this inversion and reach their LFC, either by sustained lift or removal of the inversion. Therefore, the updraft velocity and the strength of the inversion are important in determining whether deep, moist convection will occur.

Updraft velocity is a function of CAPE, with large CAPE associated with stronger updrafts. Therefore CAPE is one of the main convective forecasting

tools. However, the location and distribution of CAPE needs to be considered when forecasting convection. Two profiles with the same overall magnitude of CAPE can produce vastly different forms of convection depending on the distribution of the instability. Large low level CAPE will produce a greater updraft velocity than skinny CAPE that is distributed more evenly vertically (Blanchard 1998; McCaul and Cohen 2002). Updrafts with greater vertical momentum are thought to be less affected by water loading and entrainment (Lucas et al. 1994b,a).

An elevated mixed layer (EML) often acts as a capping inversion, especially over the moist air to the east of a dryline. Strong mixing commonly occurs in arid regions of New Mexico, Arizona and northern Mexico where sensible heating is large, resulting in a deep mixed layer. If this air is advected towards lower terrain in the east, it essentially rides over the top of the moist air mass near the surface. The EML is typically close to dry adiabatic, resulting in large lapse rates.

The presence of an EML above a moist boundary layer is of interest to weather forecasters as large mid-level lapse rates can create an atmosphere that is conditionally unstable. The EML is often warm enough to provide a significant capping inversion that can lead to a loaded gun sounding (Fig. 1.7). Strong inversions can lead to a build-up of CAPE and deep convection if the instability can be released. The magnitude of instability regulates the strength of any convection that may occur. If CAPE is small then updraft velocities will be small and convection may be shallow. Deep convection requires the build-up of instability, which can occur when there is a capping inversion.

Inversions can help create conditions favourable for deep moist convection; however they can also lead to convection initiation failure. If an inversion layer is too warm, saturated parcels may not become more buoyant than the surrounding environment and be able to ascend freely. If the inversion is deep, parcels may not acquire enough vertical momentum to be able to reach the LFC. Additionally, the convective environment is dynamic. It should not be assumed that the mid-level temperatures and lapse rates will remain constant when waiting for the updrafts to reach the LFC. For instance, strong advection from an EML source region could reinforce the capping inversion by increasing mid-level temperatures.

Deep, moist convection can only occur if instability can be released. The
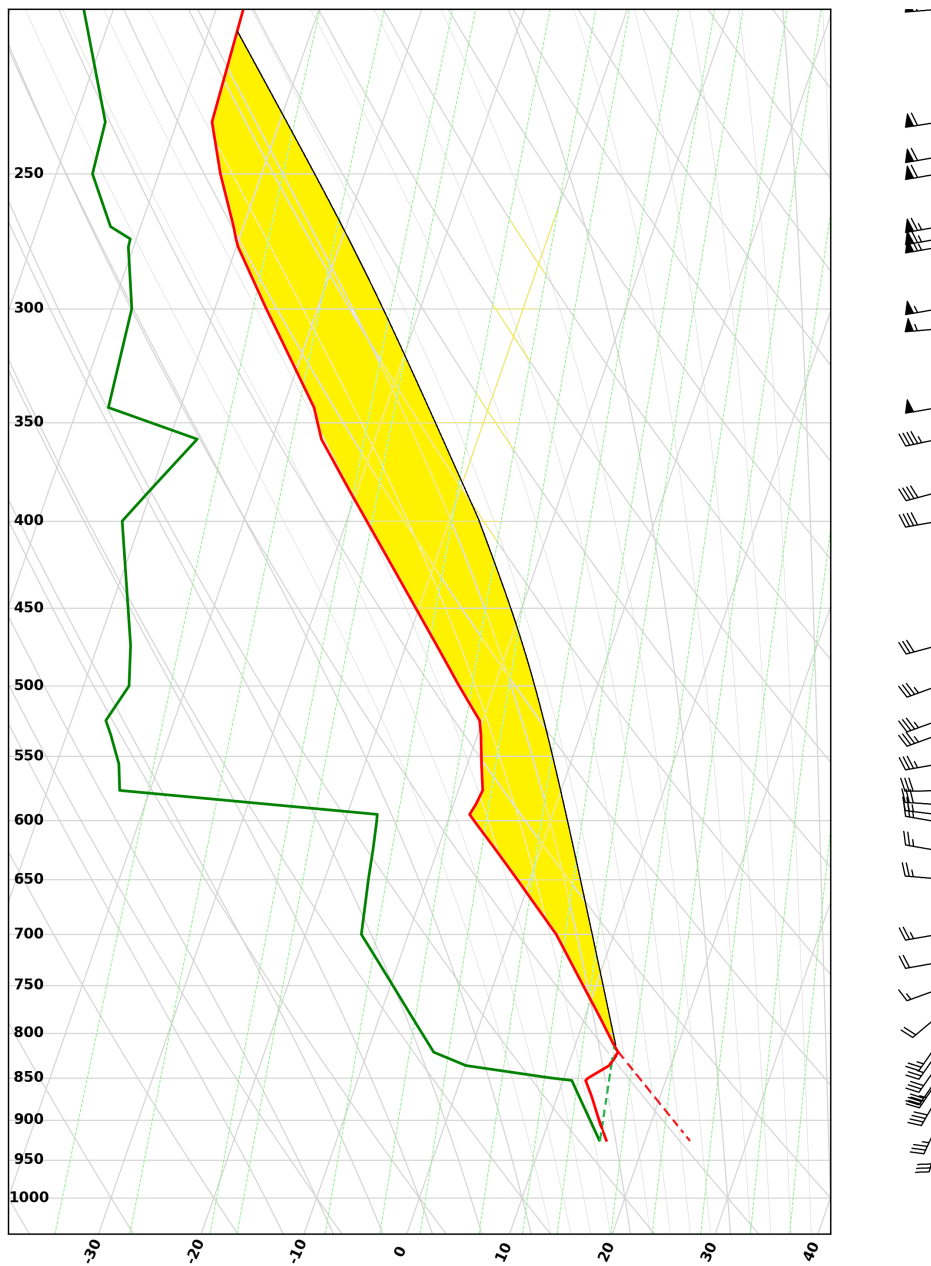
Figure 1.7:   A vertical profile of temperature (red) and dewpoint (green) for Oklahoma City, Oklahoma at 1200 UTC on 3rd May 1999. An inversion existed at approximately 850hPa, sitting atop moist air sourced from the Gulf of Mexico. If an air parcel can be lifted above the inversion, or the inversion weakened or removed, convective instability can be released. The red dotted line shows the theoretical ascent of a heated surface parcel to its lifting condensation level. Once it reaches the level of free convection it can utilize large amounts of CAPE (shown in yellow).

release of instability is primarily caused by weakening or erosion of a capping inversion. However, the removal of convective inhibition (CIN) is not necessarily enough for sustained convection. Sometimes no storms will form despite negligible CIN (Ziegler and Rasmussen 1998; Markowski et al. 2006). Conversely, storms can sometimes form despite the presence of CIN (Mueller et al. 1993). Air parcels can acquire enough vertical momentum that they can ascend through the inversion and reach the LFC. Convection can initiate once CIN becomes minimal (Colby 1980, 1984). However, convection does not necessarily occur in areas where CAPE is maximized Colby (1984).

A capping inversion can be weakened by cooling such as provided by synoptic-scale ascent. To explain how synoptic-scale processes can produce ascent, it is beneficial to introduce some concepts of quasi-geostrophic theory.

## 1.5.1   Quasi-geostrophic theory

Quasi-geostrophic theory describes a theoretical atmosphere which is 'almost' geostrophic. That is, the flow is mostly geostrophic apart from small accelerations which act to restore thermal wind balance after thermal advection by the geostrophic wind. The atmospheric response to departure from thermal wind balance is adjustment of the height field via ageostrophic flow and vertical motion. The associated ageostrophic circulation is typically of order 1 cm s$^{-1}$ for synoptic-scale motion.

The quasi-geostrophic (QG) equations that represent the atmospheric response are shown below. They are known as the QG omega and QG height tendency equations respectively. A full derivation of these equations is provided by Bluestein (1992).

$$\left[\nabla_p^2 + \frac{f_0^2}{\sigma}\frac{\partial}{\partial p^2}\right]\omega = -\frac{f_0}{\sigma}\frac{\partial}{\partial p}\left[-\vec{V_g} \cdot \nabla_p(\zeta + f)\right] - \frac{R}{\sigma p}\nabla_p^2\left[-\vec{V_g} \cdot \nabla_p T\right] \qquad (1.1)$$

$$\left[\nabla_p^2 + \frac{f_0^2}{\sigma}\frac{\partial}{\partial p^2}\right]\chi = f_0\left[-\vec{V_g} \cdot \nabla_p(\zeta + f)\right] - \frac{f_0^2}{\sigma}\frac{\partial}{\partial p}\left[\frac{R}{p}(-\vec{V_g} \cdot \nabla_p T)\right] \qquad (1.2)$$

where p is pressure, f$_0$ is the coriolis parameter, $\sigma$ is a static stability parameter, V$_g$ is geostrophic wind, $\zeta$ is relative vorticity, f is planetary vorticity, R is the gas constant, T is temperature, $\omega$ is pressure velocity and $\chi$ is geopotential height

tendency.

The equations above can be useful in diagnosing synoptic-scale motion, as vertical motion and geopotential height tendency can be estimated by evaluating the right hand side of the equations. The left hand side of each equation may be considered to be the Laplacian of $\omega$ and $\chi$ respectively. For sinusoidal flow, the Laplacian of $\omega$ can be approximated to be proportional to $\omega$. This approximation is reasonably valid in the mid-troposphere (Trenberth 1978). Therefore by analysing the right hand side of (1.1) it is possible to get a sense of the vertical motion because vertical velocity and pressure velocity have opposing signs. The first term on the right hand side of (1.1) is the differential advection of absolute vorticity by the geostrophic wind. Taking this term in isolation, if cyclonic vorticity advection (CVA) increases with height then the term will be positive and $\omega$ negative (Note: $\frac{\partial}{\partial p}$ is analogous to $-\frac{\partial}{\partial z}$). Therefore positive vertical velocity would be expected.

A physical interpretation of the effect of cyclonic vorticity advection is that an air parcel exiting a vorticity maximum will have to reduce its vorticity to reach equilibrium with the surrounding environment. A reduction in vorticity at a given level is associated with divergence at that level. Assuming the vorticity maximum is aloft (500 hPa is a commonly analysed height), divergence will be associated with air ascending from beneath. Conversely, a parcel entering a vorticity maximum needs to gain vorticity to reach equilibrium, so there is convergence and air moves towards the surface.

Alternatively, the effects of vorticity advection can be considered in terms of layer thickness, as relative vorticity is proportional to the Laplacian of the height field. Consider a case where vorticity advection is analysed at two heights. Vorticity advection is typically of greater magnitude aloft than near the surface. Therefore, the vorticity tendency aloft is likely to be positive whereas close to the surface it may be considered to be negligible. The Laplacian of the height field will increase with time, and the layer thickness reduces. A reduction in layer thickness implies cooling in the layer, which must be associated with rising motion because diabatic processes are neglected in the omega equation.

Differential vorticity advection must be analysed with caution. Anticyclonic

vorticity advection (AVA) decreasing with height will be associated with the
same vertical motion tendency as CVA increasing with height. It is usually a
safe assumption that vorticity advection dominates at upper levels and thermal
advection dominates lower down. However, that is not always the case. Even if
CVA is occurring aloft, if it is also occurring at low levels then differential vortic-
ity advection could be minimal or even negative. Therefore, the vertical profile
of vorticity advection should be considered when diagnosing vertical motion.

The second term on the right hand side of (1.1) is a Laplacian of tempera-
ture advection by the geostrophic wind. Taking this term in isolation, if there is
warm advection the overall term will be positive. Therefore $\omega$ will be negative
and vertical velocity will be positive. Warm air advection is associated with ris-
ing motion because air parcels move along isentropes in the absence of diabatic
processes. Isentropes tend to be sloped, with parcels increasing in height as they
move up the sloped surfaces.

Warm air advection within a layer increases the layer thickness. An increase
in layer thickness implies height rises above the region of maximum thermal ad-
vection, with height falls below. Because relative vorticity is proportional to the
Laplacian of height field, there should be a relative increase in vorticity at low
levels and a relative decrease aloft. A local increase in vorticity at a given height
is associated with convergence and a local decrease with divergence. Mass con-
tinuity implies that divergence aloft with convergence beneath will be associated
with ascent.

The QG omega equation must also be interpreted with care as terms can
counteract each other. For instance, CVA might be increasing with height, while
cold advection occurs at low-levels, resulting in a cancellation of terms. Hoskins
et al. (1978) developed the Q-vector form of the omega equation, which avoids
the problem of cancellation. Q-Vectors represent the rate of change of tempera-
ture gradient on an isobaric surface. Vertical motion can be estimated from the
convergence of the Q-field (Sanders and Hoskins 1990).

Durran and Snellman (1987) reviewed methods of diagnosing vertical motion

for different forms of the omega equation. They found that there was good agreement between implied vertical motion and precipitation patterns when the QG omega equation was applied at 500 hPa level, agreeing with the assumption of Trenberth (1978). However, they emphasized the importance of analysing several heights as the traditional form of the QG omega equation performed less well at 700 hPa.

The QG omega equation is an elliptic partial differential equation that is computationally expensive to solve. Care must be taken when choosing a finite differencing method as poor implementation can produce noisy fields (Dunn 1991). In a review of techniques for estimating vertical motion, Durran and Snellman (1987) stated "The traditional form of the omega equation is not well suited for practical calculation." If calculating vertical motion numerically, they recommended using the Q-vector from of the omega equation.

## 1.5.2  Synoptic-scale Convective Processes

How can the omega equation be related to synoptic-scale systems? As upper-level troughs are associated with relative vorticity maxima, CVA usually occurs downstream. Given the assumption that vorticity is generally greater in magnitude in the upper-levels of the atmosphere than near the surface, CVA typically increases with height downstream of a trough. Therefore the region downstream of a trough is a preferred region for ascent. Lifting of an unsaturated layer will result in adiabatic cooling as it ascends. Therefore, large-scale ascent is a mechanism by which inversions can be weakened by cooling.

It may be tempting to think that cold air advection around the height of the inversion should reduce CIN. However, diagnosis of the QG omega equation shows that cold air advection within a layer is associated with descent. Colby (1984) examined convection initiation along a Kansas dryline. He found that the weakening of CIN associated with cooling aloft was somewhat offset by subsidence and warming just above the convective boundary layer.

Synoptic-scale ascent reduces CIN by lifting and cooling. However, the magnitude of synoptic scale lift is not large enough to explain dryline convection initiation on its own. Synoptic scale ascent is typically in the range of 1–10 cm

$s^{-1}$. With ascent of that magnitude it will take many hours to lift a parcel from the surface to the LFC (Doswell and Bosart 2001). Additionally, ascent would be likely to lead to widespread convection, rather than the more isolated modes that often occur at the dryline e.g (Bluestein and Parker 1993). However, mesoscale lift is usually much stronger, typically of order 1 $ms^{-1}$. At this velocity, raising a parcel to an LFC at an altitude of 2 km would take approximately 30 minutes. Clearly mesoscale ascent will be far more efficient at raising parcels to the LFC. It is thought that synoptic scale lift provided helps prime the profile so that mesoscale lift can initiate storms (Doswell and Bosart 2001; Markowski et al. 2006). A mesoscale region of ascent can either raise parcels through the remaining inversion or mesoscale processes can further weaken the capping inversion.

Reduction of CIN due to adiabatic cooling is not the only mechanism by which synoptic-scale processes can prime the atmosphere for deep moist convection. Lift helps deepen the low-level moist layer by vertically transporting moisture from the surface where it is more abundant. Additionally, lee cyclogenesis or the passage of a baroclinic system can promote northward moisture transport in the Southern Plains. This moisture advection can help offset the reduction in low-level moisture caused by mixing. Increased low-level moisture tends to increase CAPE and reduce CIN, increasing the likelihood of deep, moist convection.

### 1.5.3   Mesoscale Convective Processes

Cumulus congestus are often found in the vicinity of the dryline, implying that parcels frequently reach the LCL. However, this does not always result in deep convection, suggesting parcels often fail to reach the LFC e.g.(e.g., Demoz et al. 2006). The length of time parcels spend in a zone of ascent is important in determining the likelihood of deep convection. Ziegler and Rasmussen (1998) suggested that deep convection requires parcels to be lifted to the LFC before leaving a mesoscale updraft. Similarly, Markowski et al. (2006) concluded that the lack of a persistent area of mesoscale lift contributed to the failure of convection initiation along an outflow boundary. Convection initiation is favoured when lift occurs over a mesoscale region. In convection-allowing simulations of a dryline, Trier et al. (2015) found that sustained convection occurred within a broader zone of vanishing negative buoyancy as opposed to a narrower zone. Within the narrower zone, cells moved into an area of more hostile buoyancy

before they could mature.

Wind shear is important in determining the residence time of an air parcel in an area of mesoscale ascent. Peckham and Wicker (2000) studied the effects of wind shear and topography on a dryline. They found that stronger line-normal flow leads to shorter residence time for updrafts in the main zone of ascent. Furthermore, cross-dryline flow can strengthen the capping inversion due to subsidence and warming as air moves downslope. Demoz et al. (2006) speculated that convection initiation failure along a dryline in Oklahoma may have been due to a strong inversion and detrimental effects of entrainment in strong westerly wind shear. Wind shear affects the rate of entrainment by tilting updrafts; the rate of entrainment is larger for more tilted updrafts. However, the tilting of an updraft is also a function of buoyancy. Stronger updrafts are less affected by entrainment than weaker updrafts.

Although roughly vertical near the surface, the dryline tilts eastward over the moist air aloft due to the westerly wind shear that is typical in the dryline environment. The westerly shear also enables moisture transport across the dryline. Hane et al. (1997) noted an elevated moist layer to the east of the dryline. They proposed that moist air from the top of an updraft plume is transported in the westerly flow. Ziegler et al. (1995) found a similar moist layer to the east of a modelled dryline. They suggested that the EML may be moistened by its journey over dryline. This raises the question of whether the presence of updrafts helps weaken the capping inversion by moistening the EML.

Updrafts transport moisture vertically through the boundary layer. Demoz et al. (2006) analysed a dryline during IHOP primarily using radar and LIDAR data. They observed updrafts exceeding 3 ms$^{-1}$ with widths typically between 2 and 5 km. These updrafts were more moist than the surrounding environment and extended through the depth of the boundary layer. Moistening of the boundary layer is likely to reduce the reduction of buoyancy in updrafts due to entrainment of drier air. Entrainment can also be caused by downdrafts transporting air from the inversion into the boundary layer below. In the absence of moisture advection, strong mixing within the boundary layer tends to reduce surface dew points as the moisture is redistributed through the boundary layer. However, replenishment

of low-level moisture by evapotranspiration or moisture advection can counteract the reduction of surface moisture. The maintenance of low-level moisture may be important in determining whether deep, moist convection can occur. Weldegaber et al. (2011) hypothesized that deep and persistent boundary layer moisture is required for convection initiation at the dryline.

### 1.5.4   Small-scale Convective Processes

Low-level moisture quantities are not uniform along the dryline. Drylines often feature localized areas of enhanced convergence. In a modelling study of a dryline in Texas, Xue and Martin (2006b) found that updrafts grew faster and were more intense when they stayed within the zone of enhanced convergence. This result is consistent with the findings of Ziegler and Rasmussen (1998) and Markowski et al. (2006) who emphasized the importance of updraft residence time within an area of mesoscale lift.

Convergence bands can form both ahead of and behind the dryline due to the presence of horizontal convective rolls (HCRs) within the boundary layer. Preferential cloud development occurs in locations where HCRs intersect the dryline (Weckwerth et al. 1997; Peckham et al. 2004). Xue and Martin (2006b) developed a conceptual model of the interaction between horizontal convective rolls and the dryline (Fig. 1.8). They propose that enhanced convergence occurs where HCRs intersect the dryline. Along the dryline, misocyclones are found in between the areas of enhanced convergence.

Other studies have also found misocyclones along the dryline e.g (Pietrycha and Rasmussen 2004; Wakimoto and Murphey 2009; Murphey et al. 2006). Murphey et al. used LIDAR and Airborne Doppler radar to study a dryline in Kansas. They found that updrafts were generally located to the north of misocyclones and initiated in regions of enhanced mixing ratio. As storms grew the updrafts became more collocated with the misocyclones. Recent studies have suggested that dryline misocyclones move northwards along the dryline with a speed similar to that of the mean wind in the boundary layer (Buban et al. 2007, 2012).

Convergence and cloud formation may also be related to the motion of the dryline. Pietrycha and Rasmussen (2004) used mobile mesonet data to study three drylines in Texas. They noted that cloud dissipation was observed in tandem with
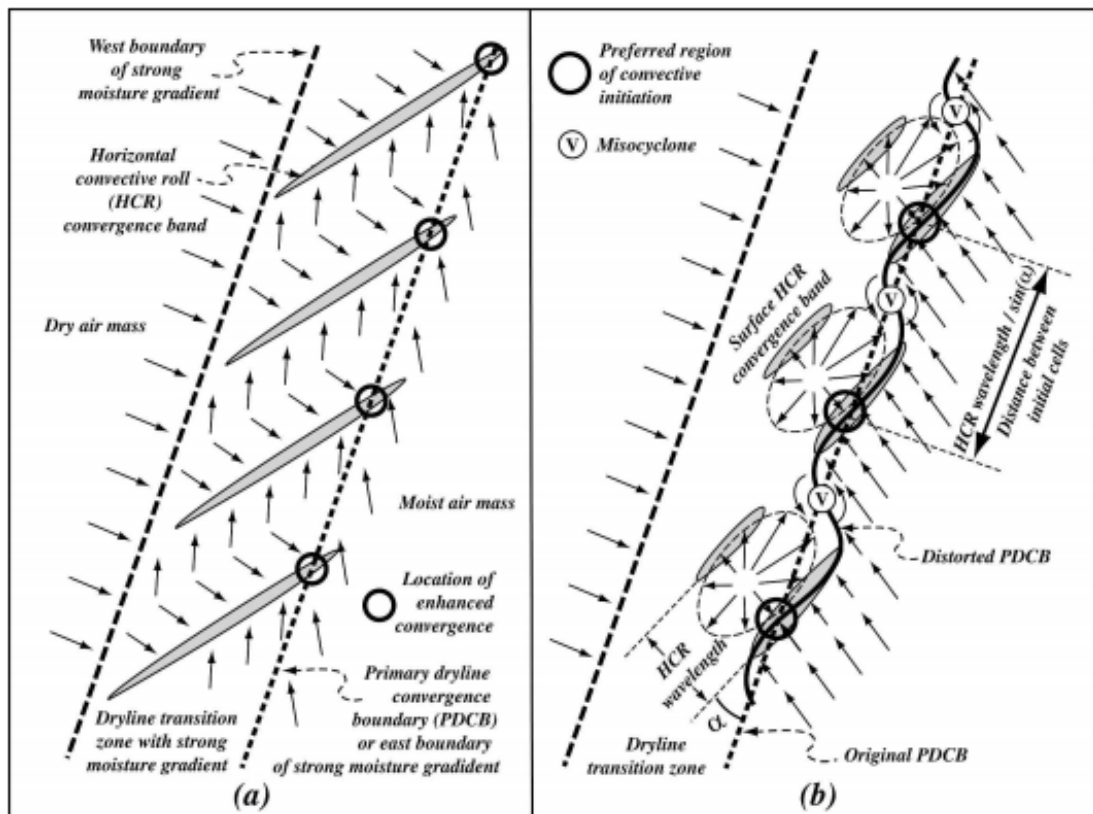
Figure 1.8: A conceptual model of the interaction between the dryline and horizontal convective rolls. In (a), the black circles denote where convergence is enhanced where the dryline intersects the horizontal convective roles. Misocyclones are located along the dryline in-between the rolls, and are denoted by the circles containing "V" in (b). Circulations associated with the misocyclones act to enhance convergence to their north and south. Figure from Xue and Martin (2006b).

sudden increases in dryline motion and weakening moisture gradients. However, they offered the caveat that with such a small sample size one must be careful in interpreting this result. Does the westward propagation of a density current oppose the westerly shear at the dryline as suggested by Ziegler et al. (1995)? If so, it is possible that a weakening of the density current decreases convergence and cloud formation and leads to eastward propagation of the dryline.

**Summary**

Convection initiation at the dryline occurs due to processes across a wide-range of spatial scales. Synoptic-scale ascent is thought to condition the atmosphere to deep, moist convection. Ascent is associated with adiabatic cooling which

reduces CIN. Furthermore, large-scale convergence provided by lee cyclogenesis can bring air parcels close to saturation (Crook and Moncrieff 1988). However, synoptic-scale ascent is not sufficient to reduce CIN to zero and initiate deep convection.  The mesoscale dryline circulation enhances moisture convergence and helps to further reduce CIN. Deep, moist convection can occur if parcels can reach the LFC before leaving a mesoscale updraft (Ziegler and Rasmussen 1998). If mesoscale convergence and ascent is not sufficient, localized convergence such as that provided from the interaction of the dryline with HCRs can aid convection initiation (Xue and Martin 2006a,b).

## 1.6   Dryline Climatologies

Much research into drylines has been performed by way of case studies, such as events documented during IHOP (Weckwerth et al. 2004), but there have been very few climatologies of drylines or dryline convection. Rhea (1966) performed the first substantial analysis of drylines within the Great Plain. He used three-hourly surface charts to identify drylines, and required a dew point discontinuity of at least $10°$ F between a station and its nearest neighbour. Drylines were identified in approximately 45% of April, May and June days for 1959–1962. Schaefer (1973) extended these criteria to include an approximately uniform moisture field to the east of the dryline with a mean value of $50°$. He also required the dew point discontinuity last at least six hours and a diurnal reversal of sign of the temperature gradient across the dryline. A dryline was present for all or part of 41% of the days he examined.

Peterson (1983) used similar criteria to Schaefer in a dryline study spanning 1970–79; however the domain chosen was much smaller, mainly covering the west Texas region. Dryline days occurred on average 30 days a year, ranging between 26 and 58 days in any given year. The study by Peterson (1983) found a lower dryline frequency than that of Rhea (1966) and Schaefer (1973). This lower frequency may be due to the smaller domain chosen, which is likely to have excluded drylines in some regions of Oklahoma and Kansas.

The most recent climatology of drylines was created by Hoch and Markowski (2005). They used 0000 UTC surface observations for April, May and June for

1973–2002. A dryline required confluent flow and a specific humidity gradient of
at least $3 \, \mathrm{g \, kg^{-1}}$ per 100 km. They found that drylines occurred on approximately
32% of days, a lower frequency than found by Rhea (1966) and Schaefer (1973).
It is possible that the criteria used by Hoch and Markowski were more restrictive,
as they only identified drylines that were present at 0000 UTC. Comparison of the
studies is difficult due to the different criteria used. The American Meteorological
society defines the dryline as "A low-level mesoscale boundary or transition zone
hundreds of kilometers in length and up to tens of kilometers in width separating
dry air from moist air"(AMS 2016). However, there is no universal standard of
required moisture gradient for a dryline. Gradients in surface dew point are con-
venient to use in dryline identification because they are reported by most weather
stations. However, dew point varies with pressure, so large altitude differences
between stations could lead to a misleading analysis. A conserved quantity such
as water vapour mixing ratio which Hoch and Markowki used is arguably better.

Although the dryline studies performed by Rhea, Schaefer and Peterson were
all within the southern Great Plains, only two studies have specifically investi-
gated dryline position. Hoch and Markowski (2005) found the dryline was most
commonly located around 101°W at 0000 UTC. They found that the average dry-
line position trended westwards through the season. Drylines can progress farther
east than the Great Plains, especially in synoptically active cases. Although they
did not distinguish between quiescent and synoptically active drylines, Hoch and
Markowski noted that "virtually all" dryline cases east of 97°W were associated
with a migrating cyclone.

There have been fewer climatologies of dryline convection than drylines in
general. The first, and arguably most comprehensive study of dryline convection
was performed by Rhea (1966). He used microfilm of hourly radar summaries
to identify convective echoes associated with the dryline. Convection occurred
within 200 nautical miles (370 km) of the dryline on 70% of days with an identi-
fiable dryline, with most new echoes forming within 50 n mi of the dryline. After
examining a subset of these cases he found that 45% of new radar echoes formed
between 1400 and 1600 Central time.

Bluestein and Parker (1993) examined hourly summaries of radar data to create a climatology of isolated severe storms and their initial storm mode when formed at the dryline. Over half of the storms they studied were initially isolated cells. However, they were unable to discern significant differences in the environment for storms that were not initially isolated. Cells were identified by analysing WSR-57 microfilm, but they speculated that the use of new techniques such as Doppler radar or high resolution surface observations may help with resolving smaller scale variations. The WSR-57 radar network is now obsolete and has been replaced by WSR-88D radar. A primary advantage of the WSR-88D network is Doppler capability which aids in the detection of severe weather (Whiton et al. 1998).

Although not a climatology of dryline convection, Ziegler and Rasmussen (1998) examined the processes associated with convection initiation at the dryline using special mesoscale observations obtained from three field projects. They found that "Cumuli were concentrated within the interval from 10 km west to 40 km east of the dryline", with these cloudy regions being associated with deep mesoscale moisture convergence. However, on two occasions deep moist convection did not occur despite the presence of zero CIN. The authors determined that deep convection requires parcels to be lifted to the LFC before leaving a mesoscale updraft, and suggest that strong deep layer convergence is a better predictor of convection initiation than surface convergence alone.

The work of Ziegler and Rasmussen (1998) greatly improved knowledge of mesoscale dryline processes. However, there has been very little work that directly relates to the influence of synoptic-scale processes on dryline convection. Rhea (1966) documented the distance of an upstream feature from the first initiation point based upon 1200 UTC charts. He found that only 29% of convective drylines did not have some form of discrete feature in the wind or temperature field at 500 hPa. The most common feature associated with convective drylines was a weak upstream trough. The availability of reanalysis data in recent years means that a study similar to that performed by Rhea could now be conducted in more detail. Furthermore, digital radar archives can provide a much more precise estimate of initiation time and location than the microfilm summaries he had available.

More recently, Schultz et al. (2007) investigated the relationship between synoptic processes and dryline intensity. Drylines were categorised by the magnitude of the west-east dewpoint gradient. "Strong" drylines featured a short wave trough in the westerly flow to the west of the region with an attendant strong surface low over eastern New Mexico leading to southerly flow over the southern Great Plains. The authors determined that synoptic-scale forcing has an important contribution in providing the confluence and intensification of drylines. A linear relationship exists between confluence and the west-east dewpoint gradient.

The work of Schultz et al. did not differentiate between convective and non-convective drylines. However, "strong" dryline cases in their study were generally associated with greater convergence. What does this imply for convection initiation? All other things being equal, stronger convergence is more likely to raise parcels to the LFC. Crook and Moncrieff (1988) proposed that largescale convergence is a mechanism that brings parcels close to saturation. Furthermore, Ziegler et al. (1997) found that convection initiation is more likely in locations where moisture convergence reduces CINH to zero. Therefore the "strong" dryline cases may be more likely to produce deep moist convection.

Schumann and Roebber (2010) investigated the role of synoptic features and how they affect storm mode. Potential temperature advection on the dynamic tropopause was used as a measure of synoptic forcing, and synoptic-scale features were classified by shape and areal extent. They found that more organized convective modes (multicellular lines and isolated rotating) were associated with weaker forcing. The study also investigated the influence of the wind profile on the initial convective mode. It was found that multicellular storms were more likely with an increase in the meridional component of the 0–6 km mean wind. They also found that convective lines were more likely with an increase in the zonal component of the 0–6 km wind, when all convection and transitions were considered.

Schumann and Roebber did not study drylines specifically, and the study only contained 56 individual storm days. Their study did not discriminate between the types of initiating boundary, so it is uncertain how many of these cases were dryline convection. How might their results vary if they considered only dryline

convection? Drylines are generally north-south oriented, as opposed to fronts
which are far more variable. Therefore, the influence of the mean wind is likely
to be different for dryline storms than for the storms they studied, especially
when it is considered that they did not account for the orientation of the initiat-
ing boundary.

Dial et al. (2010) investigated convective mode for storms that form along
boundaries. They investigated the effects of shear, wind and ascent on convective
mode and transition. Drylines favoured discrete storm modes more than cold
fronts. It was suggested that one reason for the difference in preferred mode is
that drylines often feature strong westerly flow and capping. A narrow zone of
ascent and strong westerly flow can lead to shorter residence times for updrafts
at the boundary. Cloud layer boundary normal shear was found to be skilful
in determining whether storms would stay discrete. Higher values of boundary
normal shear resulted in longer lasting discrete storms.

The study by Dial et al. did not include non-convective drylines, therefore no
comparison could be made between convective and non-convective dryline envi-
ronments. However, some of the arguments they provided for why drylines favour
discrete modes may also apply to cases of convection initiation failure. Strong
westerly flow could reinforce a capping inversion and inhibit convection. Addi-
tionally, strong shear relative to buoyancy would lead to short residence times
and parcels would not reach the LFC before leaving the zone of ascent.

The synoptic environment plays a large role in determining the environmen-
tal wind shear. Forcing for ascent is likely to be associated with changes in the
wind shear in the vicinity of the dryline. How would changes in wind shear affect
entrainment rates and residence time of updrafts in mesoscale regions of ascent?
Additionally, would the advection of mid-level air from the west reinforce the
capping inversion?

Many questions remain as to the role of large-scale processes in dryline con-
vection initiation. For instance, are there large-scale differences between drylines
that produce deep, moist convection and those that do not? Or is dryline con-
vection initiation determined only by smaller-scale processes which are harder to

predict at short time scales? These questions suggest the need for a climatology of dryline convection initiation.

[Blank page]

# Chapter 2

# Paper 1: A Synoptic Climatology of Dryline Convection in the Southern Great Plains

# A Synoptic Climatology of Spring Dryline Convection in the Southern Great Plains

TREVOR MITCHELL[a]

*Centre for Atmospheric Science, Department of Earth and Environmental Sciences, University of Manchester,*
*Manchester, United Kingdom*

DAVID M. SCHULTZ

*Centre for Atmospheric Science, Department of Earth and Environmental Sciences, and Centre for Crisis*
*Studies and Mitigation, University of Manchester, Manchester, United Kingdom*

## ABSTRACT

A dataset of drylines within a region of the southern Great Plains was constructed to investigate the large-scale environments associated with the initiation of deep moist convection. Drylines were identified using NOAA/NWS Weather Prediction Center surface analyses for all April, May, and June days 2006–15. Doppler radar and visible and infrared satellite imagery were used to identify convective drylines, where deep, moist convection was deemed to have been associated with the dryline circulation. Approximately 60% of drylines were convective, with initiation most frequently occurring between 2000 and 2100 UTC. Composite synoptic analyses were created of 179 convective and 104 nonconvective dryline days. The composites featured an upper-level long-wave trough to the west of the Rockies and a ridge extending across the northern and eastern United States. At the surface, the composites featured a broad surface cyclone over western Texas and southerly flow over the south-central states. Convective drylines featured more amplified upper-level flow, associated with a deeper trough in the western United States and a stronger downstream ridge than nonconvective drylines up to 5 days preceding a dryline event. By the day of a dryline event, the convective composite features greater low-level specific humidity and higher CAPE than the nonconvective composite. These results demonstrate that synoptic-scale processes over several days help create conditions conducive to deep, moist convection along the dryline.

## SIGNIFICANCE STATEMENT

The southern Great Plains dryline separates moist air from the Gulf of Mexico from drier air farther west. Drylines sometimes initiate convective storms, that is, storms that produce lightning, tornadoes, and other extreme weather. We wanted to know if we could tell the difference between days when such storms occur and days when they do not. We found that there were distinctive weather patterns in the middle and upper troposphere that distinguished these two sets of days. These differences were apparent 3–5 days ahead of time, suggesting an opportunity for more lead time in forecasting such storms.

## 1. Introduction

The southern Great Plains dryline is a boundary that separates moist air originating over the Gulf of Mexico from drier air originating from the desert southwest (e.g., Rhea 1966; Schaefer 1973, 1974; McCarthy and Koch 1982; Schaefer 1986; Hane 2004; Hoch and Markowski 2005). The dryline is often associated with a

---

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

---

[a] Current affiliation: MetDesk, Wendover, United Kingdom.

zone of convergence beneath the ascending portion of a thermally direct circulation (e.g., Sun and Ogura 1979; Ziegler et al. 1995). This convergence zone is a preferred region for convection initiation (e.g., Fujita 1958; Beebe 1958; Miller 1959; McGuire 1962; Rhea 1966). However, low-level moist air east of the dryline is typically capped by warmer, drier air. The environment east of the dryline can also be hostile to deep convection because dry-air entrainment can reduce the buoyancy of ascending air parcels. For deep convection to occur, moist air parcels need to be lifted to their level of free convection before leaving the mesoscale updraft region (Ziegler and Rasmussen 1998). As incipient convection moves away from the dryline, it requires sufficient instability for deep convection to develop. Moisture, instability and lift are necessary ingredients for deep, moist convection to occur (Doswell 1987), and the dryline is often a region where these ingredients overlap. However, despite the apparent presence of these ingredients, sometimes deep, moist convection does not develop along or near the dryline (Ziegler et al. 1997).

Uncertainty over whether deep convection will occur is not limited to drylines. Convection initiation in general is sensitive to small changes in conditions, such as changes in temperature and moisture within the planetary boundary layer (Mueller et al. 1993; Crook 1996) and changes in lapse rate above the level of free convection (Houston and Niyogi 2007). Improving understanding of convection initiation has been a motivation of research projects such as the International H$_2$O Project (IHOP_2002; Weckwerth and Parsons 2006) and the Spring Forecast Experiment of 2011 (Kain et al. 2013). Although not focused on dryline convection specifically, some cases of convection initiation during those projects occurred along the dryline. However, robust conclusions regarding dryline convection initiation are hard to draw due to the small number of dryline events. For instance, the Spring Forecast Experiment of 2011 included only five cases of dryline convection initiation. Despite this limitation, the literature on dryline convection contains several examinations of the causes of convection initiation failure along the dryline, particularly during IHOP.

One such example of initiation failure during IHOP was examined by Demoz et al. (2006). They speculated that a relatively dry near-surface layer, a strong capping inversion, and moisture detrainment in a dry layer between the LCL and the LFC were detrimental to initiation. The presence of a dry layer between the LCL and LFC was also thought to have contributed to initiation failure during the Verification of the Origins of Rotation in Tornadoes Experiment (VORTEX;

Weiss and Bluestein 2002) and in an IHOP case examined by Cai et al. (2006). Cai et al. (2006) also found that midlevel subsidence contributed to a strong capping inversion over the dryline. Midlevel subsidence and warming were also blamed for initiation failure in a case from VORTEX studied by Richter and Bosart (2002) as a short-wave ridge migrated over the Texas Panhandle region. The importance of migratory mesoscale features for convection initiation was also demonstrated by Hill et al. (2016). In a simulation of two dryline events, they found that convection initiation was sensitive to the location of a 700-hPa short-wave trough.

Although the passage of features such as short-wave ridges and troughs are primarily mesoscale processes, variables such as the strength of a capping inversion, the vertical distribution of moisture, and the magnitude of moisture detrainment can also be affected by large-scale processes. Schultz et al. (2007) showed that synoptic-scale processes are important in regulating the strength of the dryline. Stronger drylines were associated with passage of a short-wave trough in the ambient westerlies, favoring surface cyclogenesis, increased confluence, and tighter moisture gradients. Short-wave troughing may also increase the likelihood of convection given the presence of a dryline. Rhea (1966) found that 71% of cases of storm formation within 100 n mi (185 km) of the dryline were associated with a discrete feature at 500 hPa in the temperature or wind field at 1200 UTC. The most frequent of these features was an upstream trough. However, Rhea did not examine the synoptic environment for drylines that did not produce convection. Questions remain as to the role of large-scale processes in dryline convection initiation. Does short-wave troughing west of the dryline increase the chances of convection? More generally, are there large-scale differences between drylines that produce deep, moist convection and those that do not? Or is dryline convection initiation determined only by smaller-scale processes that are harder to predict at short time scales? These questions suggest the need for a climatology of dryline convection initiation.

Although there have been several studies of the mechanisms by which dryline convection occurs, there has only been one previous climatology of dryline convection specifically. Rhea (1966) identified thunderstorm formation along drylines on April, May, and June days for 1959–62. The dryline location was estimated using 3-h surface charts, and thunderstorms were identified by examining hourly radar summaries for new echoes within 200 n mi (370 km) of the dryline. However, Rhea's work was somewhat limited by the lack of meteorological data available. Present-day observation networks such as the state mesonets in Texas and Oklahoma make it easier to determine dryline

location and subsequent convection today. Furthermore, the availability of atmospheric reanalyses provides more options for examining dryline conditions, including the potential for automated detection (Clark et al. 2015).

To address the question of whether there are large-scale differences between convective and nonconvective drylines, we present a climatology of dryline convection initiation in the southern Great Plains by separating days with drylines into those that produced deep, moist convection and those that did not. In section 2, dryline days are identified using NOAA/NWS Weather Prediction Center surface analyses, then days with or without convection are identified using NEXRAD radar mosaics and visible and infrared satellite imagery. Section 3 presents a climatology of convection at the dryline. In section 4, the differences between convective and nonconvective drylines are discussed by examining composite dryline environments constructed from North American Regional Reanalysis (NARR) data. Section 5 examines composite dryline environments in the days preceding a dryline event. Finally, section 6 summarizes the main results of this article.

## 2. Data and methods

### a. Dryline selection

Drylines have commonly been identified using gradients in moisture obtained from surface station data (e.g., Rhea 1966; Schaefer 1973; Peterson 1983). Dewpoint temperature has often been used as a measure of surface moisture because it is easily obtained from NWS observations (Schaefer 1986). However, the dryline is not always identified by exclusively considering moisture gradients. In addition to requiring a 10°F (5.6°C) dewpoint discontinuity between neighboring stations, Rhea (1966) also required an organized line of veering surface wind. Schaefer (1973) expanded these criteria to include a diurnal reversal of the direction of the temperature gradient across the dryline. More recently, in their climatology of drylines, Hoch and Markowski (2005) excluded the necessity for a wind shift or diurnal reversal of temperature gradient. They also considered specific humidity rather than dewpoint, because specific humidity is conserved when gradients in pressure are present due to sloping terrain. Finding a reliable method to identify drylines is an ongoing problem (e.g., Coffer et al. 2013; Clark et al. 2015), with no widely used method to automate dryline identification. Because the focus of the present study was to identify dryline convection rather than develop a method to identify

drylines, we therefore used an existing dataset that includes previously identified drylines by an independent organization.

The Weather Prediction Center (WPC) combines analyses from the Hydrological Prediction Center, Ocean Prediction Center, National Hurricane Center, and the Honolulu Weather Forecast Office to form the Unified Surface Analysis, available at 3-h intervals (Berg et al. 2007). These analyses are created by experienced surface analysts using a library of conceptual models with the aid of data from surface observations, satellite data, and model analyses, and the analyses depict synoptic and mesoscale features such as fronts, troughs, outflow boundaries, and drylines (NOAA 2013). According to the Unified Surface Analysis Manual, "A tight 14°C (25°F), or a broader 17°C (30°F), dewpoint gradient is used to help determine the existence of a dryline. The dryline does not have to be the leading edge of all the change in the dewpoint, merely where the best gradient/leading edge of foehn winds exists." Dryline data were obtained from an online archive of surface analyses produced by the WPC, with analyses available from May 2005 onward. The choice of the WPC analysis for our dryline provided an independent analysis of dryline occurrence and location, minimizing any bias on our analysis. It also avoided us having to develop, verify, and implement criteria for dryline occurrence from gridded model output. Thus, in this instance, a manual method of detection was deemed superior to an automated method (Schultz 2009, 210–212).

WPC analyses were obtained for all April, May, and June days for 2006–15. These months were chosen for two reasons: 1) consistency with the months examined in previous dryline climatologies (e.g., Rhea 1966; Schaefer 1973; Peterson 1983; Hoch and Markowski 2005) and 2) convective storms most often initiate along the dryline during these months (Hoch and Markowski 2005).

Although most dryline studies have focused on late spring and early summer, previous climatologies differ in the size and location of the region in which drylines were considered. Schaefer (1974) only selected drylines in the south-central United States. Peterson (1983) was even more restrictive than Schaefer, only studying drylines in the West Texas region. However, Hoch and Markowski (2005) chose a broader region, identifying drylines within the Great Plains. The domain chosen for the present study (Fig. 1) was guided by the work of Hoch and Markowski (2005) who found that drylines were most frequent around 101°W, with approximately 98% of drylines located between 97° and 104°W. Thus, to limit the domain to be considered, we studied drylines
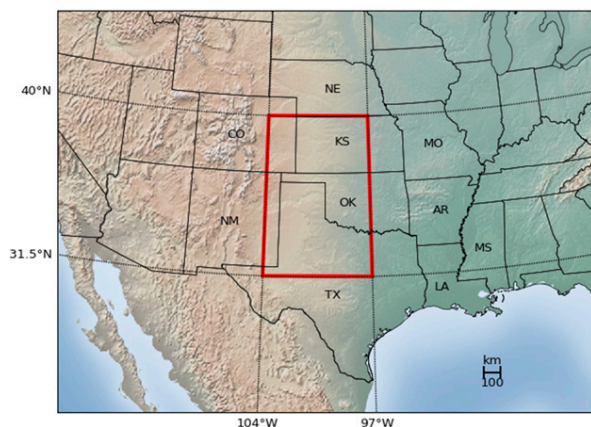
FIG. 1. The domain chosen for the dryline study is enclosed by the red box.

within a domain with a western boundary imposed along a line at 104°W, and an eastern boundary imposed at 97°W. The western boundary was chosen to incorporate dryline convection in western Texas and eastern New Mexico while avoiding convection initiation associated with upslope flow. Although dryline formation is dependent on the sloping terrain, topographic features can lead to localized regions of enhanced convergence, increasing the likelihood of convection (e.g., Banta 1984). The effects of local features on dryline convection are beyond the scope of this study. For similar reasons, the southern boundary was restricted to 31.5°N to avoid including the higher terrain in the Trans Pecos region (e.g., Nielsen et al. 2016). The northern boundary was imposed at 40°N, coincident with the Kansas–Nebraska state line. Although drylines can be observed north of this boundary, drylines are more frequently observed in Texas, Oklahoma, and Kansas (Schaefer 1986).

A day was classified as having a dryline if at least 100 km of a WPC-analyzed dryline was present within the domain in any analysis between 1500 and 0300 UTC the following day. Although drylines can occur outside of this time frame (e.g., see Fig. 5b in Hane 2004), drylines typically have a diurnal cycle with a peak in intensity in the late afternoon and early evening. Convection associated with the dryline also has a diurnal cycle. Rhea (1966) found that the first occurrence of convective echoes in the vicinity of the dryline was most common between 1900 and 2100 UTC, with no occurrences before 1500 UTC or after 0200 UTC. Therefore, the use of analyses between 1500 and 0300 UTC seems appropriate to capture most events.

Most dryline days featured drylines being present in more than one of the WPC analyses produced.

However, occasionally there were inconsistencies between consecutive analyses. An example of one such inconsistency is a feature analyzed as a cold front in the morning, but identified as a dryline in a subsequent analysis. We believe that these cases represent the same feature. Overnight, dry air west of the dryline often cools more quickly than the moist air to the east overnight, resulting in a west–east temperature gradient by morning. This temperature gradient may be misidentified as a cold front, before the gradient reduces or even reverses during the day. Because a dryline is only required to be present on one analysis on a given day, these instances were included, despite the duration of the dryline being in question.

Drylines were identified on 33 days per season on average, corresponding to approximately 36% of all days analyzed. This frequency of dryline occurrence is greater than that obtained by Hoch and Markowski (2005), who found drylines on 32% of the days they analyzed over a 30-yr period. They identified drylines using gradients of specific humidity obtained from surface observations, requiring a horizontal gradient of $3\,\mathrm{g\,kg^{-1}}\,(100\,\mathrm{km})^{-1}$. The most likely explanation why Hoch and Markowski identified a lower proportion of drylines is the use of a more restrictive time constraint. They required a dryline to be present at 0000 UTC, whereas the criteria used in this study allowed for a dryline at any time between 1500 and 0300 UTC. Sometimes a dryline can be present in the late morning or early afternoon, but may no longer exist by 0000 UTC. One such example is when a cold front moves east faster than the dryline, and they eventually merge.

### b. Identifying dryline convection

The Iowa Environmental Mesonet generates and archives national radar mosaics, which are derived from base reflectivity output from the Next Generation Weather Radar (NEXRAD) network. The mosaics are available at 5-min intervals and can be viewed using the NCEI/NOAA Interactive Radar Map Tool (NCEI 2016). Potential dryline convection was identified using the radar tool, with a minimum radar reflectivity factor of 40 dBZ required for a continuous period of at least one hour. Echoes that satisfied these criteria were most likely indicative of deep, moist convection.

Echoes were not considered if they were deemed to have been caused by a feature other than a dryline. Preexisting convection such as a mesoscale convective system was excluded. Cases of convection that could be identified as being initiated along an outflow boundary were also excluded, as well as that associated with

frontal boundaries. However, convection that occurred at the intersection of a dryline and convective outflow or a front was considered. On occasion, a dryline was present during the afternoon, but was later overtaken by a cold front. Echoes were rejected if convection was not determined to have initiated along the dryline before it was overtaken by a cold front, even if subsequent echoes fulfilled the reflectivity criterion.

Visible and infrared satellite imagery were used to help distinguish convection initiating along the dryline from convection initiating along outflow boundaries or fronts. The dryline can often be spotted in visible imagery as a thin line of cumulus or the western edge of a cumulus field. When it was not possible to confidently identify the dryline using visible satellite imagery, infrared images were used. Infrared satellite images were useful in distinguishing drylines from fronts, particularly when the frontal temperature gradient was strong.

When a combination of visible and infrared satellite images failed to distinguish dryline convection, Doppler radar imagery was sourced for the nearest radar site with the aim of identifying radar fine lines. Radar fine lines are lines of low reflectivity returns that are not associated with precipitation. The returns are associated with scattering by insects and typically occur in areas of low-level convergence (Russell and Wilson 1997). The presence of fine lines is helpful in detecting boundaries such as drylines, fronts, and outflow (e.g., Geerts and Miao 2005).

On 41 occasions, the use of satellite and radar imagery failed to distinguish dryline convection from other types of convection. In these cases, Storm Prediction Center mesoscale discussions and National Weather Service (NWS) area forecast discussions were used as a supplement to the WPC analyses. Mesoscale discussions are usually issued in advance of severe weather and contain a forecaster's thoughts about a mesoscale event within the forecast area. The product is often a combination of a text forecast plus a mesoscale analysis, which can sometimes offer greater detail than the WPC surface analyses. However, mesoscale discussions are issued on an ad hoc basis and are not available every day, as opposed to area forecast discussions, which are issued by NWS offices several times a day. Area forecast discussions usually give a detailed breakdown of the forecast for a NWS forecast area and often include a summary of the challenges and general thoughts of the forecaster when making their forecast for the area (NOAA 2016).

Once dryline convection had been established, echoes were also required to have initiated no farther than 100 km from the dryline in a perpendicular distance to the dryline orientation. The 100 km is consistent with Fig. 4 in Rhea (1966) who found first new radar echoes formed within 50 n mi (93 km) of 69 nonconsecutive dryline days. More recently, Ziegler and Rasmussen (1998) used special mesoscale observations from three drylines studied during separate field projects. One of the cases they examined featured high-based cumulus 55 km east of the dryline. They found a peak cumulus frequency approximately 15 km east of the dryline, but urged caution when interpreting the results of such a small sample. Given the time it takes for initial cumuli to develop to 40-dB$Z$ echoes, movement of the convective storms of the order of tens of minutes and tens of kilometers could have placed the storms farther east, as much as 100 km eastward.

Initiation was deemed to occur at the first occurrence of a 40-dB$Z$ echo that maintained or exceeded that intensity for a continuous period of at least one hour. This threshold is consistent with previous literature on convection initiation (e.g., Parker and Johnson 2000; Fowle and Roebber 2003; James et al. 2005; Trapp et al. 2005; Hocker and Basara 2008; Grams et al. 2012). Because radar scans are available at 5-min intervals, the recorded time of initiation was at worst a late estimate by four minutes, but never early. At the time of initiation, the location of initiation was defined as the center of a local maximum in reflectivity.

## 3. Climatology of dryline convection

Three-hourly WPC surface analyses were analyzed between 1500 and 0300 UTC for all April, May, and June days 2006–15. Of the 329 drylines identified, 199 (60%) were associated with convective storms as defined in section 2b. The other 130 days did not fulfill the criteria of deep, moist convection defined in section 2b. The proportion of dryline days that produced deep, moist convection was less than that found by Rhea (1966) in his study of thunderstorm formation along drylines. He found new radar-echo development within 200 nautical miles (370 km) of the dryline for 70% of drylines identified. The most likely explanation for the lower frequency of dryline convection found in this study is the more restrictive criteria used. Echoes were required to be within 100 km of the dryline, rather than 370 km used by Rhea. The lower frequency may also be explained by the different durations examined. Rhea's study spanned only 4 years and may have sampled a period where dryline convection was more frequent than the long-term average.

The importance of the dryline sampling period is demonstrated when assessing the interannual variability
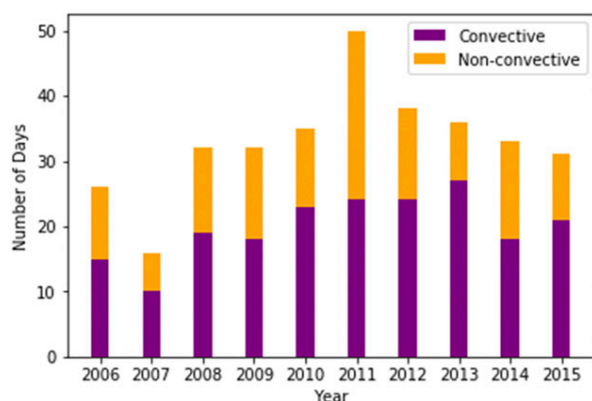
FIG. 2. The number of convective (purple) and nonconvective (orange) dryline days that occurred in a given year.
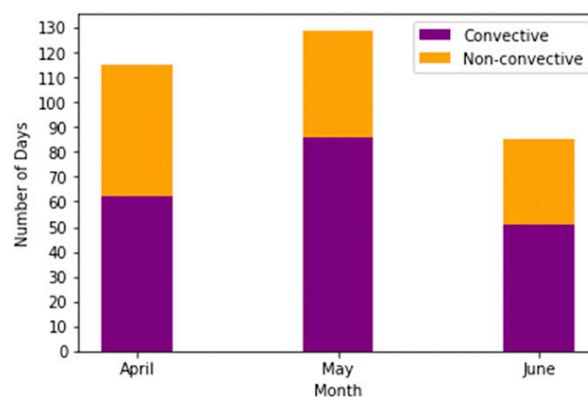


FIG. 3. The number of convective (purple) and nonconvective (orange) dryline days (2006–15) separated by the month in which they occurred.

of dryline convection (Fig. 2). Dryline frequency shows large year-to-year variation with as few as 16 dryline days in 2007 to as many 50 in 2011. As with dryline frequency, there appears to be large interannual variability in dryline convection. The number of convective dryline days ranged from as few as 10 in 2007 to as many as 27 in 2013.

In addition to the variability in dryline frequency between years, the frequency of both drylines and dryline convection also varied within a season. Both dryline days and convective dryline days were most frequent in May, followed by April then June (Fig. 3). These results are similar to that of Hoch and Markowski (2005) who found that peak dryline frequency occurred in mid- to late May. When broken down by week (Fig. 4), dryline frequency appears to reduce toward the end of June. Hoch and Markowski also found a reduction in dryline frequency at the end of June. They hypothesized that the onset of the North American monsoon season moistening air over the elevated mixed-layer source regions such as the desert southwest may be responsible for the reduction in dryline occurrence. The moistening of these regions would likely reduce the moisture gradient between the tropical continental air and moist air from the Gulf of Mexico.

## a. Convection initiation

The first occurrence of dryline convection initiation most commonly occurred in the midafternoon to early evening (Fig. 5a). In total, 71% of convective drylines had first initiation between 1900 and 2300 UTC, with a peak in frequency between 2000 and 2100 UTC. This result is not unexpected, as the conditions for dryline convection are often optimal around the time of maximum heating. Convergence

at the dryline is often at a maximum in the mid- to late afternoon (Ziegler et al. 1995). Furthermore, instability generally increases with surface temperature while surface heating can help reduce convective inhibition.

A midafternoon peak in convection initiation is consistent with previous studies of both drylines and surface-based convection more generally. Rhea (1966) found 45% of first echoes associated with the dryline occurred between 1900 and 2100 UTC. More recently, the International H$_2$O Project studied convection initiation and evolution in the southern Great Plains. During the field campaign, surface-based initiation most commonly occurred between 1800 and 2100 UTC (Wilson and Roberts 2006). However, only 6 out of 55 initiation episodes featured a dryline.

In the present study, the first occurrence of convection initiation was generally most common in the Texas Panhandle, but much less common farther
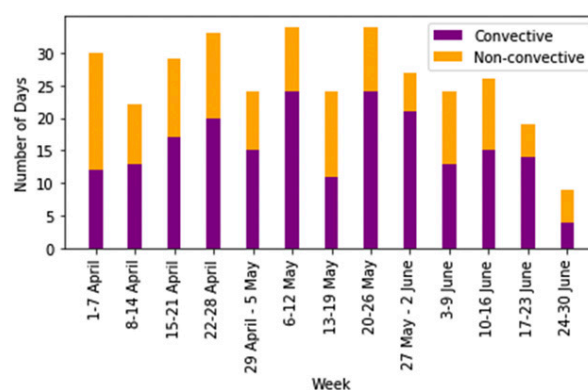


FIG. 4. The number of convective (purple) and nonconvective (orange) dryline days (2006–15) separated by the week in which they occurred.
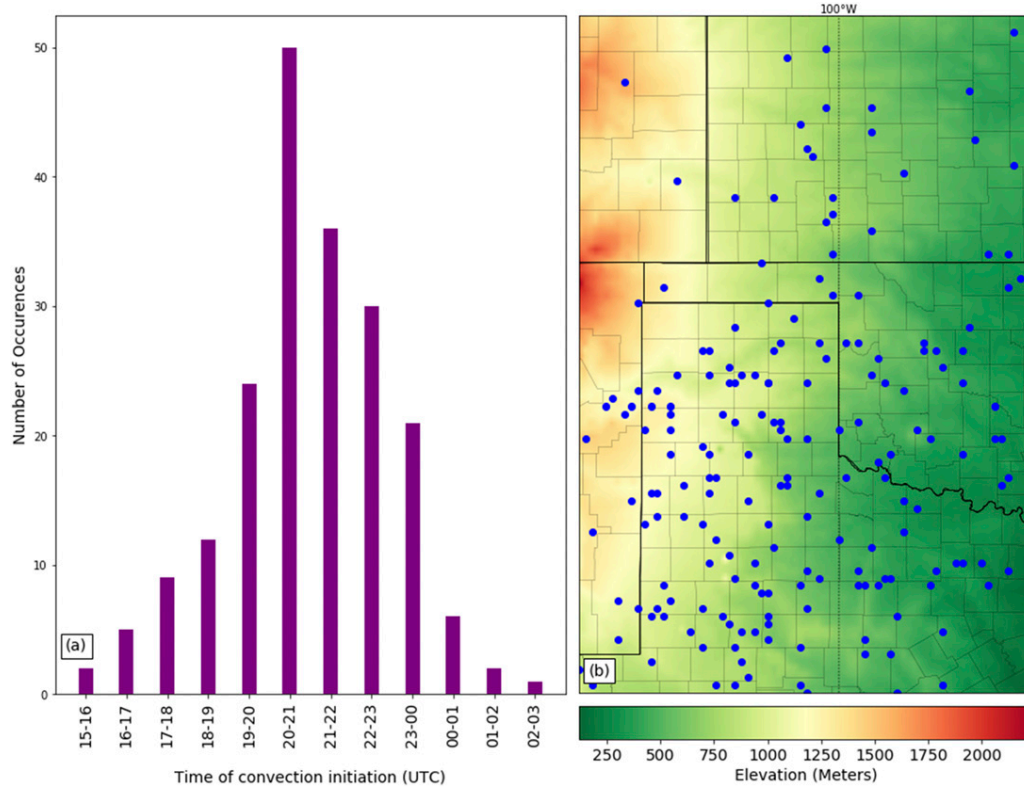
FIG. 5. (a) The frequency of convection initiation separated by the time at which it occurred. (b) The location of first occurrence of convection initiation for all convective dryline days (2006–15). Each blue circle represents an initiation location for a dryline day. Background shading represents topography.

north (Fig. 5b). Eastern Colorado and far western Kansas were rarely locations where storms first initiated. Overall, the first occurrence of initiation was more common in the southern half of the domain. Convection often initiated west of 100°W in Texas. Initiation in that region is consistent with Hoch and Markowski (2005) who found that the 0000 UTC dryline position had a peak frequency around 101°W. This longitude approximately coincides with the Caprock escarpment, which extends southward for around 200 miles from the eastern Texas panhandle. Across the escarpment, elevation changes by as much as 300 m. The elevation change may be a contributing factor as to why there is a peak in dryline frequency at that longitude. The dryline can be considered as the location where the western edge of the moist air mass intersects with the sloping terrain (Jones and Bannon 2002). Fewer initiations occur in Colorado and Kansas, consistent with synoptic experience that the dryline is more common in Texas than farther north (e.g., Schaefer 1986).

There are many questions that remain unanswered with regard to orography. For instance, do storms initiate earlier in certain locations than surrounding areas? Does convection preferentially occur in some locations,

but not others? However, answering these questions is beyond the scope of this study.

### b. Temporal clustering

Drylines often occurred in multiday sequences. The frequency of consecutive dryline days is shown in Fig. 6.
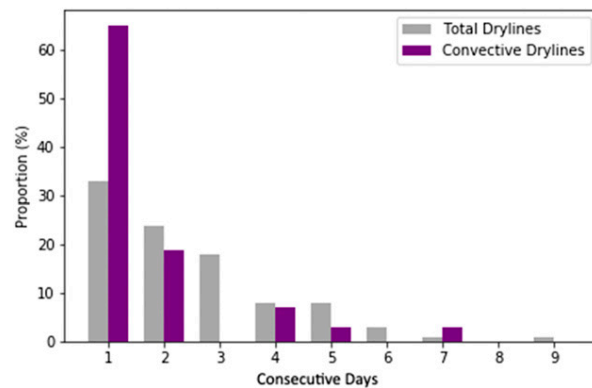


FIG. 6. The proportions of the total (gray) drylines and convective (purple) drylines that occurred for a given number of consecutive days (2006–15).
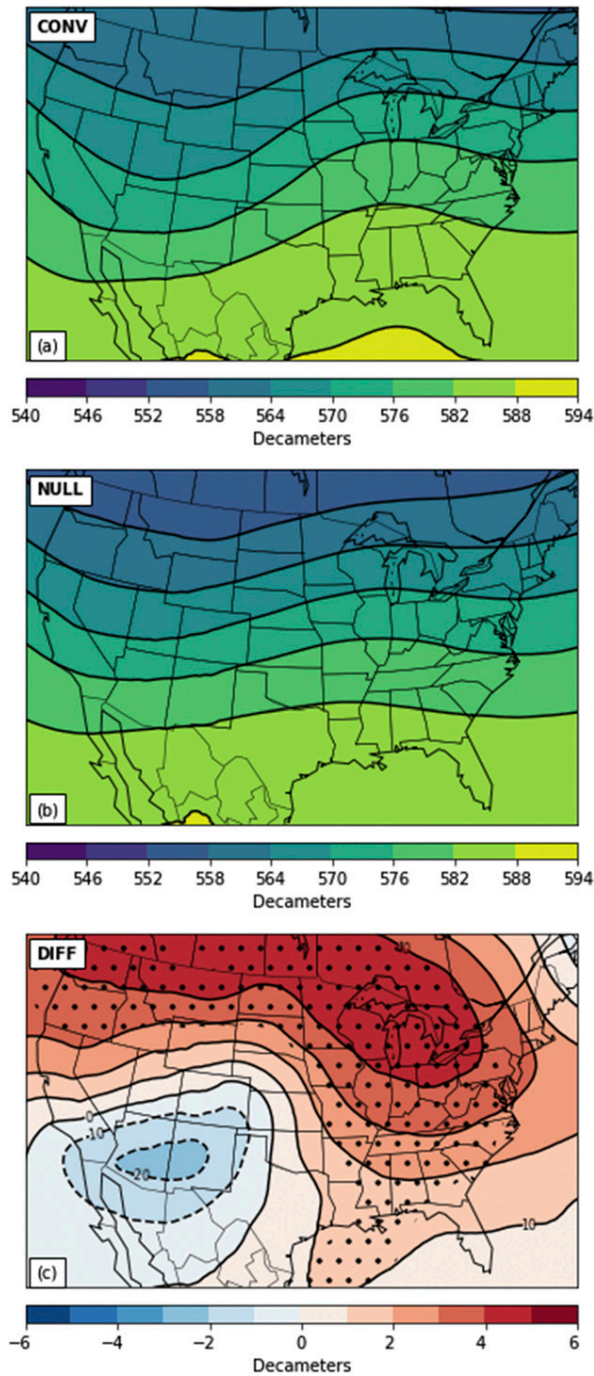
FIG. 7. 500-hPa geopotential height for (a) convective and (b) null days at 2100 UTC. (c) The difference between the two fields, calculated by subtracting the null from convective. Stippling indicates significant points controlled for a false discovery rate of 0.1.



FIG. 8. Mean sea level pressure for (a) convective and (b) null days at 2100 UTC. (c) The difference between the two fields, calculated by subtracting the null from convective. Lack of stippling indicates that no differences were significant using a false discovery rate of 0.1.

Over 66% of drylines occurred over two or more consecutive days, and drylines were identified as many as nine days in succession. Consecutive dryline days were typically due to a dryline persisting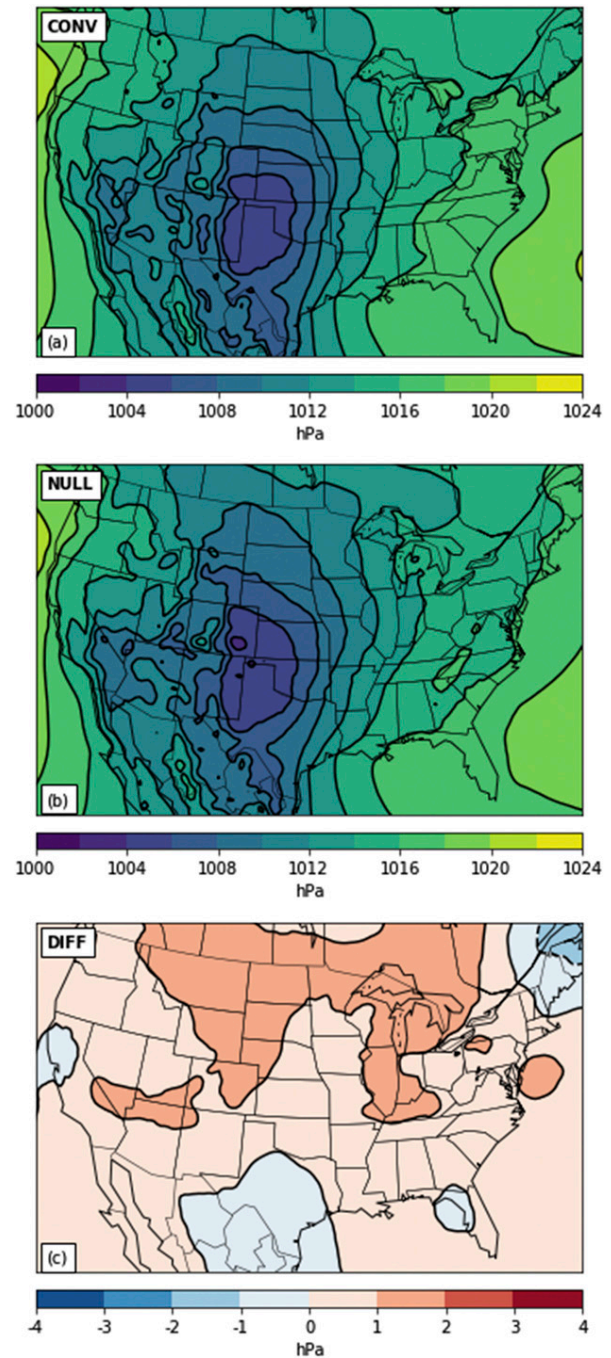 overnight, rather than regeneration of a new dryline each day. The occurrence of dryline days in temporal sequences has previously been studied by Schaefer (1974). He found that 22 distinct dryline events over 3 years resulted in 114 dryline days.
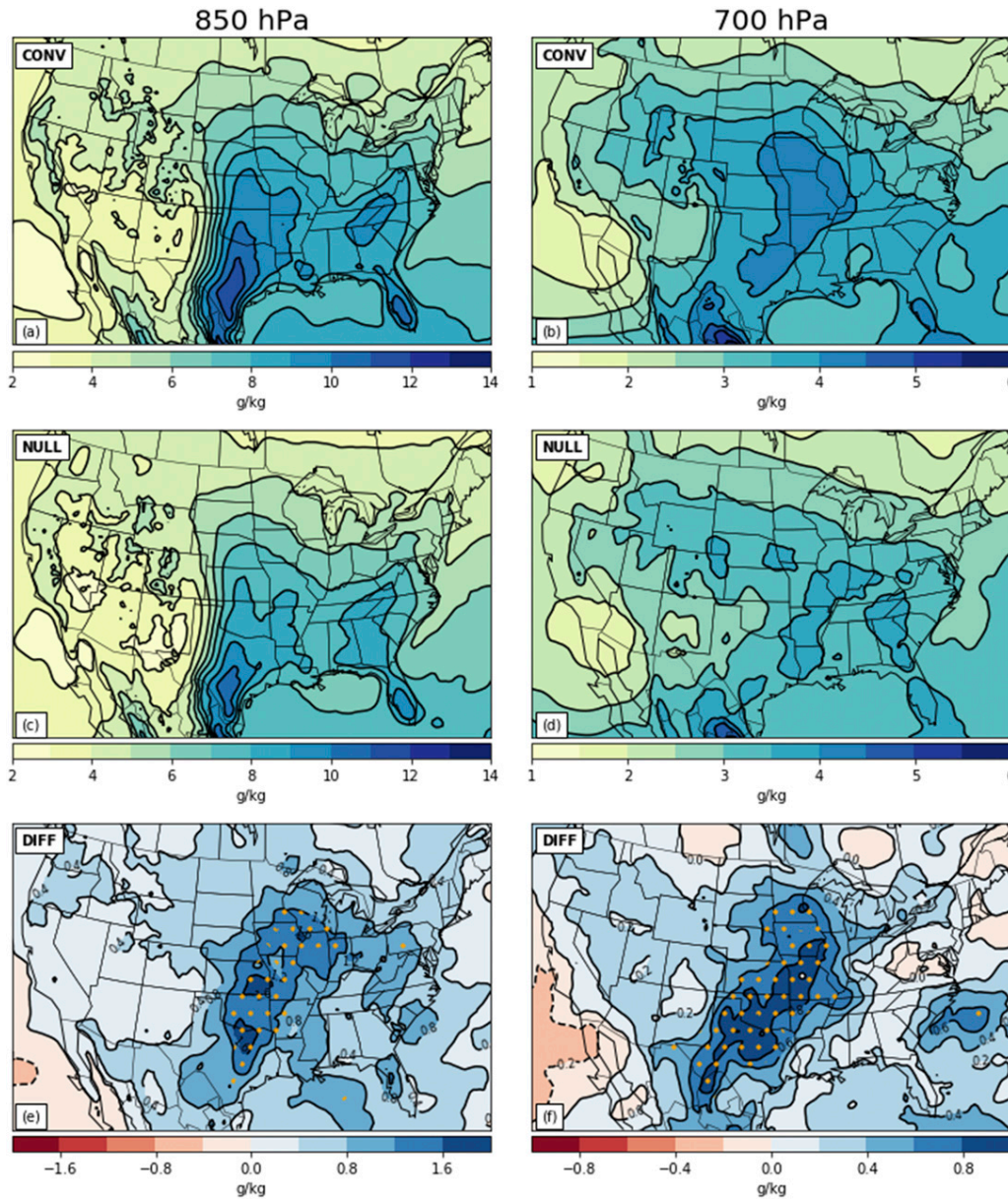
FIG. 9. (left) 850- and (right) 700-hPa specific humidity composite fields. (a),(b) The convective composites; (c),(d) the null; and (e),(f) the difference between the two. Stippling indicates significant points controlled for a false discovery rate of 0.1.

These results suggest that conditions conducive to dryline development often exist for more than one day in the Great Plains.

Convective drylines were less likely to occur in multiday sequences. Only 34% of convective drylines occurred in sequences of two or more consecutive days. On 52 of the 199 convective dryline days (26%), a convective dryline occurred on the second day of a dryline sequence or later. However, even if conditions were conducive to deep, moist convection on one day,

it does not mean that subsequent dryline days would produce deep, moist convection. On 43 occasions (22%), a convective dryline day was followed by a dryline day with no convection. Thus, the ingredients for deep, moist convection are either not consistently present following a convective dryline day, or present in insufficient quantity or magnitude. Are any ingredients missing on days that do not produce deep, moist convection? The following section will explore synoptic composites on days with and without
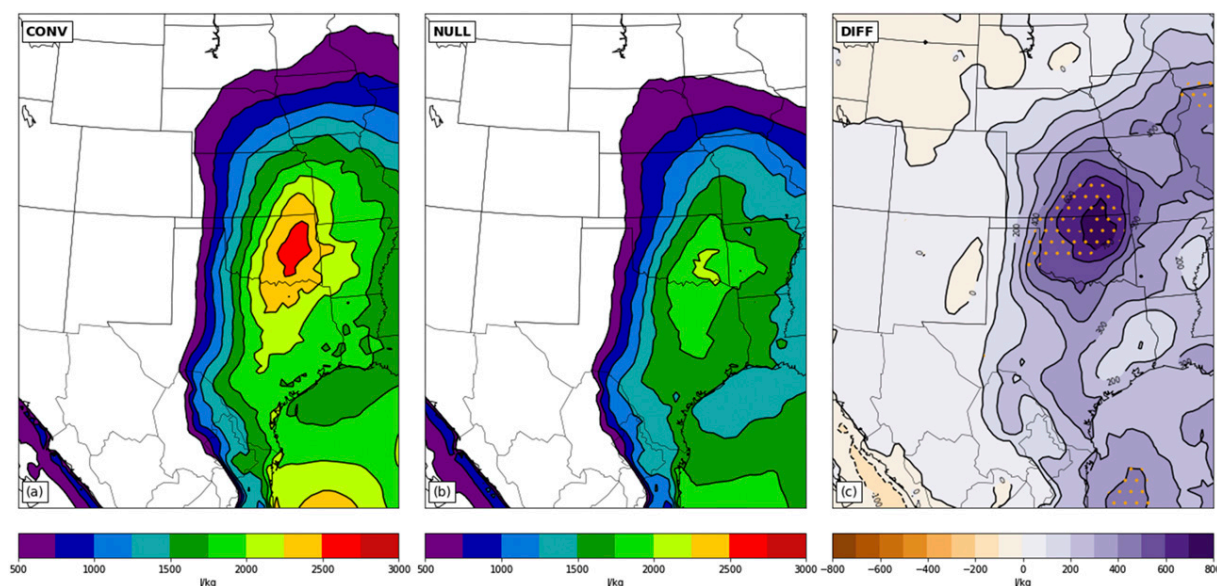
FIG. 10. Surface-based CAPE for (a) convective and (b) null days at 2100 UTC. (c) The difference between the two fields, calculated by subtracting the null from convective. Stippling indicates significant points controlled for a false discovery rate of 0.1.

dryline convection to identify any differences in these ingredients.

## 4. Synoptic composites of days with and without dryline convection

To examine differences between drylines that produce deep, moist convection and those that do not, synoptic composites were created using NCEP North American Regional Reanalysis (NARR) data. The NARR is available from 1979 and consists of 45 vertical levels of data at a 32-km horizontal grid spacing (Mesinger et al. 2006). Composites of hourly mean conditions were created from 2100 UTC data, chosen because dryline convection most commonly initiates between 2000 and 2100 UTC. Days where a dryline was not present at 2100 UTC were excluded. Two composites were created: convective and null. The convective composite consisted of drylines that satisfied the criteria for convective storms defined in section 2, whereas the null composite was created from all the remaining dryline cases. The convective composite contained 179 days, and the null composite contained 104.

Composite differences were calculated by subtracting the null composite from the convective composite. Statistical significance of the differences was initially calculated using a two-tailed t test applied to each grid point. Field significance was tested by applying the false discovery rate (FDR) method recommended by Wilks (2016). The FDR method accounts for random rejections

of local null hypotheses that occur when performing multiple hypothesis tests and is robust to spatial autocorrelation of underlying data. An acceptable proportion of incorrect rejections of local null hypotheses was chosen by controlling the FDR using the Benjamini–Hochberg procedure (Benjamini and Hochberg 1995). The procedure calculates a local threshold $p$ value for each grid point. Unless otherwise stated, field significance presented in this section is calculated using an FDR of 0.1. The remainder of this section will present a comparison of the synoptic-scale conditions in the convective and null composites, before focusing on the ingredients necessary for deep, moist convection.

### a. Synoptic overview

At 500 hPa, both the convective and null composites show a long-wave trough with an axis extending from north to south, west of the Rockies (Figs. 7a,b), while there is long-wave ridging in the eastern United States. However, the convective composite shows a more amplified pattern than the null. Figure 7c shows the difference between the two composite height fields, calculated by subtracting the null from the convective. The heights in the convective composite are lower over Arizona, New Mexico, and southern Colorado, but higher in the northeastern United States, suggestive of both a deeper trough and stronger ridge in the convective composite.

At the surface, there is a high over southern Florida and low pressure in the lee of the Rockies (Figs. 8a,b). In both composites, the low pressure area extends
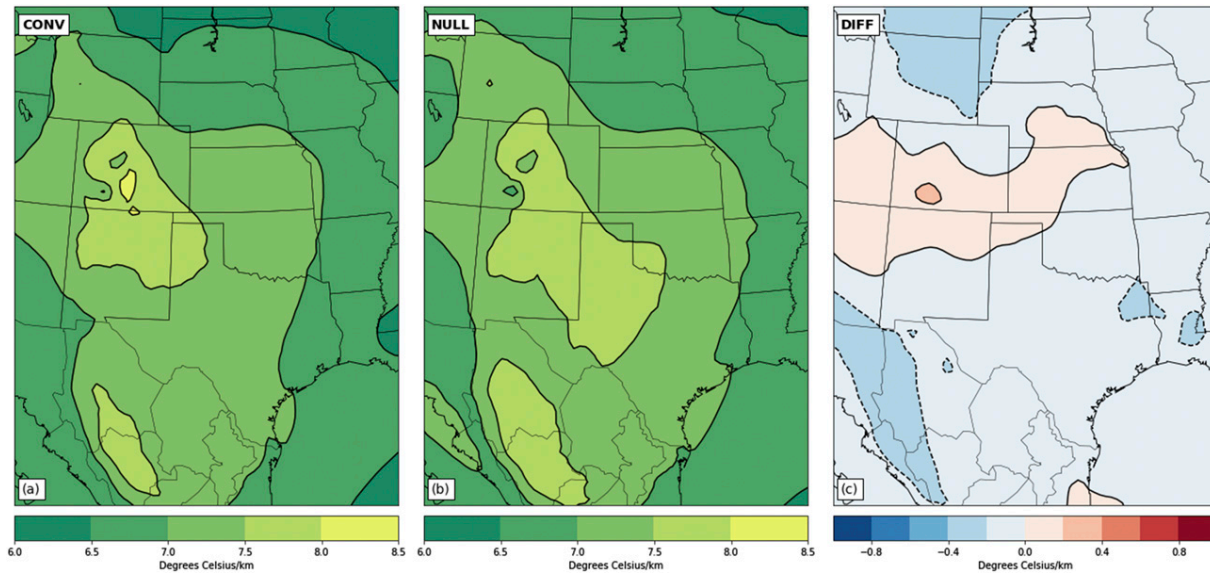
FIG. 11. 700–400-hPa lapse rate for (a) convective and (b) null days at 2100 UTC. (c) The difference between the two fields, calculated by subtracting the null from convective. Regions where the convective composite has steeper lapse rates are shown in red. Lack of stippling indicates that no differences were significant using a false discovery rate of 0.1.

through the Texas and Oklahoma panhandles, southwest Kansas, southeast Colorado, and eastern New Mexico. However, in the null composite, troughing extends farther north into Wyoming. This northward extension may be explained by the greater zonal component to winds aloft over the northern Rockies (Fig. 7), which would induce stronger lee troughing. Figure 8c shows

the effect of this troughing, with higher surface pressure in northern Wyoming and southern Montana in the convective composite.

### b. Moisture

The synoptic-scale conditions in the composites provide a favorable environment for dryline development
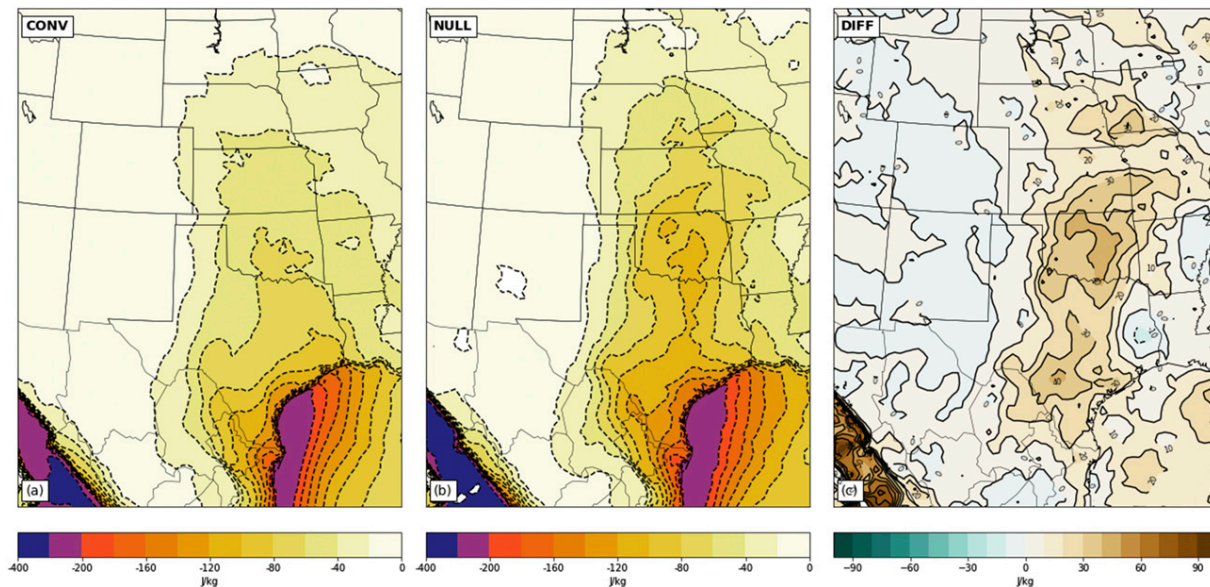


FIG. 12. Surface-based convective inhibition for (a) convective and (b) null days at 2100 UTC. (c) The difference between the two fields, calculated by subtracting the null from convective. Lack of stippling indicates that no differences were significant using a false discovery rate of 0.1.
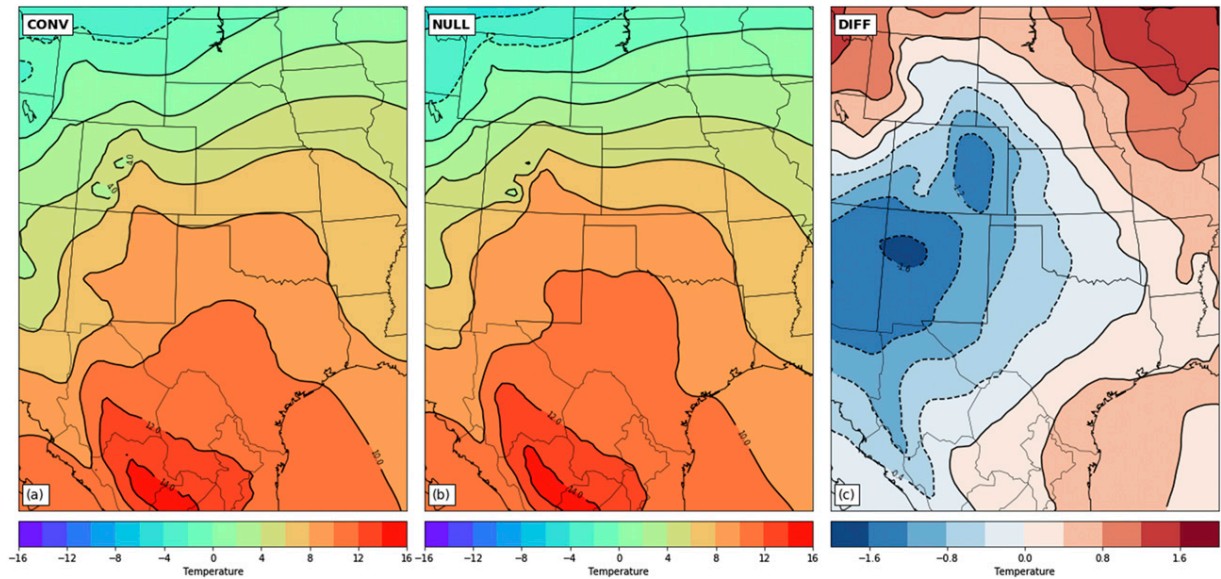
FIG. 13. As in Fig. 12, but for 700-hPa temperature.

and deep, moist convection. Lee troughing east of the Rockies results in large-scale confluence that facilitates the development of a dryline. The average dryline position can be approximated as the location of strongest west–east moisture gradient at 850 hPa (Figs. 9a,b). The strongest gradient is at the western edge of a tongue of moisture that extends northward from the Gulf of Mexico. In the convective composite, the moist air east of the dryline extends farther

north and is more moist than in the null composite. Differences of over $1 \, g \, kg^{-1}$ extend from northern Texas into the Midwest (Fig. 9c). These differences in moisture are not restricted to east of the dryline, however. In fact, the convective composite is more moist almost everywhere east of the Rockies.

At 700 hPa, the poleward-extending moist tongue is not as well defined, especially in the null composite (Figs. 9d,e). The highest specific humidity values in the
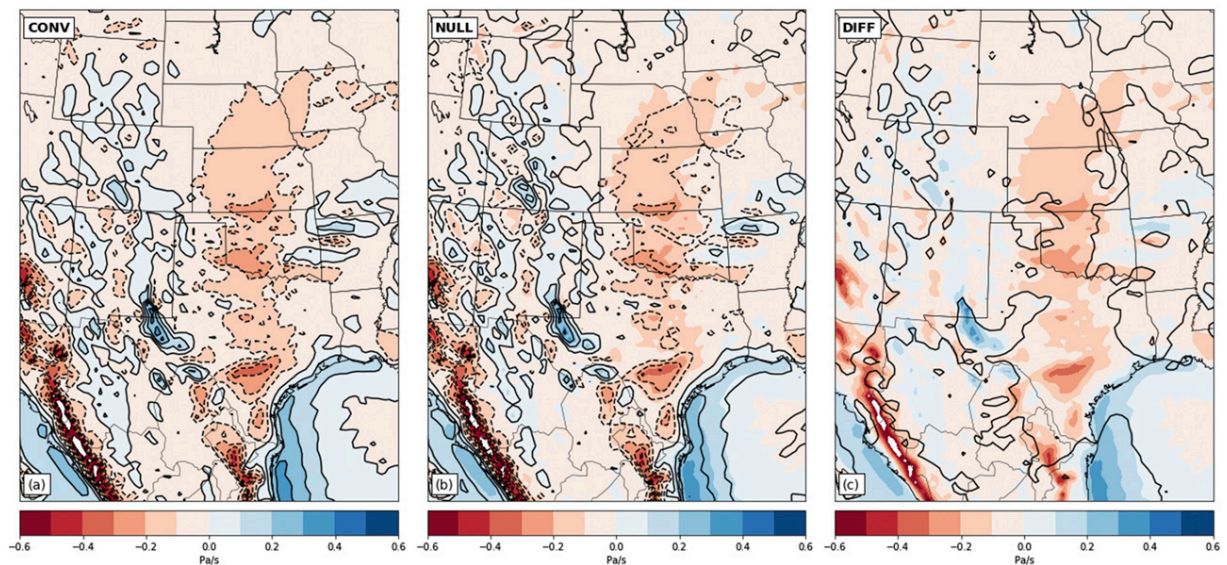


FIG. 14. Vertical velocity at 850 hPa for (a) convective and (b) null days at 2100 UTC. (c) The difference between the two fields, calculated by subtracting the null from convective. Regions where the convective composite has more ascent are shown in red. Lack of stippling indicates that no differences were significant using a false discovery rate of 0.1.
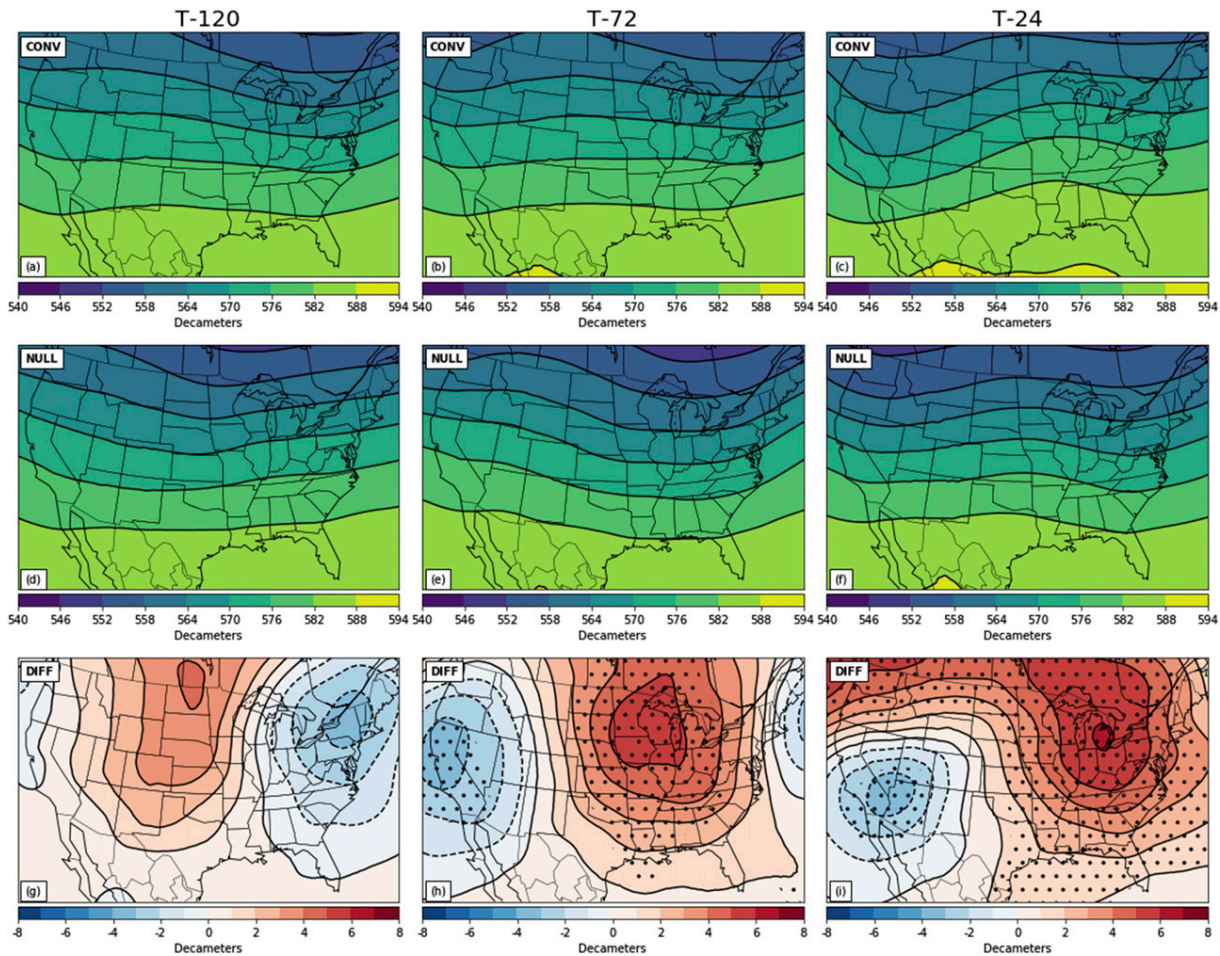
FIG. 15. 500-hPa geopotential height for (left) 120, (center) 72, and (right) 24 h before a dryline event. (a)–(c) The convective composite, (d)–(f) the null, and (g)–(i) the difference between the two. Stippling indicates significant points controlled for a false discovery rate of 0.1.

convective composite are observed in the central states, but the null composite is drier in this region and lacks an obvious region of higher values. As at 850 hPa, the 700-hPa convective composite is more moist than the null over a large swath extending from Texas into the Midwest (Fig. 9f). However, the region of significant difference between the composites extends farther south and west than at 850 hPa, and overspreads some of the surface dryline. All other things being equal, incipient convection would be less affected by the detrimental effects of dry-air entrainment.

*c. Convective instability*

As with the 850-hPa specific humidity field, the location of the dryline in both composites can be identified by the west–east gradient in surface-based CAPE (Figs. 10a,b). A region of high CAPE exists east of the dryline, extending north and east from the Gulf of Mexico. There is a maximum in CAPE in eastern

Oklahoma, where values exceed $2600 \, \mathrm{J \, kg^{-1}}$ in the convective composite and $2000 \, \mathrm{J \, kg^{-1}}$ in the null composite. Overall, the convective composite has higher CAPE values than the null composite over Texas, Oklahoma, Kansas, Nebraska, and all locations farther east (Fig. 10c). The differences in CAPE are most pronounced east of the dryline, where values differ by as much as $700 \, \mathrm{J \, kg^{-1}}$. The regions of largest difference in CAPE are somewhat collocated with the regions of largest difference in specific humidity, suggesting that the CAPE differences are caused by differences in low-level moisture.

The argument that the CAPE differences between the composites are driven by low-level moisture is supported by the 700–400-hPa lapse-rate difference shown in Fig. 11. The steepest lapse rates occur over southern Colorado, northern New Mexico and northwest Texas. However, areas where the convective composite has steeper lapse rates than the null are separated from the
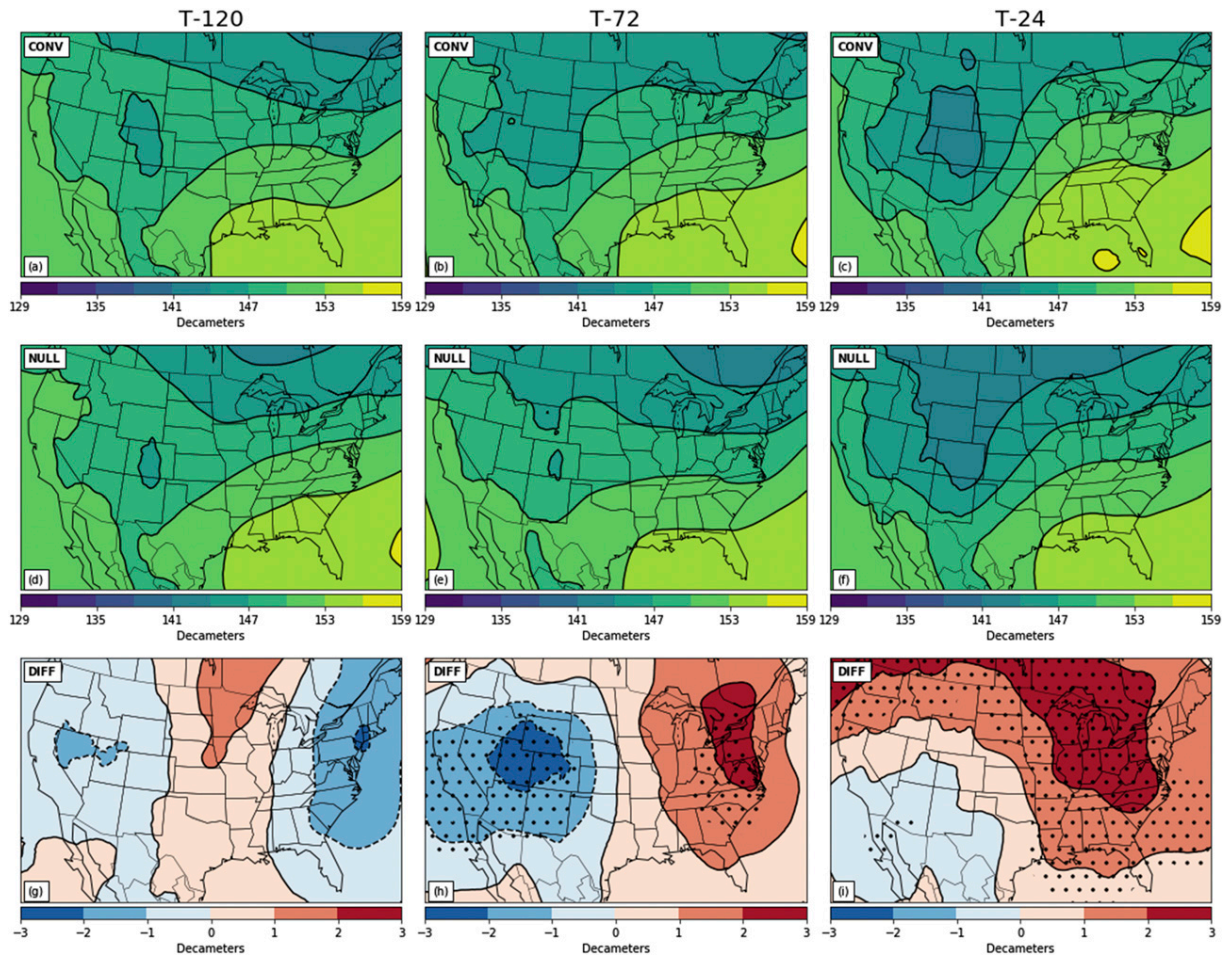
FIG. 16. As in Fig. 15, but for 850-hPa geopotential height.

region of higher CAPE values, implying that the differences in CAPE are primarily driven by moisture differences.

*d. Convective inhibition*

The presence of moisture and instability are necessary, but not sufficient, ingredients to initiate deep, moist convection as enough lift must be provided to overcome any existing convective inhibition. In both composites, inhibition is weak near the dryline and increases in an easterly direction toward central Texas and Oklahoma (Figs. 12a,b). The strongest inhibition is found over the Gulf of Mexico near the Texas coastline. The convective composite has inhibition values that rarely exceed $80 \, \mathrm{J \, kg^{-1}}$, with most of the dryline region exhibiting less than $60 \, \mathrm{J \, kg^{-1}}$, and has weaker inhibition than the null composite over a large area east of the dryline. The largest differences in inhibition occur in eastern Oklahoma where the difference widely exceeds $30 \, \mathrm{J \, kg^{-1}}$.

What effect do midlevel temperatures have on the magnitude of inhibition? At 700 hPa, temperatures are lower in the convective composite over Arizona, New Mexico, Colorado, and western regions of Kansas, Oklahoma, and Texas (Fig. 13). In fact, the convective composite is colder over almost all of the domain defined in section 2. However, as with convective inhibition, there are no areas where the differences are significant. Furthermore, the regions of weaker inhibition are somewhat separated from regions of lower 700-hPa temperatures. This separation is evident in the Midwest where the convective composite is warmer than the null (Fig. 13c), but inhibition is weaker (Fig. 12c). The differences in inhibition are primarily controlled by low-level moisture rather than midlevel temperature. Examination of 850-hPa specific humidity (Fig. 9e) reveals that regions of higher moisture are roughly collocated with regions of lower CIN (Fig. 12c), suggesting that moisture is
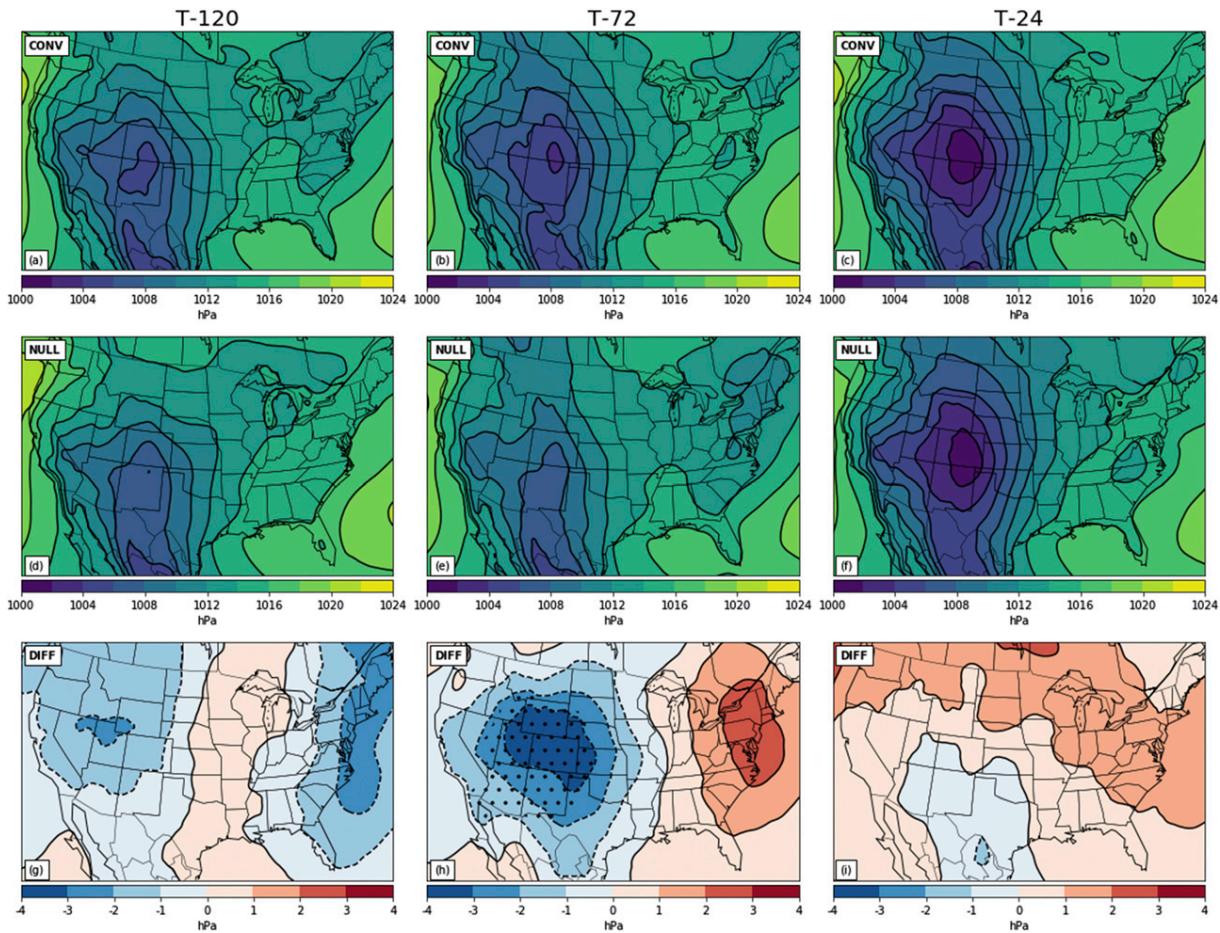
FIG. 17. As in Fig. 15, but for mean sea level pressure.

the primary factor driving differences in inhibition as well as CAPE.

### e. Lift

Weak inhibition will not always result in deep, moist convection along the dryline. The inhibition must be overcome by low-level convergence and ascent through the inversion, weakened by large-scale ascent and cooling, or a combination of the two. Even before the smoothing effect of compositing, the dryline circulation cannot be resolved by the NARR, which has a horizontal grid spacing of 32 km. However, vertical motion over larger spatial scales can be analyzed.

Ascent is observed in both the convective and null composite 850-hPa vertical velocity fields over a large area of the Great Plains (Figs. 14a,b). The strongest ascent over the Plains is located through central regions of Texas, Oklahoma, and Kansas. Larger ascent is observed in the convective composite than the null to the

immediate east of the dryline, with the difference widely exceeding $0.1 \, \text{Pa s}^{-1}$. However, these differences are not significant.

A similar pattern is also observed at higher altitudes (not shown), but the magnitude of the ascent over the Great Plains generally weakens with increasing altitude. The ascent in both the convective and null composites is likely associated with warm-air advection. However, there may be some contribution from ongoing convection given that the composites are created from 2100 UTC data.

## 5. Synoptic composites of days preceding a dryline

Many differences are observed between convective and nonconvective composites of 2100 UTC NARR data on the day of a dryline, but is it possible to distinguish between convective and nonconvective drylines at longer lead times? Applying the same criteria as in section 4, composites were created of
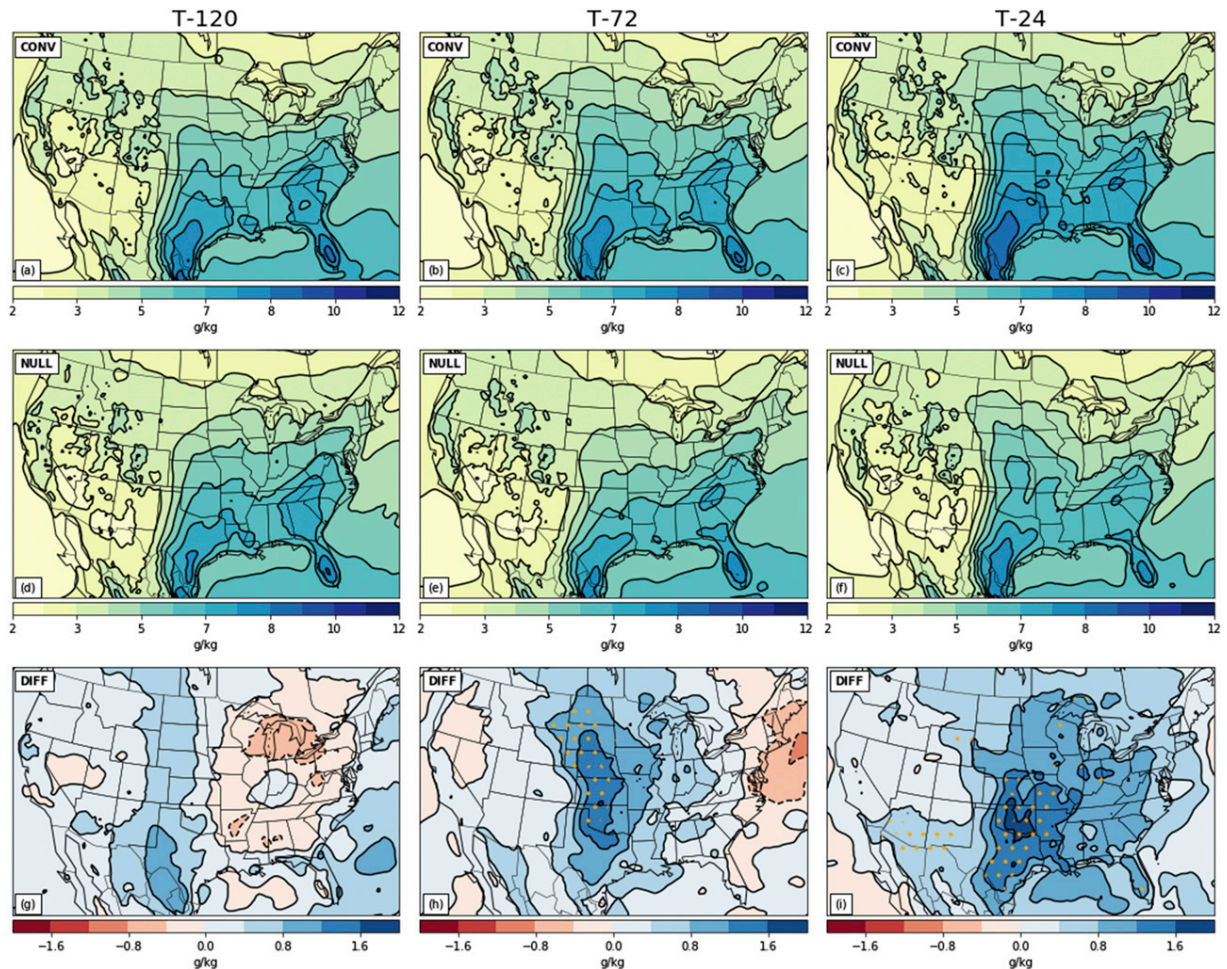
FIG. 18. As in Fig. 15, but for 850-hPa specific humidity.

conditions at 2100 UTC for 120, 72, and 24 h before a dryline day.

*a. Synoptic overview*

Figure 15 shows the difference in 500-hPa geopotential height, for 120, 72, and 24 h before a dryline day (T). In both the convective and null composites, heights rise in the upper Midwest and Great Lakes region as the lead time reduces, while heights fall over the west coast between T–72 and T–24. At T–72, significant differences can be observed between the composites. The convective composite features a more amplified west coast trough and higher heights in the east than the null. By T–24, the convective composite features a prominent trough west of the Rockies and a ridge in the Midwest. However, the null composite has a less amplified pattern. The trough over the west coast is more shallow, and the ridge is weaker and located slightly farther west than in the convective composite.

The building of a ridge in the east in combination with stronger troughing in the west is also seen at lower altitudes (Fig. 16). At T–120, both 850-hPa composites have a height minimum in Colorado and higher heights in the southeast (Figs. 16a,d). At T–96 (not shown) and T–72, the null composite also features a height minimum over Colorado, but the convective composite has significantly lower heights over the central Rockies in addition to higher heights in the northeastern United States than the null. By T–24, the convective composite has higher heights over the entire eastern United States, with the largest differences occurring over the upper Midwest. However, the differences in the strength of the western trough have all but disappeared by T–24.

A similar trend is observed in mean sea level pressure as the differences between the composites are maximized at T–72 and reduced by T–24 (Fig. 17). At T–120, both composites feature lower pressure in New Mexico and southern Colorado, and higher pressure in the
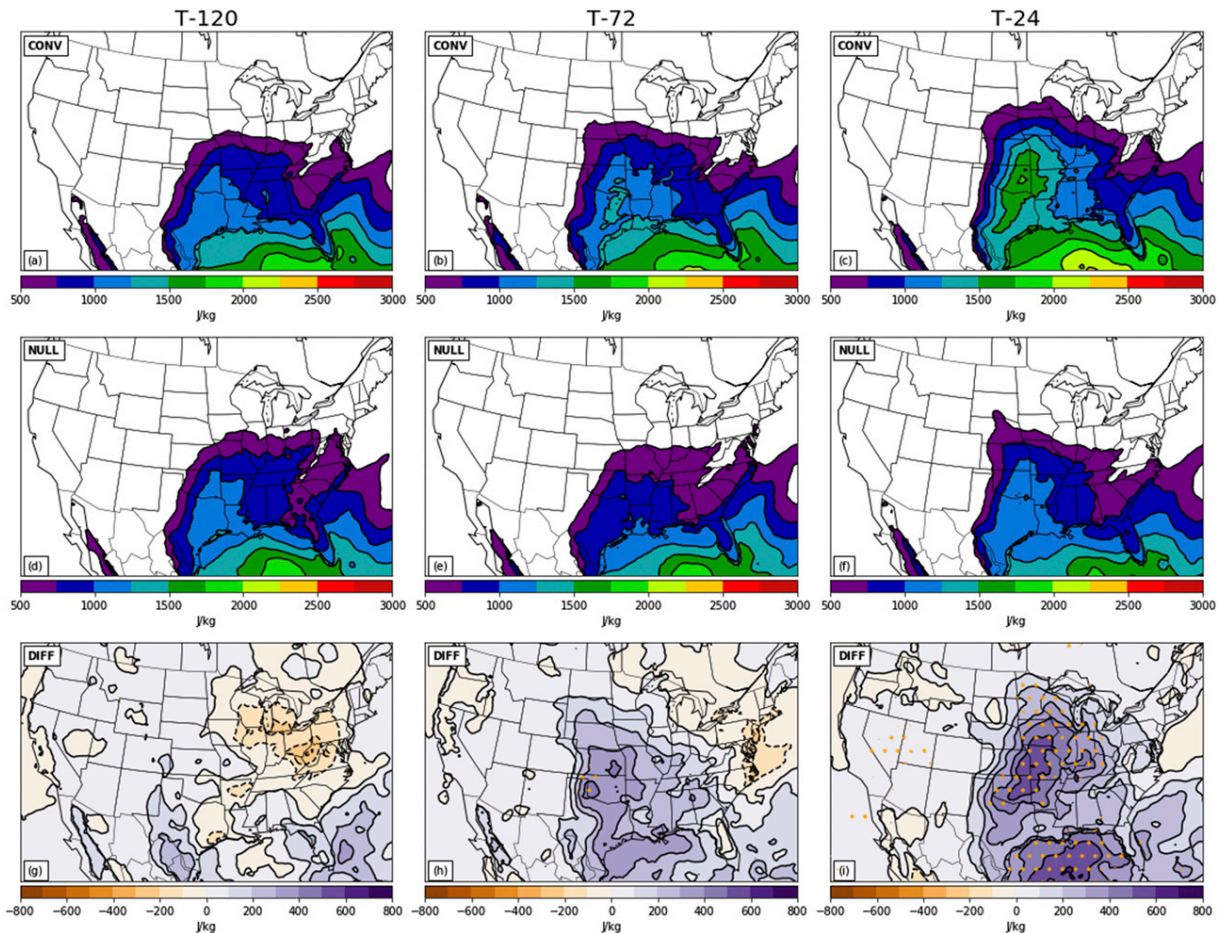
FIG. 19. As in Fig. 15, but for surface-based CAPE.

southeast with southerly flow into the Great Plains. By T–72, the convective composite has a deeper low pressure region than at T–120, but the low pressure region in the null composite is relatively unchanged. As a result, the mean sea level pressure in the convective composite is over 3 hPa lower over Colorado and parts of adjacent states. At T–24, both composites have a deeper low pressure area than at T–72; however, the differences between the composites have all but disappeared.

### b. Moisture

The presence of significant differences in mean sea level pressure at T–72, but none on the day of a dryline, suggests that the large-scale pattern in the three or four days before a dryline event can help create the differences observed on a dryline day. The composites at T–72 feature a large-scale pattern that is conducive to moisture advection into the Great Plains from the Gulf of Mexico. The convective composites feature a stronger trough, likely associated with greater moisture advection than the null.

Composites of 850-hPa specific humidity (Fig. 18) confirm that convective cases are more moist in the days preceding a dryline event. For all the times shown, the highest specific humidity values are found in the south and southeastern United States, with drier air in the southwest. At T–120 h, the convective composite is more moist over the western Plains. By T–72, a tongue of moisture extends farther north in the convective composite and the convective composite is more moist over the entire Great Plains region. This difference is significant over much of Oklahoma and Kansas, and significant differences extend as far north as North Dakota. By T–24, the convective composite has specific humidity values over $1.6 \, \mathrm{g \, kg^{-1}}$ higher than the null composite in eastern Oklahoma.

### c. Convective instability

A result of this difference in specific humidity is that the convective composite has greater surface-based CAPE values in the Plains from T–72 onward (Fig. 19).
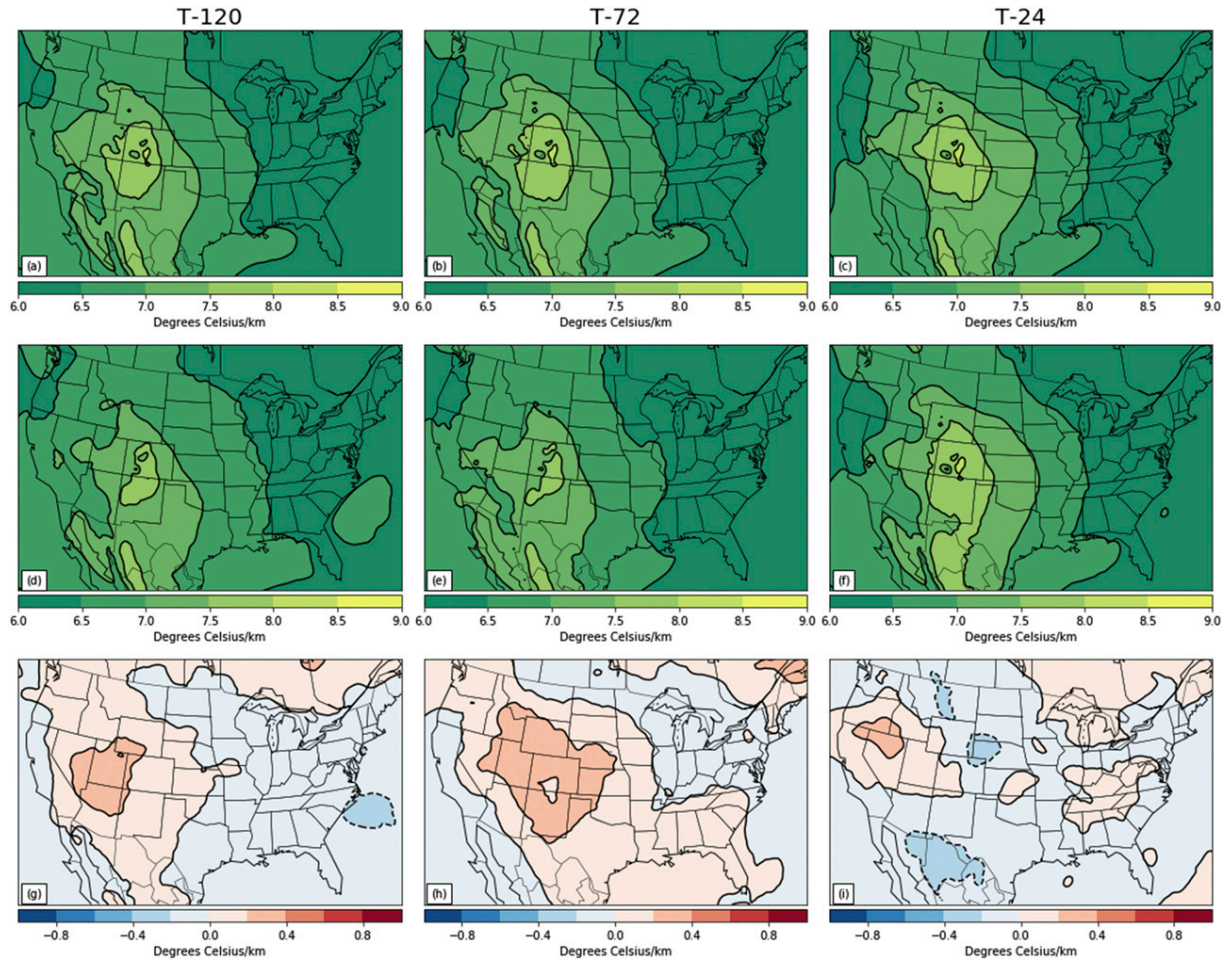
FIG. 20. 700–400-hPa lapse rate for (left) 120, (center) 72, and (right) 24 h before a dryline event. (a)–(c) The convective composite, (d)–(f) the null, and (g)–(i) the difference between the two. Regions where the convective composite has steeper lapse rates are shown in red. Lack of stippling indicates that no differences were significant using a false discovery rate of 0.1.

At T–72, the convective composite has CAPE values over $300 \, \mathrm{J \, kg^{-1}}$ greater than the null over most of Oklahoma and Kansas. By T–24, the convective composite has larger CAPE values over a large swath extending from Texas into the Midwest. The differences in CAPE are greatest over eastern Oklahoma and Kansas, and western Missouri.

The steepest lapse rates (exceeding $7.5°\mathrm{C \, km^{-1}}$) exist over Colorado and New Mexico (Fig. 20), appearing to be collocated with the 850-hPa trough (Fig. 16), while the $7°\mathrm{C \, km^{-1}}$ isopleth extends eastward to central Oklahoma. By T–24, the region of lapse rates greater than $7°\mathrm{C \, km^{-1}}$ extends farther east, reaching the Kansas–Missouri border. The steepening lapse rates are likely explained by the arrival of cool air aloft due to the approaching trough. However, the increase in lapse rates farther east as the lead time shortens may be also be an indicator of advected elevated mixed-layer air.

The argument that the increase in lapse rates is associated with the elevated mixed layer is supported by an increase in convective inhibition (not shown) between T–72 and T–24. Convective inhibition increases over almost all of Texas, Oklahoma, and Kansas between those times. However, even if the elevated mixed layer is responsible for the increase in convective inhibition, the characteristics of the elevated mixed layer do not appear to be important in distinguishing between convective and nonconvective drylines. No significant differences are observed between the convective and null composites of 700–400-hPa lapse rate (Figs. 20g–i).

### d. Lift

There are also no significant differences between the convective and null composites of vertical velocity in the days preceding a dryline event (Fig. 21). Composites of
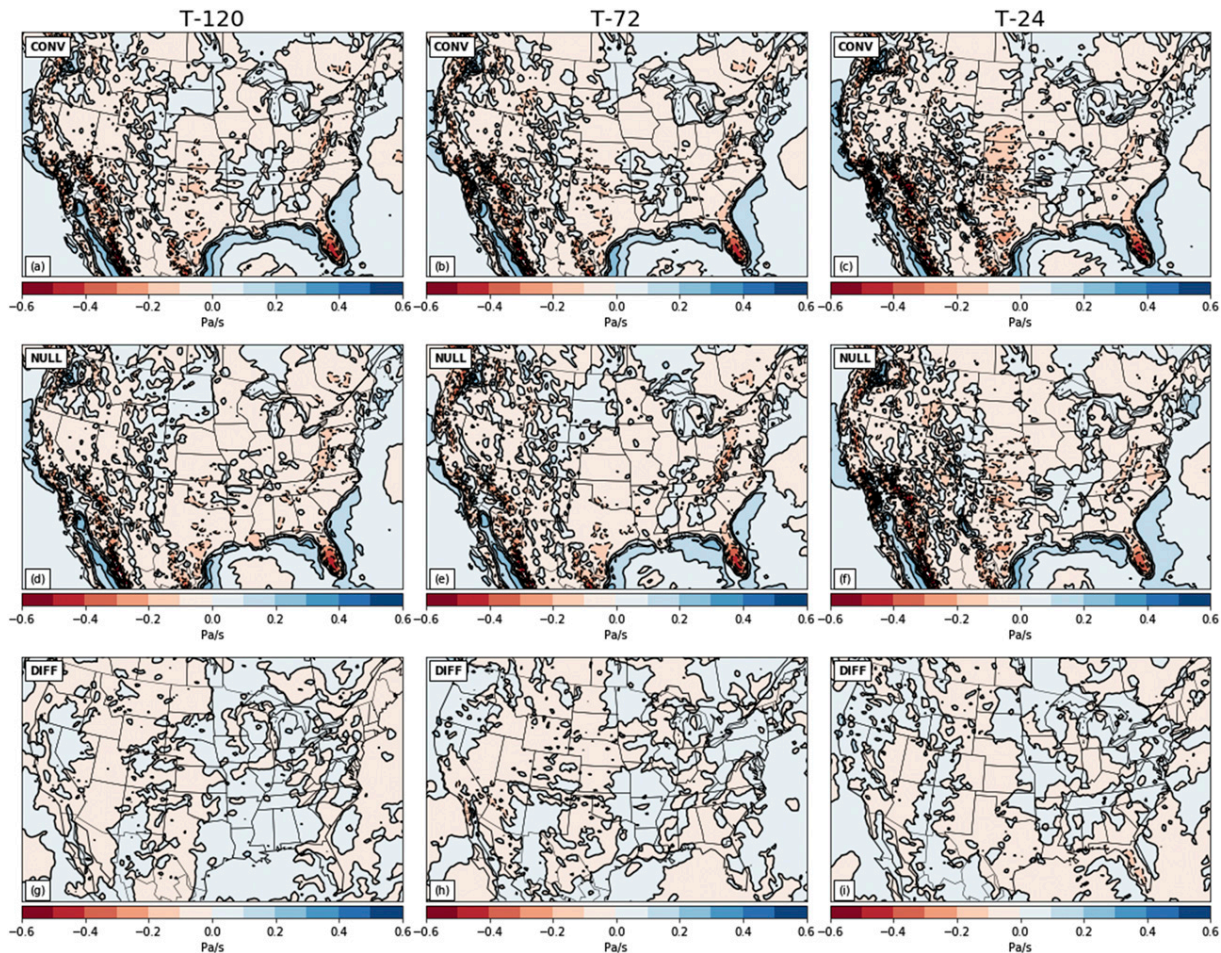
FIG. 21. 850-hPa vertical velocity for (left) 120, (center) 72, and (right) 24 h before a dryline event. (a)–(c) The convective composite, (d)–(f) the null, and (g)–(i) the difference between the two. Regions where the convective composite has more ascent are shown in red. Note the different scale used in (g)–(i). Lack of stippling indicates that no differences were significant using a false discovery rate of 0.1.

850-hPa vertical velocity feature weak ascent ($\sim$0.1 Pa s$^{-1}$) over most of the southern Great Plains. The magnitude of this ascent is greatest at T–24 with values exceeding 0.1 Pa s$^{-1}$ over much of Oklahoma and Kansas. Although the magnitude of ascent is slightly greater ($<$0.1 Pa s$^{-1}$) in the convective composite at T–24, this difference is not significant.

Ascent is also present over much of the southern Plains at higher altitudes (not shown), but the magnitude of ascent weakens with height. The strongest ascent is at low levels, and the southerly winds over the southern Plains in the mean sea level pressure composites (Fig. 17) suggest the ascent in both composites is associated with warm-air advection. In the midlevels (not shown), there is no clear signal for vertical motion in the Plains, even at short lead-times. Although it appears that the large-scale pattern is conducive to warm-air advection and ascent in the low-levels, the large-scale influence of vertical motion at higher altitudes is unclear.

## 6. Summary

A dataset of drylines within a region of the southern Great Plains was constructed to investigate the importance of large-scale processes in the initiation of deep moist convection. Drylines were identified using WPC surface analyses, then radar and satellite imagery were used to establish whether deep, moist convection initiated along the dryline. Over the 10 years examined in this study, approximately 60% of drylines produced deep, moist convection, with convection most frequently initiating between 2000 and 2100 UTC. Convective drylines were most common in May.

Synoptic composites were created in an attempt to identify differences between convective and nonconvective

drylines. Analysis of the composites reveals that the large-scale environment appears to not only facilitate the development of a dryline, but also create conditions conducive to deep, moist convection. Three days before a dryline event, convective drylines feature a stronger 500-hPa ridge in the eastern United States and a deeper west coast trough. Meanwhile, convective drylines feature a deeper surface low over the central Rockies. As a result of greater poleward advection of low-level moisture into the Great Plains, the convective composite is more moist over large areas east of the Rockies.

On the day of a dryline, the synoptic composites feature a long-wave trough over the western Rockies and a ridge in the east. However, convective drylines have more amplified flow, a deeper trough, and a stronger downstream ridge. At the surface, both composites feature low pressure over the Texas panhandle region with associated poleward moisture advection to the east of the low. However, convective drylines have more abundant low-level moisture over the Plains, which results in greater values of CAPE east of the dryline.

Although CAPE differences appear to be primarily caused by differences in low-level moisture, the influence of the elevated mixed layer in creating greater CAPE values in convective cases is unclear. No significant differences were found in 700–400-hPa lapse rates or midlevel temperatures between the composites. Although convective inhibition was weaker in the convective composite east of the dryline, this result was also insignificant. It is possible that convection initiation is affected by variations in convective inhibition on a smaller spatial scale than the NARR can resolve.

Of course, composites cannot help answer all of the questions regarding dryline convection initiation. Whether or not the magnitude of convergence, and hence ascent, along the dryline can help distinguish between convective and nonconvective drylines remains unclear. Drylines vary in location and orientation within the Plains, and convergence is often confined to a narrow band along or ahead of the dryline. Future work could benefit from accounting for dryline location, orientation, and strength to help establish the importance of both convergence, any frontal circulation, and the strength of the moisture gradient in determining whether convection will initiate.

Despite these unanswered questions about finescale dryline variations, we have shown that the large-scale pattern is not only important in creating conditions conducive to dryline development, but may also help determine whether or not deep, moist convection initiates along the dryline. Convective drylines are associated with greater low-level moisture and higher values of surface-based CAPE than nonconvective drylines, a result of greater moisture advection into the Plains in the preceding days. Our results indicate that synoptic-scale processes may be important in determining whether dryline convection will occur.

Data availability statement. The WPC and NARR datasets are publicly available (https://www.wpc.ncep.noaa.gov/archives/web_pages/sfc/sfc_archive.php and https://psl.noaa.gov/data/gridded/data.narr.html, respectively).

REFERENCES

Banta, R. M., 1984: Daytime boundary-layer evolution over mountainous terrain. Part I: Observations of the dry circulations. *Mon. Wea. Rev.*, **112**, 340–356, https://doi.org/10.1175/1520-0493(1984)112<0340:DBLEOM>2.0.CO;2.

Beebe, R. G., 1958: An instability line development as observed by the tornado research airplane. *J. Meteor.*, **15**, 278–282, https://doi.org/10.1175/1520-0469(1958)015<0278:AILDAO>2.0.CO;2.

Benjamini, Y., and Y. Hochberg, 1995: Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B*, **57**, 289–300, https://doi.org/10.1111/J.2517-6161.1995.TB02031.X.

Berg, R., J. Clark, D. Roth, and T. Birchard, 2007: The National Weather Service Unified Surface Analysis. *22nd Conf. on Weather Analysis and 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.

Cai, H., W. C. Lee, T. M. Weckwerth, C. Flamant, and H. V. Murphey, 2006: Observations of the 11 June dryline during IHOP_2002—A null case for convection initiation. *Mon. Wea. Rev.*, **134**, 336–354, https://doi.org/10.1175/MWR2998.1.

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.

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.

Crook, N. A., 1996: Sensitivity of moist convection forced by boundary layer processes to low-level thermodynamic fields. *Mon. Wea. Rev.*, **124**, 1767–1785, https://doi.org/10.1175/1520-0493(1996)124<1767:SOMCFB>2.0.CO;2.

Demoz, B., and Coauthors, 2006: The dryline on 22 May 2002 during IHOP_2002: Convective-scale measurements at the

profiling site. *Mon. Wea. Rev.*, **134**, 294–310, https://doi.org/10.1175/MWR3054.1.

Doswell, C. A., 1987: The distinction between large-scale and mesoscale contribution to severe convection: A case study example. *Wea. Forecasting*, **2**, 3–16, https://doi.org/10.1175/1520-0434(1987)002<0003:TDBLSA>2.0.CO;2.

Fowle, M. A., and P. J. Roebber, 2003: Short-range (0–48 h) numerical prediction of convective occurrence, mode, and location. *Wea. Forecasting*, **18**, 782–794, https://doi.org/10.1175/1520-0434(2003)018<0782:SHNPOC>2.0.CO;2.

Fujita, T. 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.

Geerts, B., and Q. Miao, 2005: The use of millimeter Doppler radar echoes to estimate vertical air velocities in the fair-weather convective boundary layer. *J. Atmos. Oceanic Technol.*, **22**, 225–246, https://doi.org/10.1175/JTECH1699.1.

Grams, J. S., R. L. Thompson, D. V. Snively, J. A. Prentice, G. M. Hodges, and L. J. Reames, 2012: A climatology and comparison of parameters for significant tornado events in the United States. *Wea. Forecasting*, **27**, 106–123, https://doi.org/10.1175/WAF-D-11-00008.1.

Hane, C. E., 2004: Quiescent and synoptically-active drylines: A comparison based upon case studies. *Meteor. Atmos. Phys.*, **86**, 195–211, https://doi.org/10.1007/s00703-003-0026-y.

Hill, A. J., C. C. Weiss, and B. C. Ancell, 2016: Ensemble sensitivity analysis for mesoscale forecasts of dryline convection initiation. *Mon. Wea. Rev.*, **144**, 4161–4182, https://doi.org/10.1175/MWR-D-15-0338.1.

Hoch, J., and P. Markowski, 2005: A climatology of springtime dryline position in the US Great Plains region. *J. Climate*, **18**, 2132–2137, https://doi.org/10.1175/JCLI3392.1.

Hocker, J. E., and J. B. Basara, 2008: A geographic information systems–based analysis of supercells across Oklahoma from 1994 to 2003. *J. Appl. Meteor. Climatol.*, **47**, 1518–1538, https://doi.org/10.1175/2007JAMC1673.1.

Houston, A. L., and D. Niyogi, 2007: The sensitivity of convective initiation to the lapse rate of the active cloud-bearing layer. *Mon. Wea. Rev.*, **135**, 3013–3032, https://doi.org/10.1175/MWR3449.1.

James, R. P., J. M. Fritsch, and P. M. Markowski, 2005: Environmental distinctions between cellular and slabular convective lines. *Mon. Wea. Rev.*, **133**, 2669–2691, https://doi.org/10.1175/MWR3002.1.

Jones, P. A., and P. R. Bannon, 2002: A mixed-layer model of the diurnal dryline. *J. Atmos. Sci.*, **59**, 2582–2593, https://doi.org/10.1175/1520-0469(2002)059<2582:AMLMOT>2.0.CO;2.

Kain, J. S., and Coauthors, 2013: A feasibility study for probabilistic convection initiation forecasts based on explicit numerical guidance. *Bull. Amer. Meteor. Soc.*, **94**, 1213–1225, https://doi.org/10.1175/BAMS-D-11-00264.1.

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.

McGuire, E., 1962: The vertical structure of three dry lines as revealed by aircraft traverses. Tech. Rep., National Severe Storms Project Rep. 7, 11 pp.

Mesinger, F., and Coauthors, 2006: North American Regional Reanalysis. *Bull. Amer. Meteor. Soc.*, **87**, 343–360, https://doi.org/10.1175/BAMS-87-3-343.

Miller, R. C., 1959: Tornado-producing synoptic patterns. *Bull. Amer. Meteor. Soc.*, **40**, 465–472, https://doi.org/10.1175/1520-0477-40.9.465.

Mueller, C. K., J. W. Wilson, and N. A. Crook, 1993: The utility of sounding and mesonet data to nowcast thunderstorm initiation. *Wea. Forecasting*, **8**, 132–146, https://doi.org/10.1175/1520-0434(1993)008<0132:TUOSAM>2.0.CO;2.

NCEI, 2016: Interactive map tool—National Centers for Environmental Information (NCEI) formerly known as National Climatic Data Center (NCDC). Accessed 13 June 2016, https://www.ncdc.noaa.gov/data-access/radar-data/radar-map-tool.

Nielsen, E. R., R. S. Schumacher, and A. M. Keclik, 2016: The effect of the Balcones Escarpment on three cases of extreme precipitation in central Texas. *Mon. Wea. Rev.*, **144**, 119–138, https://doi.org/10.1175/MWR-D-15-0156.1.

NOAA, 2013: Unified Surface Analysis manual. NOAA Rep., 33 pp., http://www.wpc.ncep.noaa.gov/sfc/UASfcManualVersion1.pdf.

——, 2016: NOAA's National Weather Service—Glossary. Accessed 13 June 2016, https://w1.weather.gov/glossary/.

Parker, M. D., and R. H. Johnson, 2000: Organizational modes of midlatitude mesoscale convective systems. *Mon. Wea. Rev.*, **128**, 3413–3436, https://doi.org/10.1175/1520-0493(2001)129<3413:OMOMMC>2.0.CO;2.

Peterson, R. E., 1983: The west Texas dryline: Occurrence and behavior. Preprints, *13th Conf. on Severe Local Storms*, Tulsa, OK, Amer. Meteor. Soc., J9–J11.

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.

Richter, H., and L. F. Bosart, 2002: The suppression of deep moist convection near the southern Great Plains dryline. *Mon. Wea. Rev.*, **130**, 1665–1691, https://doi.org/10.1175/1520-0493(2002)130<1665:TSODMC>2.0.CO;2.

Russell, R. W., and J. W. Wilson, 1997: Radar-observed "fine lines" in the optically clear boundary layer: Reflectivity contributions from aerial plankton and its predators. *Bound.-Layer Meteor.*, **82**, 235–262, https://doi.org/10.1023/A:1000237431851.

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

——, 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.

——, 1986: The dryline. *Mesoscale Meteorology and Forecasting*, P. S. Ray, Ed., Amer. Meteor. Soc., 549–572.

Schultz, D. M., 2009: *Eloquent Science: A Practical Guide to Becoming a Better Writer, Speaker, and Atmospheric Scientist*. Amer. Meteor. Soc., 440 pp., http://www.eloquentscience.com./.

——, C. C. Weiss, and P. M. Hoffman, 2007: The synoptic regulation of dryline intensity. *Mon. Wea. Rev.*, **135**, 1699–1709, https://doi.org/10.1175/MWR3376.1.

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.

Trapp, R. J., S. A. Tessendorf, E. S. Godfrey, and H. E. Brooks, 2005: Tornadoes from squall lines and bow echoes. Part I: Climatological distribution. *Wea. Forecasting*, **20**, 23–34, https://doi.org/10.1175/WAF-835.1.

Weckwerth, T. M., and D. B. Parsons, 2006: A review of convection initiation and motivation for IHOP_2002. *Mon. Wea. Rev.*, **134**, 5–22, https://doi.org/10.1175/MWR3067.1.

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.

Wilks, D. S., 2016: "The stippling shows statistically significant grid points": How research results are routinely overstated and overinterpreted, and what to do about it. *Bull. Amer. Meteor. Soc.*, **97**, 2263–2273, https://doi.org/10.1175/BAMS-D-15-00267.1.

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.

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.

——, 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.

——, T. J. Lee, and R. A. Pielke, 1997: Convective initiation at the dryline: A modeling study. *Mon. Wea. Rev.*, **125**, 1001–1026, https://doi.org/10.1175/1520-0493(1997)125<1001:CIATDA>2.0.CO;2.

# Chapter 3

# Objective Analysis of the Dryline Environment

## 3.1 Overview

Chapter 2 identified the large scale conditions conducive to dryline convection primarily by the creation of synoptic composites of variables understood to be representative of the ingredients of deep, moist convection. The selection of variables for compositing was heavily influenced by previous literature recommending an ingredients-based approach to forecasting deep, moist convection. Moisture, instability and lift are necessary, but not necessarily sufficient for deep, moist convection to occur (Doswell 1987).

However, which variables best represent the aforementioned ingredients is open to question. When nowcasting, tools such as surface observations, atmospheric soundings and satellite imagery at different wavelengths can be useful in establishing whether the ingredients for deep, moist convection are present. When forecasting at longer lead times, forecasters are forced to make more use of numerical weather prediction. But which of the variables output by models are most useful? Some models can output vast amounts of data. For instance, the most recent version of the Rapid Refresh model (RAP) (Benjamin et al. 2016) outputs nearly 300 variables, many of which are either directly or indirectly related to moisture, instability and lift. Does a hierarchy of variable importance exist?

Given the number of variables available at multiple vertical levels in North American Regional Reanalysis (NARR) (Mesinger et al. 2006) used in Chapter 2, the composites were unable to be comprehensive. It is possible that variables that were not examined also differ between convective and non-convective drylines. Can the variables be examined more objectively?

Many dryline studies have focused on the near-dryline environment (e.g. Ziegler and Hane 1993; Atkins et al. 1998; Buban et al. 2007). The environment along and immediately ahead of the dryline helps to determine whether convection can initiate and sustain. Therefore, it may be useful to examine conditions immediately ahead of the dryline. One drawback of the creation of composites in Chapter 2 is that they do not specifically provide information on the pre-dryline environment. The dryline location was not taken into account because of a focus on synoptic-scale conditions.

While creating a climatology of dryline convection, Weather Prediction Center (WPC) analyses featuring drylines were obtained. Additionally, a database of convection initiation times and locations was created. By extracting dryline position data from the WPC analyses and combining it with the existing database on convection initiation, conditions in the vicinity of the dryline could be examined more thoroughly. How does the environment in the vicinity of the dryline affect the likelihood of convection initiation? Furthermore, are the results consistent with those in chapter 2?

## 3.2 Dryline data extraction

### a. WPC Dryline Locations

The WPC analyses are image files depicting features such as fronts, troughs, outflow boundaries, and drylines (Fig. 1.1). However, they do not have embedded geographical data, so obtaining dryline location data from the analyses is not a trivial task. Dryline locations were obtained in three stages.

Firstly, the images were filtered by colour to isolate drylines, troughs and outflow boundaries, all of which are represented by orange symbols in the WPC

analyses, from other features. Secondly, the images were converted to greyscale (Fig. 3.1), before a Gaussian filter was applied to isolate drylines (solid lines) from troughs and outflow boundaries (dashed lines). Finally, the images were overlaid on a Matplotlib (Hunter 2007) map with the same map projection so that accurate end points of each dryline could be obtained.
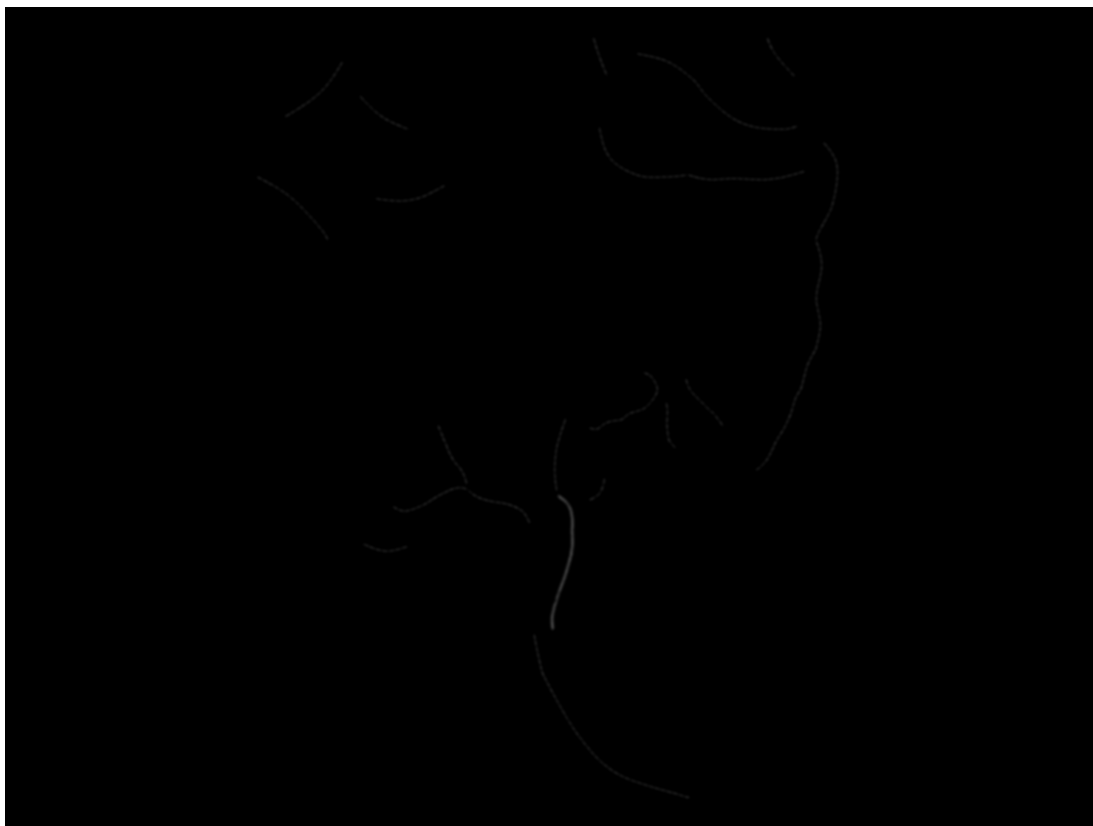


Figure 3.1: An example of a Weather Prediction Center surface analysis after all boundaries except drylines, troughs and outflow have been excluded. The image was converted to greyscale before a Gaussian filter was applied to isolate the dryline (the brightest line on the image) from troughs ad outflow boundaries (the fainter lines).

## b. Model Analysis Data

Given that cumulus associated with the dryline is most frequent within 25 km of the dryline itself (Ziegler et al. 1997), the NARR was clearly not suitable for examining the near-dryline environment as it only has a horizontal grid-spacing of 32 km. In recent years convection-allowing models such as the 3-km High

Resolution Rapid Refresh model (HRRR) (Benjamin et al. 2016) have been developed, however, data were not available for the entire period 2006–2015. The highest resolution model available throughout that period was the RAP, (formerly known as the Rapid Update Cycle (RUC)). The RAP has a horizontal grid-spacing of 13 km, has 50 vertical levels and typically outputs around 300 variables (see Appendix A). However, the precise number of variables output varies as minor updates are applied. The RAP was deemed to be the most suitable model to choose with regards to analysis frequency. The RAP analyses are available for every hour, as opposed to another mesoscale model, the North American Mesoscale Forecast System (Rogers et al. 2009), which is only available at intervals of six hours with a grid-spacing of 20 km. The RAP was also considered the most suitable source in terms of dryline positional accuracy. The accuracy of experimental convection-allowing models has been previously verified using the RAP as "truth" (Coffer et al. 2013), most likely because it uses a data assimilation scheme that incorporates radar reflectivity, satellite, and surface data hourly.

The RAP replaced the RUC (Benjamin et al. 2004) in May 2012. The RUC had a horizontal grid-spacing of 20 km, therefore the model data had to be preprocessed to maintain consistency throughout the ten years examined. Because no grid points were shared between the RAP and the RUC, data were obtained for the nearest model point to a relative location. Although convective drylines have an obvious reference point, namely the location of convection initiation, nonconvective drylines do not. By choosing a relative position for each dryline, it is possible to perform a fair comparison of days with and without convection. The reference location chosen was the northern end of the dryline. Dryline lengths ranged between approximately 100km and 1500km, so to ensure all drylines were included in analysis and to allow fair comparison, it seemed prudent to choose either the northern or southern end of the dryline. The northern end of the dryline (herein referred to as the dryline origin) often intersects a frontal boundary and can be an area of enhanced convergence. Therefore it is of great interest when considering convection initiation.

Data were initially obtained at 0.2° intervals, extending 3° north, south, east and west of the dryline origin, creating a 31 x 31 square. For each location, a value was obtained for every available variable at 1800 UTC on all dryline days,

excluding variables which were not present in all RUC and RAP analyses. After removing the aforementioned variables, 305 were left remaining (see Appendix B for further details). The time of 1800 UTC was chosen because it was the earliest analysis time available that had a large number of drylines (205 in total).

## 3.3    Machine Learning for Binary Classification

The combination of dryline locations and model data within the vicinity of the dryline provided a useful dataset for analysis. Given this dataset, dryline convection initiation can be reduced to a binary classification problem. Either a dryline initiated deep, moist convection, or it did not. Can an objective classification technique show any skill in determining whether convection will initiate along the dryline? If so, which variables are most important in making that determination? The results could confirm existing conceptual models, or perhaps reveal something new.

Recently, machine learning has been increasingly employed as a tool for analysing the increasingly large amount of meteorological data available. They are often used to automate tasks that may take a long time to perform manually. There are many potential methods for performing these analyses, however, the choice of technique applied is influenced by factors such as the size and type of input data, the interpretability of how a model makes its predictions, and the complexity and cost of implementation.

Linear regression is a relatively simple form of machine learning used to predict a non-categorical output variable. It has been used for such purposes as forecasting probabilities of severe weather from radar data (Kitzmiller et al. 1995), and to predict hail size from a combination of radar and modified upper-air soundings (Billet et al. 1997). Both of these studies were outputting probabilities, however, logistic regression is required when predicting binary outcomes. Mecikalski et al. (2015) used logistic regression to convective initiation at lead times of less than an hour using satellite and numerical weather prediction (NWP) parameters. They found that the logistic regression method incorporating NWP performed better than predicting initiation using satellite alone.

Linear regression is relatively easy to implement and can be a valuable tool when the relationships between the independent and dependent variables are not complex. However, it tends to perform poorly when the data is not linearly separable or there is a large degree of multicollinearity between the independent variables. Furthermore, logistic regression is prone to overfitting when there is a large number of input variables.

Another relatively simple form of machine learning is the use of decision trees (Breiman et al. 1984). Decision trees are constructed to predict an output category or value based on a set of input variables. At each branch node, the data is split by finding the variable and threshold value that produces the most information gain. Decision trees have also been used in meteorological classification problems. Gagne et al. (2009) determined that decision trees are a viable method to classify storm type using simulated and observed radar reflectivity. However, although decision trees are relatively easy to implement and interpret, they are prone to overfitting of training data (Schaffer 1993).

Ensembles of decision trees can reduce variance and the problem of overfitting. Two prominent ensemble methods are random forest classifiers (Breiman 2001), and gradient boosting (Friedman 2001). Random forest classifiers use an ensemble of decision trees trained on different subsets of data (McGovern et al. 2019). They have been used for a wide range of meteorological problems in recent years. Williams et al. (2008) used random forests to train a forecasting tool to predict thunderstorm intensity using multiple data sources including satellite, radar, weather model output and weather station data. Random forests have also been used to forecast the probability of hail from convection-allowing model output (Gagne et al. 2017), to forecast aircraft turbulence (Williams 2014), and to help identify drylines from weather model output (Clark et al. 2015). Clark et al. (2015) created an algorithm to identify drylines using gradients of specific humidity and dewpoint. Random forests were then used on the remaining drylines and model data to reduce the number of false positives.

In contrast to the independent trees generated using random forests, gradient boosting sequentially adds trees which are trained using the residual error of

the previous ensemble (Natekin and Knoll 2013). This learning is achieved by minimizing a loss function via gradient-descent. Far fewer meteorological studies have applied gradient boosting in comparison to random forests. McGovern et al. (2015) created a competition in which they invited artificial intelligence researchers and environmental scientists to produce solar energy forecasts. They found that the winning entries all used gradient boosting and that it outperformed random forests, linear regressions, and neural networks. Although these results may not apply to some other datasets, the authors hoped they would encourage more atmospheric scientists to use gradient boosting.

Although gradient boosting has historically lagged behind random forests in terms of how often it is applied, it has been used far more frequently of late. For instance, Ma et al. (2018) created an hourly rain area delineation scheme that applies gradient boosting to multispectral geostationary satellite and topographic elevation data. Gradient boosting has also been applied directly to weather model output. Xu et al. (2020) improved Weather Research and Forecasting model (WRF) near-surface wind speed forecasts using model output of wind speed and direction, temperature, pressure and vorticity.

More complex, but powerful machine learning techniques such as neural networks (Rumelhart and McClelland 1986) have also been used in the atmospheric sciences. A comprehensive description of neural network design and function is provided by Svozil et al. (1997). Neural networks have been used to predict hail size using radar-derived parameters (Marzban and Witt 2001), and using sounding-derived predictors (Manzato 2013). Neural networks (NN) have frequently been used to solve classification problems. For instance, Bankert (1994) classified cloud types from high resolution radiometer data, and Lakshmanan et al. (2014) classified precipitation type from polarimetric radar scans.

Several binary classification problems have also been addressed by using NN. Marzban and Stumpf (1996) applied NN to output from a mesocyclone detection algorithm to attempt to establish whether or not a tornado occurred. Similarly, NN was applied to the same alogrithm to identify if a damaging wind event occurred (Marzban and Stumpf 1998). Neural networks have also been used in conjunction with radar data for binary classification. Anagnostou (2004) used

radar reflectivity data to separate convective and stratiform rainfall. Lakshmanan et al. (2007) used NN for quality control purposes, separating meteorological and non-meteorological echoes.

Although NN has been used for many meteorological classification problems and tends to perform well when dealing with a large amount of input variables (Rumelhart and McClelland 1986), it suffers from poor interpretability (Kotsiantis et al. 2007). Furthermore, NN are difficult to implement and can be computationally expensive.

Machine learning is being increasingly used in the atmospheric sciences and is a valuable tool for objectively analyzing data. It can be used in conjunction with the data obtained when creating the climatology with a goal of providing further insight into dryline convection initiation. In particular, tree-based classifiers such as random forests and gradient boosting appear to be suitable for modelling dryline convection initiation. They are relatively simple to implement, do not suffer from the poor interpretability of techniques such as neural networks and are not as prone to overfitting as using singular decision trees.

Gradient boosting has been shown to outperform random forests in a variety of binary classification problems (Caruana and Niculescu-Mizil 2006). Therefore, it was chosen to test its viability in determining whether it can predict whether dryline convection will occur. Chapter 4 describes the training and testing of a gradient boosting model applied to drylines for the very first time.

[Blank page]

# Chapter 4

# Paper 2: Using Machine Learning to Predict Dryline Convection

# Using Machine Learning to Predict Dryline Convection

Trevor Mitchell* and David M. Schultz

*Centre for Atmospheric Science, School of Earth and Environmental Sciences, University of*

*Manchester, Manchester, United Kingdom*

*Corresponding author address:* Trevor Mitchell, Centre for Atmospheric Science, School of Earth

and Environmental Sciences, University of Manchester, Manchester, United Kingdom

E-mail: trevor.mitchell@manchester.ac.uk

# ABSTRACT

The viability of using machine learning to predict deep, moist convection along the Great Plains dryline was explored with a method known as gradient boosting. Using a dataset of 10 years of drylines occurring within a region of the Southern Great Plains, gradient boosting was applied to Rapid Update Cycle (Benjamin et al. 2004) and Rapid Refresh (Benjamin et al. 2016) model data obtained for 205 dryline days. The model was tested across two domains, firstly using data from locations relative to the dryline, and secondly from fixed locations across a domain spanning the contiguous United States. Tuned to maximize the harmonic mean of precision and recall, the gradient boosting model resulted in a high probability of detection of convective drylines (mean = 0.993). Low precision (mean = 0.638) resulted in a mean F1 score of 0.778 and accuracy of 0.638. Analysis of feature importance revealed a spatial variation in the performance of variables. In the dryline-relative domain, surface lifted index exhibited the highest values of feature importance east of the dryline. However, 675-hPa relative humidity was the leading variable in the most locations. The importance of instability and mid-level moisture in distinguishing between convective and non-convective drylines is consistent with the synoptic composites created in Chapter 2. In the large domain the 100 and 125-hPa u-component of the wind were leading variables at the highest number of locations, and also had the highest values of feature importance, mostly in the northwest US. Our results suggest that gradient boosting is not viable as a tool for directly predicting dryline convection, but may provide further avenues of research. We recommend that future work uses a larger dataset to determine whether gradient boosting could be improved with more training data.

## 1. Introduction

Determining whether deep, moist convection will initiate along the dryline is a difficult forecasting problem. Often, an ingredients-based approach is used to help forecast deep, moist convection with moisture, instability and lift being necessary ingredients (Doswell 1987). Forecast model output can be examined to identify whether these ingredients will be present in the vicinity of the dryline, but which variables best represent moisture, instability and lift? Furthermore, are any variables better discriminators of whether a dryline will produce deep, moist convection? In this paper we apply a machine learning technique to gridded model output to attempt to identify variables important for determining whether or not convection will occur on any day a dryline is present.

In recent years, machine learning has been increasingly used to solve other multi-variate problems in atmospheric science. Learning systems can be trained on samples of data where the outcome is already known, in what is known as supervised learning (Kotsiantis et al. 2007). Supervised learning has been used to solve regression problems such as predicting quantities of precipitation (e.g. Hall et al. 2002; Gagne et al. 2014), but also to automate classification tasks such as identifying cloud type (e.g. Lee et al. 2004; Xia et al. 2015), storm type (Gagne et al. 2009) and sudden stratospheric warming events (Blume et al. 2012).

Given a dataset of drylines and knowledge of whether there was convection, dryline convection initiation can be reduced to a binary classification problem. However, the selection of an appropriate classification method depends on many factors. The performance of a classification method can be sensitive to factors such as the number of input features, the balance of the dataset and the separation between the categories (Kotsiantis et al. 2007). However, even if one technique is found to outperform others in terms of accuracy, the practicality of implementation may be prohibitive. For example, although support vector machines (Cortes and Vapnik 1995) and neural networks

(Rumelhart and McClelland 1986) tend to perform well with a large amount of features, both techniques suffer from poor interpretability (Kotsiantis et al. 2007). Because the purpose of this study is to provide insight into the classification process rather than solely focus on the accuracy of predictions, the selection of a technique that allows the interpretation of variable importance is vital.

Tree-based classifiers such as decision trees (Breiman et al. 1984) fulfil this interpretability criterion and are also relatively simple to implement. However, complex trees are prone to overfitting of the training data (Schaffer 1993a). The problem of overfitting can be reduced by employing ensemble methods such as random forests (Breiman 2001) or gradient boosting (Friedman 2001), with minimal loss of interpretability. Random forest classifiers use an ensemble of decision trees trained on random samples of data. Each tree predicts a class, and the most common class is used as the model prediction. In contrast to the independent trees generated using random forests, gradient boosting sequentially adds trees which are trained using the residual error of the previous ensemble (Natekin and Knoll 2013). This learning is achieved by minimizing a loss function via gradient-descent. Gradient boosting generally outperforms random forests, provided that the boosting algorithm is correctly tuned.

The purpose of this paper is to explore the viability of gradient boosting for predicting whether a dryline will produce deep, moist convection. Using a dataset of 10 years of drylines occurring within a region of the Southern Great Plains, gradient boosting is applied to Rapid Update Cycle (Benjamin et al. 2004) and Rapid Refresh (Benjamin et al. 2016) model data obtained for each dryline day. Gradient boosting is tested with regards to two primary questions. Firstly, does the method have any skill in predicting whether a dryline will produce deep, moist convection? And if so, are any variables better discriminators of whether convection will initiate?

5

The layout of the paper is as follows. Section 2 describes the dryline dataset, the Rapid Refresh Model analyses and how gradient boosting was implemented. Section 3 explains how the algorithm was optimized to maximize performance and avoid overfitting. The viability of gradient boosting is tested over a small near-dryline domain in section 4. Section 5 analyses the performance of gradient boosting when applied to data obtained from the entire contiguous United States, before the paper is concluded in section 6.

## 2. Data and method

### a. Dryline dataset

A dataset of drylines was created from Weather Prediction Center analyses for all April, May and June days 2006–2015. Deep moist convection initiating along the dryline was subjectively determined from visible and infrared satellite, and radar imagery. A detailed description of the methods used to identify dryline convection is provided in chapter 2. Rapid Refresh (Benjamin et al. 2016) and Rapid Update Cycle (Benjamin et al. 2004) model data were obtained for 1800 UTC on each of the convective (123) and non-convective (82) dryline days where a dryline was present at 1800 UTC. Rapid Update Cycle (RUC) data was obtained for drylines occurring before May 2012. The model was replaced by the Rapid Refresh (RAP) in 2012, therefore RAP data was obtained for the remainder of the period.

Data were obtained at the nearest $1°$ interval for all variables available within each model analysis. This horizontal grid-spacing was chosen to allow for changes in resolution of the model data within the period of study. The RUC upgraded from a horizontal grid-spacing of 20 km to 13 km in 2007, therefore using fixed points was preferred to obtaining data from all grid points. For each location, a value was obtained for every available variable at 1800 UTC on all dryline days,

6

excluding variables which were not present in all RUC and RAP analyses. Using the remaining

model data from each dryline day, and the knowledge of whether or not that dryline produced con-

vection, a gradient boosting model was trained to predict whether a dryline would produce deep,

moist convection.

*b. Gradient boosting model*

Gradient boosting is an ensemble-based method that combines many decision trees to produce an

improved model. Individual decision trees are weak learners, often unable to adequately separate

and classify data. However, gradient boosting sequentially adds these weak learners to create

an ensemble of trees, improving the model with each iteration. A comprehensive description of

gradient boosting is provided by Natekin and Knoll (2013).

Each tree is constructed to predict an output category based on a set of input variables. At the

root of the tree, the data is split by finding the variable and threshold value that produces the most

information gain, obtained by calculating the entropy of the data before and after the split is made.

This process is then repeated for the branches created from the original split. The splitting process

continues in an iterative fashion, typically until a predetermined type of tree is reached.

After each tree is constructed, predictions are made for each sample of data using the class

probabilities from all the leaves in the ensemble of trees. A cost function is then calculated for

the residuals of the predictions and the known outcomes. The cost function used is the log loss

function, which punishes confident incorrect predictions more than it rewards confident correct

predictions. At each iteration, a decision tree is fit to the negative gradients of the cost function

from the previous iteration. The contribution of each tree to the ensemble is scaled by the learning

rate, a parameter that shrinks the contribution of each tree and slows the rate of convergence to a

solution. Gradient boosting can also be considered to follow the process outlined below:

7

124    1. Create an initial decision tree using observed categories

125    2. Calculate the error of the predictions of the initial tree

126    3. Adjust the predictions via gradient descent

127    4. Add the new tree to the ensemble

128    5. Create a new set of predictions from the expanded ensemble and calculate the errors

129    6. Construct a new tree

130    7. Adjust the predictions via gradient descent

131    8. Add the new tree to the ensemble

132    9. Repeat steps 5 to 8 until a predetermined number of trees have been created

133    A gradient boosting model was trained using RUC and RAP data obtained for 205 dryline days

134    using the Python package scikit-learn (Pedregosa et al. 2011). Although there were 329 dryline

135    days within the period of study, to maintain a consistent dataset, 89 days where a dryline was

136    not present at 1800 UTC were excluded. The number of drylines was reduced further by missing

137    data in the RUC and RAP archives. No 1800 UTC analyses were available for 11 days. Using

138    the remaining 205 days of data, each of the $1° \times 1°$ locations was treated independently, with a

139    gradient boosting model trained on data acquired from each location, and predictions only made

140    for that location.

141    The choice of metric to measure the skill of classification predictions is a topic of much dis-

142    cussion (e.g. Doswell 2004; Roebber 2009). If the event is rare or has dire consequences, false

143    positives may be acceptable, hence thresholds are often used that increase the probability of detec-

144    tion. If a missed event is unacceptable, a model will benefit from being tuned to maximize recall,

145    defined as TP/(TP+FN) where TP and FN stand for true positives and false negatives respectively.

8

Whereas if false positives are unacceptable, it is typical to maximize the precision, defined as TP/(TP+FP). However, increasing the probability of detection is also likely to increase the number of false positives (e.g. Brooks 2004).

Dryline convection is not rare. Approximately 60% of drylines identified in chapter 2 produced deep, moist convection. It could be argued that a high false alarm rate is acceptable because deep, moist convection can produce severe weather. However, not all deep, moist convection produces severe weather. We choose not to make a determination whether false positives or false negatives have greater cost with regards to dryline convection. We elect to tune the model by using the harmonic mean of precision or recall, the F1 score (Chinchor 1992), because it punishes large differences between them. The F1 score is defined as 2 x (precision x recall) / (precision + recall).

## 3. Parameter tuning

Maximizing the forecast skill can be achieved in a variety of ways because the scikit-learn package provides many different parameters which can be modified. The choice of values of these parameters is somewhat dependent on the goals of the project. For instance, factors such as the speed of training, the size of the dataset and the independence of variables may all influence the degree of complexity of the model implemented. Complex trees may produce higher skill, but can be computationally expensive and may also be prone to overfitting. There is no one "correct" method of implementing any machine learning technique, and the performance of the model is affected by variations in parameters to different degrees. Therefore, implementing a system of parameter tuning was deemed appropriate.

The selection of parameters to use in the gradient boosting classifier was an iterative process. The F1 score was computed for training and test data at 20 randomly selected locations. The performance of the model was optimized by varying one parameter at a time, while the other

9

parameters remained unchanged. Values were selected after examining how the model performed with training and test data (Fig. 1). The value at which the test score peaks was applied at the next iteration, when a different variable was examined. For example, in Fig. 1a, the test score is maximized using 60 trees. At that value, the training score is no longer increasing, therefore the model would not benefit from a further increase in the number of trees. Therefore, 60 trees were used for the remaining iterations.

Consideration of both the training and test scores in tandem is important because a high training score does not guarantee a well-trained model. A training performance much greater than test result can often be a sign of overfitting. Overfitting can occur as a result of a model training on noise from the training data and applying it too generally (Dietterich 1995), a common problem when dealing with a small dataset. Two techniques were used to avoid overfitting the dryline data: cross-validation and regularization.

Cross-validation (Stone 1974; Geisser 1975) is a technique that repeatedly holds out a different part of the dataset for testing at each iteration while training on the remaining data. A benefit of cross-validation is that it avoids the problem of choosing an unrepresentative sample and can increase the average performance of a model (Schaffer 1993b). K-fold cross validation is a technique that divides the dataset into k segments, with each segment being held out as test data in turn. A 5-fold stratified technique was applied to data from each location. Stratification attempts to preserve the proportions of each category in the training and test samples. The cross-validation process was performed 20 times, with the data resampled on each of the 20 occasions.

To test if the gradient boosting model was overfitting the dryline data, learning curves were plotted for 50 randomly chosen locations (Fig. 2). Learning curves show how the performance of the model on training and test data varies with the size of the sample. If the training performance does not increase with sample size, or is still increasing even at the largest sample size, it may

10

indicate the model is failing to learn enough information from the training data. This failure is known as underfitting. Conversely, a model may be overfitting if the model learns too much from noise in the training data, resulting in a training performance much higher than the test performance.

Figure 2a shows learning curves after the iterative parameter-tuning process, and provides an example of overfitting. The model performs relatively well on the training data, but much worse on test data. This overfitting was reduced using regularization; the simplification of a model. In the case of gradient boosting, this is typically achieved by reducing the learning rate, or reducing the complexity of the trees used. Varying the number of trees and the learning rate are the main methods of regularization, however, there are a number of tree attributes that can also affect the performance of the model. Other methods of reducing overfitting include, but are not limited to, reducing the depth of trees used in the ensemble, reducing the number of features randomly selected before making a splitting decision, and subsampling data before creating each tree.

The parameters used to produce the learning curves in Fig. 2a were as follows: learning rate = 0.1, number of trees = 60, maximum tree depth = 2 layers below the root, maximum features before choosing a split = 6, and subsampling = 90%. The overfitting observed in Fig. 2a was eradicated by reducing the learning rate to 0.01 (Fig. 2b). The updated parameters were then applied to data acquired from within the vicinity of the dryline to test the predictive ability of gradient boosting and identify important variables in the classification process.

## 4. Dryline-relative domain

Fair comparison of convective and non-convective dryline using machine learning requires careful consideration of which locations from which to extract data. Convective drylines provide a reasonable option for choice of location from which to obtain data: the location of first convection

11

initiation. But what location should be chosen for non-convective dryline days? Days with no convection initiation have no obvious point of reference. One possible approach would be to select a fixed location for both convective and non-convective days. However, a weakness of this approach is caused by day-to-day variations in dryline position. A fixed location may sample from ahead of the dryline in some cases, but behind it in others. In this scenario a comparison of small-scale dryline variations would be futile.

It seems reasonable to expect that fair comparison of drylines requires some sort of commonality with regards to the dryline position. Fixed locations struggle to achieve this characteristic, but using locations that are fixed relative to the dryline location offers an alternative approach. Data were obtained from locations relative to a location along the dryline at 1800 UTC, extracted from Weather Prediction Center analyses.

Dryline lengths ranged between approximately 100km and 1500km, so to ensure all drylines are included in analysis and to allow fair comparison, it seemed prudent to choose either the northern or southern end of the dryline. The northern end of the dryline often intersects a frontal boundary and can be an area of enhanced convergence. Therefore it is of great interest when considering convection initiation. The reference point chosen was the northernmost extent of the dryline, herein referred to as the dryline origin. Data were obtained at $0.2°$ intervals, extending $3°$ north, south, east and west of the dryline origin, creating a 31 x 31 square.

*a. Forecast skill*

At each point, a gradient boosting model was trained and tested on the data obtained from that location. At each point, four forecast metrics were calculated: the F1 score, precision, recall, and the accuracy, defined as the proportion of drylines that were correctly categorized (Fig. 3). The values in Fig. 3 represent the skill of predictions of whether dryline convection would occur

anywhere along the dryline, given data obtained from each location. For example, an accuracy of 0.650 was produced by using 229 samples of data only obtained from point X.

The F1 score ranged between 0.774 in areas well southwest of the dryline to 0.795 east of the dryline origin. The region with best skill was between $1°$ and $2°$ east of the dryline origin, at around $0°$ relative latitude. The environment east of the dryline is important in determining whether incipient convection can intensify as it moves east, thus the region of higher skill east of the origin is not surprising.

Regions with highest F1 score were collocated with regions of higher precision, lower recall, and higher accuracy (Fig. 3b,c,d). The higher precision implies that a higher proportion of drylines labelled as convective are correct using data from these areas. However, the lower recall suggests that in these locations thresholds are used which lower the probability of detection of convective drylines. In general, recall is high over the whole domain, with no values lower then 0.970. The model is very effective at detecting convective drylines, but the large difference between recall and accuracy suggests that the high probability of detection comes at a cost. The model appears to be labelling too many drylines as convective, and labels some non-convective drylines incorrectly.

*b. Variable importance*

Highest F1 scores are observed east of the origin, but which variables were most important in making those predictions? Figure 4 shows the variable with highest feature importance for each location. The feature importance is a measure of how important a variable is in the decision trees created, calculated by averaging the information gain produced by the variable across all trees. Values range from 0 to 1, with high values representing variables that have high importance.

Relative humidity at 675 hPa was most frequently the most important feature (n=67). A cluster of locations in the east-central portion of the domain featured 675-hPa relative humidity as the

13

best variable. In general, mid-level moisture appears to be important in determining whether dryline convection will occur. In addition to relative humidity at 675 hPa, there is also an area along and immediately east of the dryline where 825-hPa relative humidity is the leading variable. Furthermore, relative humidity at 650 and 725 hPa also appear in the top ten variables by number of locations.

The surface lifted index (herein lifted index) was also an important variable east of the dryline. The lifted index is defined as the difference between the environmental temperature at 500 hPa and the theoretical temperature of an air parcel lifted dry adiabatically from the surface until it reaches saturation and then moist adiabatically up to 500hPa. The lifted index was the leading variable in a region around a relative latitude of -0.5, and relative longitude 1 (-0.5, 1) and (-2, 2). The presence of the lifted index as a leading variable east of the dryline is most likely because it is a measure of instability, a necessary ingredient for deep, moist convection. However, west of the dryline instability is usually minimal, meaning distinguishing between convective and non-convective days is more difficult using data from locations there.

Overall, it appears that feature importance has some regional dependence. Several distinct areas are observed where a particular variable performs best. For instance, northwest of the dryline, at around (1, -2.5), there is a an area where 125-hPa temperature is the leading variable. Although 675-hPa relative humidity and the lifted index are measures of moisture and instability, it is not immediately obvious how 125-hPa temperature can influence the likelihood of convection along the dryline. Over the Great Plains, 125-hPa would typically be in the lower stratosphere. We suggest that the 125-hPa temperature may be a proxy for the height of the tropopause, and the strength of upper-level troughing. The passage of short-wave troughs is associated with stronger drylines (Schultz et al. 2007), which may increase the likelihood of convection.

Another prominent variable that does not appear to relate to the three ingredients of deep, moist convection is the v-component of the wind at 400 hPa. It is the leading variable in a region centred around (2, 1). It is unclear as to why this variable may be a useful in distinguishing between convective and non-convective drylines. However, it is likely that its appearance as a leading variable is also related to troughing. The v-component of the wind at 400 hPa may be an indicator of the amplification of the upper-level flow.

Although the v-component of the wind at 400 hPa is the leading variable in several locations, mainly in the northern half of the domain, how well does it perform relative to the other variables discussed? The feature importance was calculated at every location for four variables: 675-hPa relative humidity, lifted index, 125-hPa temperature and the 400-hPa v-component of the wind (Fig. 5). Given that 305 variables were used in the model, if each variable were equally important, they would all have a feature importance of 0.003. However, the four variables all have a feature importance in excess of 0.300 in the areas where they perform best. Each variable tends to have its highest values of importance collocated with where it was the lead variable overall.

Relative humidity at 675 hPa is the leading variable at the highest number of locations, but the lifted index is the variable that achieves the highest values of feature importance. Around $1°$ east of the dryline origin the feature importance is as high as 40%. Yet, there are areas east of the dryline where the lifted index is disregarded by the model. In fact, all of the four variables shown have large areas with zero feature importance.

How can a variable have a feature importance of 0.4 in one location and an importance of zero at a location only $1°$ away? Some variables have high feature importance because those variables were preferred when determining the best split. Other variables may also provide a large information gain if they were used to split the data. However, if a variable does not produce the best information gain, it is essentially ignored until another split is made. Hypothetically a

15

variable could be only surpassed by only one variable at each node in every tree, yet have a feature importance of zero. The splitting process outlined here may explain why a variable such as 675-hPa relative humidity leads in 67 locations, 725-hPa relative humidity in 40 locations, yet the level in between, 700-hPa relative humidity, only leads in 20 locations. Physically, there seems little reason why 700-hPa relative humidity would perform worse than the levels adjacent to it.

Overall, gradient boosting produces highest accuracy to the east of the dryline, but the accuracy and feature importance show large variation over small spatial scales. Given that at lead times longer than a few hours the precise dryline location can be difficult to predict, how does gradient boosting perform when the location is not known? The following section presents the results of the gradient boosting model applied to fixed locations across the contiguous United States.

## 5. Fixed-location domain

Data were obtained for 1800 UTC at $1°$ x $1°$ intervals for the contiguous United States, with the dryline location ignored. Rapid Refresh model data is available on a 451 x 337 grid with horizontal grid-spacing of 13 km, However, older RUC data uses a grid of 301 x 225 with a grid-spacing of 20 km. Because the two sets of data do not have grid points at the same geographical locations, data was extracted to the nearest degree of longitude and latitude. When the model contained no data at the requested grid point, the nearest model location was used.

*a. Forecast skill*

Figure 6 shows the forecast skill for the larger domain. In general, the F1 score of gradient boosting using fixed locations is similar to the score obtained using locations relative to the dryline. The model produces a small range of F1 scores, with a minimum of 0.774 and a maximum of 0.788. Highest skill is observed in the northern Rockies, but there are also regions of relatively

16

high skill in the central Plains and the Lower Midwest. As was observed in the smaller domain, the regions of higher F1 score are collocated with higher precision and accuracy, and lower recall. These results provide more evidence that the model is labelling too many drylines as convective.

*b. Variable importance*

Which variables are the most important in these high scoring regions? Figure 7 shows the leading features using data from each location in the large domain. The u-component of the wind at high altitudes dominates over most of the northwestern US. More specifically, the 100-hPa and 125-hPa u-component of the wind are the leading two variables by frequency, with 72 and 69 locations respectively. However, there is a large gap in frequency to the third best variable. The 100-hPa v-component of the wind is the leading variable in 37 locations, and those locations are widely scattered.

The importance of the zonal wind at high altitudes may be related to the effects of lee troughing. Stronger westerlies across the Rockies would induce greater lee troughing, and a stronger dryline. High-altitude winds are also important in the central Plains and lower Midwest, but most of the region had a leading variable that was not among the ten best variables in the US. Figure 8 shows the five leading variables in a region comprising some of the central Plains and Lower Midwest. In this region, the leading variables are all related to the v-component of the upper-tropospheric jet.

Other areas of note are the southwest, where 125-hPa temperature is a leading variable, and the southern Plains, where the lifted index leads in a few more scattered locations. However, Figure 7 does not inform how well these variables perform in other locations. Figure 9 shows the spatial variation in feature importance for the 100-hPa u-component of the wind, lifted index, 125-hPa temperature and the 225-hPa v-component of the wind. Apart from the latter variable, which did

17

not appear in the top ten leading variables over the contiguous US, the variables tended to have their highest importance in similar regions to where they were the leading variable. On average, the 100-hPa u-component of the wind had higher importance than the other three variables. However, the highest values of feature importance are lower than those observed in the smaller domain.

Despite these lower feature importance values, Fig. 9 shows that feature importance is strongly dependent on location. The lifted index has its highest values in southeast Colorado and northeast New Mexico, and generally performs better in the southern Plains than elsewhere. It may not be coincidence that the lifted index performs well near the typical dryline location, because it is a measure of two ingredients of deep, moist convection: moisture and instability.

Meanwhile, 125-hPa temperature has highest values over a larger area than the lifted index, extending from Wyoming down to the Mexico border Fig. 9c. This variable may be an indicator of the strength of troughing encroaching into the Plains. Not only is troughing associated with increased dryline convergence, but cyclonic vorticity advection downstream of a trough is associated with ascent and cooling, helping to reduce convective inhibition.

## 6. Discussion

This study aimed to answer two main questions. Firstly, is gradient boosting a viable method for predicting dryline convection? And secondly, which variables are the best discriminators of whether a dryline will produce convection?

*a. The viability of gradient boosting*

The forecast skill of gradient boosting was primarily measured with the F1 score, which was similar in magnitude and range in both domains, falling between 0.77 and 0.79. These scores generally occurred as a result of high recall, but much lower precision. Values of recall never fell

18

below 0.96, implying the method applied here has a high probability of detection of convective drylines. However, precision did not exceed 0.65 in any location, even close to the dryline. The model appears to use thresholds that enable a high probability of detection of convective events as the expense of a large number of false positives. These false positives explain the modest accuracy of around 0.64.

Although peak accuracy values were slightly higher in the small domain, in general the accuracy was similar for both domains. This result suggests that gradient boosting does not suffer in performance when the dryline position is not provided. However, despite the potential benefit of not requiring the dryline location to make a forecast, the accuracy would need to be increased for the method to be used to directly forecast dryline convection. Predictive accuracy did not exceed 0.675, even in areas near where a dryline would typically be located.

Why might the methods applied in this paper have such a modest accuracy? The answer could be split into two categories. Firstly, it may be possible that no model will ever be able to predict dryline convection with high accuracy because the problem is too complex. Alternatively, the methods used here were simply inadequate to produce accurate predictions of dryline convection.

If it is possible to accurately predict dryline convection, the most likely explanation for why gradient boosting does not perform better is that decision trees are too simplistic to model potentially complex relationships between dependent variables. This weakness is exacerbated when the model can only be trained on a small dataset. However, it is possible that despite parameter tuning, there is still some combination of parameters that could yield a higher performance than obtained here.

*b. Variable importance*

Although gradient boosting does not appear to be a viable technique to accurately forecast dryline convection, it may provide some insight into the variables that may best determine whether dryline convection will occur. A strong signal exists that there is a regional variation in variables important for predicting dryline convection. However, the high number of false positives suggests that although these variables may be effective in detecting convective drylines, they may not be good at distinguishing between convective and non-convective drylines.

Despite this weakness, analysis of the importance of individual variables yields interesting results. East of the dryline, mid-level relative humidity and the lifted index were leading variables. However, variables seemingly unrelated to the ingredients of deep, moist convection were also prominent. An example is provided by the 125-hPa temperature, which was the leading variable in a region northwest of the dryline origin.

The temperature at 125-hPa also appears in the top ten variables in the large domain. It is the leading variable in a few scattered locations in the southwest US. We propose that this variable is related to the height of the tropopause, a proxy for the strength of upper-level troughing. Synoptic-scale features are also proposed as an explanation for perhaps the most striking result in the analysis of the large domain. The leading two variables by number of locations are the u-component of the wind at 100 and 125 hPa respectively. These two variables dominate in the central Rockies. We suggest that the magnitude of flow across the Rockies is important because stronger flow promotes greater lee troughing, and strengthens the dryline.

However, despite our speculation about why these variables have highest feature importance, we are unable to determine the nature of the relationships between these variables and the likelihood of convection. Gradient boosting does not allow the opportunity to examine the thresholds used

20

when making splits because multiple trees are used and any given variable may be used on more than one occasion. Therefore, the significance of these results remains unclear. We recommend that similar experiments should be performed with another tree-based classifier that produces a measure of feature importance, random forests. Random forests could be used to answer several questions. Firstly, can higher forecast skill be obtained using a method other than gradient boosting? Secondly, does predictive accuracy of dryline convection have an upper limit? And finally, will the variables with highest feature importance in this study also be found using random forests?

## 7. Conclusion

A dataset of ten years of drylines within the southern Great Plains was used as "truth" to investigate whether machine learning can be used to predict whether a dryline will produce deep, moist convection. A gradient boosting model was tuned and tested using RAP and RUC data, initially obtained from locations relative to the dryline location, then from fixed locations across the contiguous United States. The model was tuned to maximize the harmonic mean of precision and recall, the F1 score. Forecast performance was similar across both domains, suggesting that knowledge of the dryline location has little effect on the performance of the model. F1 scores ranged from 0.774 to 0.795, a result of high recall (mean = 0.993), but lower precision (mean = 0.638). The model has a high probability of detection of convective drylines, but at the expense of a high false alarm rate. These results suggest that gradient boosting is not viable as a tool for directly predicting dryline convection.

Analysis of feature importance revealed a spatial variation in the performance of certain variables. In the dryline-relative domain, lifted index had the highest values of feature importance, mostly east of the dryline. However, 675-hPa relative humidity was the leading variable in the most locations. Regional variations in feature importance were also observed in the larger domain.

21

The 100 and 125-hPa u-component of the wind were leading variables at the highest number of locations, and also had the highest values of feature importance. The u-component of the wind at high altitude was the best variable in the northwest US. The prominence of these variables may be related to the effects of lee troughing. Stronger flow across the Rockies would likely lead to stronger lee troughing, and dryline intensification. However, gradient boosting is unable to reveal the nature of the relationships between variables and the likelihood of dryline convection.

The gradient boosting method applied in this study demonstrates that machine learning techniques have potential as a method for understanding dryline convection. However, future work would benefit from a larger dataset for two main reasons. Firstly, to explore whether the forecast skill of gradient boosting would be improved given more training data, and secondly, to reduce the possibility of overfitting. The model was trained on a small dataset and is susceptible to noise, despite the efforts to minimize overfitting during parameter tuning.

It is possible that alternative machine learning techniques such as support vector machines or neural networks may yield higher accuracy than gradient boosting. However, these techniques would suffer from a lack of interpretability of result with regards to feature importance in comparison with tree-based methods. Therein lies a constant dilemma of machine learning applications. A similar study that uses random forests rather than gradient boosting would help determine whether machine learning has potential as a forecasting tool. Furthermore, if similar variables were found to be important it would motivate the need for further investigation into large-scale influence on dryline convection.

## References

Benjamin, S. G., and Coauthors, 2004: An hourly assimilation-forecast cycle: The RUC. *Mon. Weather Rev.*, **132**, 495–518, doi:10.1175/1520-0493(2004)132⟨0495:ahactr⟩2.0.co;2.

Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Weather Rev.*, **144**, 1669–1694, doi:10.1175/mwr-d-15-0242.1.

Blume, C., K. Matthes, and I. Horenko, 2012: Supervised learning approaches to classify sudden stratospheric warming events. *J. Atmos. Sci.*, **69**, 1824–1840, doi:10.1175/jas-d-11-0194.1.

Breiman, L., 2001: Random forests. *Mach. Learn.*, **45**, 5–32, doi:10.1023/A:1010933404324.

Breiman, L., J. Friedman, R. Olshen, and C. Stone, 1984: *Classification and regression trees.* Wadsworth International Group, 368 pp.

Brooks, H. E., 2004: Tornado-warning performance in the past and future. *Bull. Am. Meteorol. Soc.*, **85**, 837–843, doi:10.1175/BAMS-85-6-837, URL https://journals.ametsoc.org/doi/pdf/10.1175/BAMS-85-6-837.

Chinchor, N., 1992: MUC-4 Evaluation Metrics. *Proceedings of the 4th Conference on Message Understanding*, Association for Computational Linguistics, USA, 2229, MUC4 '92, doi:10.3115/1072064.1072067, URL https://doi.org/10.3115/1072064.1072067.

Cortes, C., and V. Vapnik, 1995: Support-vector networks. *Mach. Learn.*, **20**, 273–297, doi:10.1023/A:1022627411411.

Dietterich, T., 1995: Overfitting and undercomputing in machine learning. *ACM Comput. Surv.*, **27**, 326–327, doi:10.1145/212094.212114.

Doswell, C. A., 1987: The distinction between large-scale and mesoscale contribution to severe convection: A case study example. *Weather Forecast.*, **2**, 3–16, doi:10.1175/1520-0434(1987) 002⟨0003:TDBLSA⟩2.0.CO;2.

Doswell, C. A., 2004: Weather forecasting by humans–Heuristics and decision making. *Weather Forecast.*, **19**, 1115–1126, doi:10.1175/WAF-821.1.

Friedman, J. H., 2001: Greedy function approximation: A gradient boosting machine. *Ann. Stat.*, 1189–1232.

Gagne, D. J., A. McGovern, and J. Brotzge, 2009: Classification of convective areas using decision trees. *J. Atmos. Ocean. Technol.*, **26**, 1341–1353, doi:10.1175/2008JTECHA1205.1.

Gagne, D. J., A. McGovern, and M. Xue, 2014: Machine learning enhancement of storm-scale ensemble probabilistic quantitative precipitation forecasts. *Weather Forecast.*, **29**, 1024–1043, doi:10.1175/waf-d-13-00108.1.

Geisser, S., 1975: The predictive sample reuse method with applications. *J. Am. Stat. Assoc.*, **70**, 320–328, doi:10.1080/01621459.1975.10479865.

Hall, T., H. E. Brooks, and C. A. Doswell, 2002: Precipitation forecasting using a neural network. *Weather Forecast.*, **14**, 338–345, doi:10.1175/1520-0434(1999)014⟨0338:pfuann⟩2.0.co;2.

Kotsiantis, S. B., I. Zaharakis, and P. Pintelas, 2007: Supervised machine learning: A review of classification techniques. *Emerg. Artif. Intell. Appl. Comput. Eng.*, **160**, 3–24.

Lee, Y., G. Wahba, and S. A. Ackerman, 2004: Cloud classification of satellite radiance data by multicategory support vector machines. *J. Atmos. Ocean. Technol.*, **21**, 159–169, doi:10.1175/ 1520-0426(2004)021⟨0159:CCOSRD⟩2.0.CO;2.

24

Natekin, A., and A. Knoll, 2013: Gradient boosting machines, a tutorial. *Front. Neurorobot.*, **7**, 21, doi:10.3389/fnbot.2013.00021.

Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in Python. *JMLR*, **12**, 2825–2830.

Roebber, P. J., 2009: Visualizing multiple measures of forecast quality. *Weather Forecast.*, **24**, 601–608, doi:10.1175/2008waf2222159.1.

Rumelhart, D. E., and J. L. McClelland, 1986: Learning internal representations by error propagation. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*, MIT Press, 318–362.

Schaffer, C., 1993a: Overfitting avoidance as bias. *Mach. Learn.*, **10**, 153–178, doi:10.1023/A:1022653209073.

Schaffer, C., 1993b: Selecting a classification method by cross-validation. *Mach. Learn.*, **13**, 135–143, doi:10.1023/A:1022639714137.

Schultz, D. M., C. C. Weiss, and P. M. Hoffman, 2007: The synoptic regulation of dryline intensity. *Mon. Weather Rev.*, **135**, 1699–1709, doi:10.1175/mwr3376.1.

Stone, M., 1974: Crossvalidatory choice and assessment of statistical predictions. *J. R. Stat. Soc. Ser. B*, **36**, 111–133.

Xia, M., W. Lu, J. Yang, Y. Ma, W. Yao, and Z. Zheng, 2015: A hybrid method based on extreme learning machine and k-nearest neighbor for cloud classification of ground-based visible cloud image. *Neurocomputing*, **160**, 238–249, doi:10.1016/j.neucom.2015.02.022.

# LIST OF FIGURES

26

FIG. 1. The performance of the gradient boosting model when varying (a) the number of trees, (b) the maximum depth of trees, (c) the maximum number of features considered when making a split, and (d) the size of subsample of the data. 60 trees were used in the remaining tests following (a), a maximum tree depth of 2 was carried forward after (b), and maximum features were set at 6 following (c). In addition to these values, a subsample value of 0.95 was applied for all modeling thereafter.
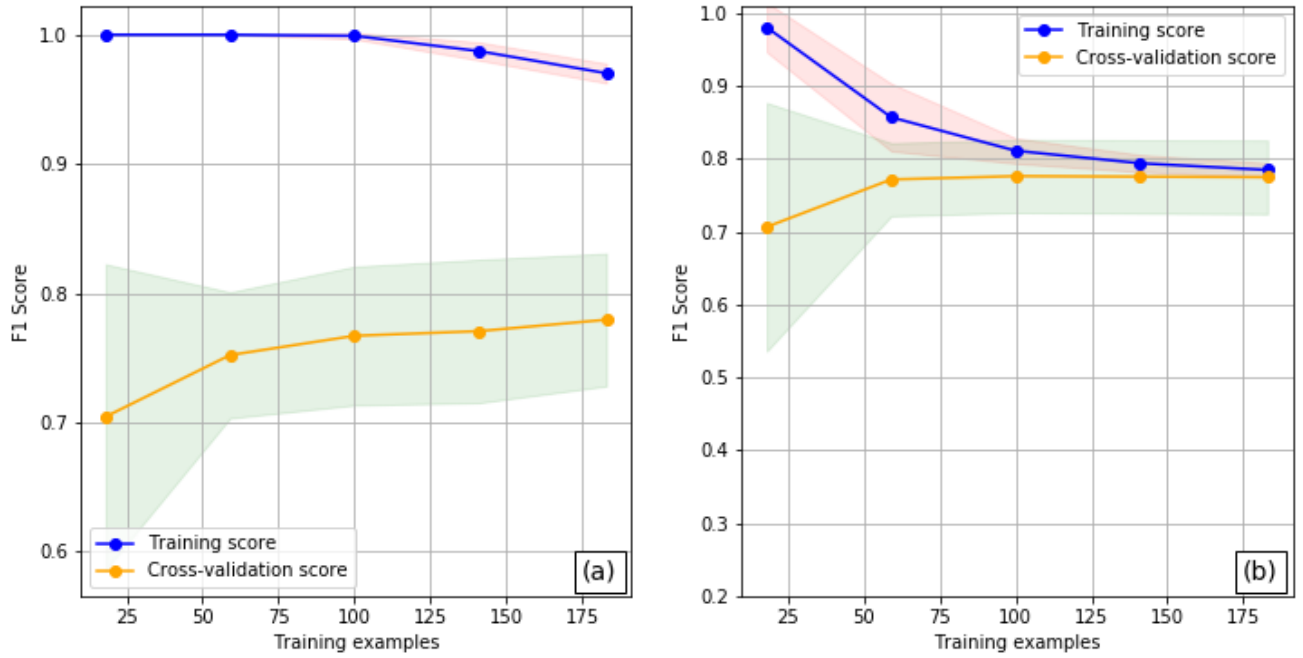
FIG. 2. Learning curves (a) before and (b) after regularization. The learning curves in (a) were created using 60 trees of maximum depth 2, with a learning rate of 0.1 and subsampling 90% of the data. The learning curves in (b) were created using the same parameters, except for a learning rate of 0.01.

28
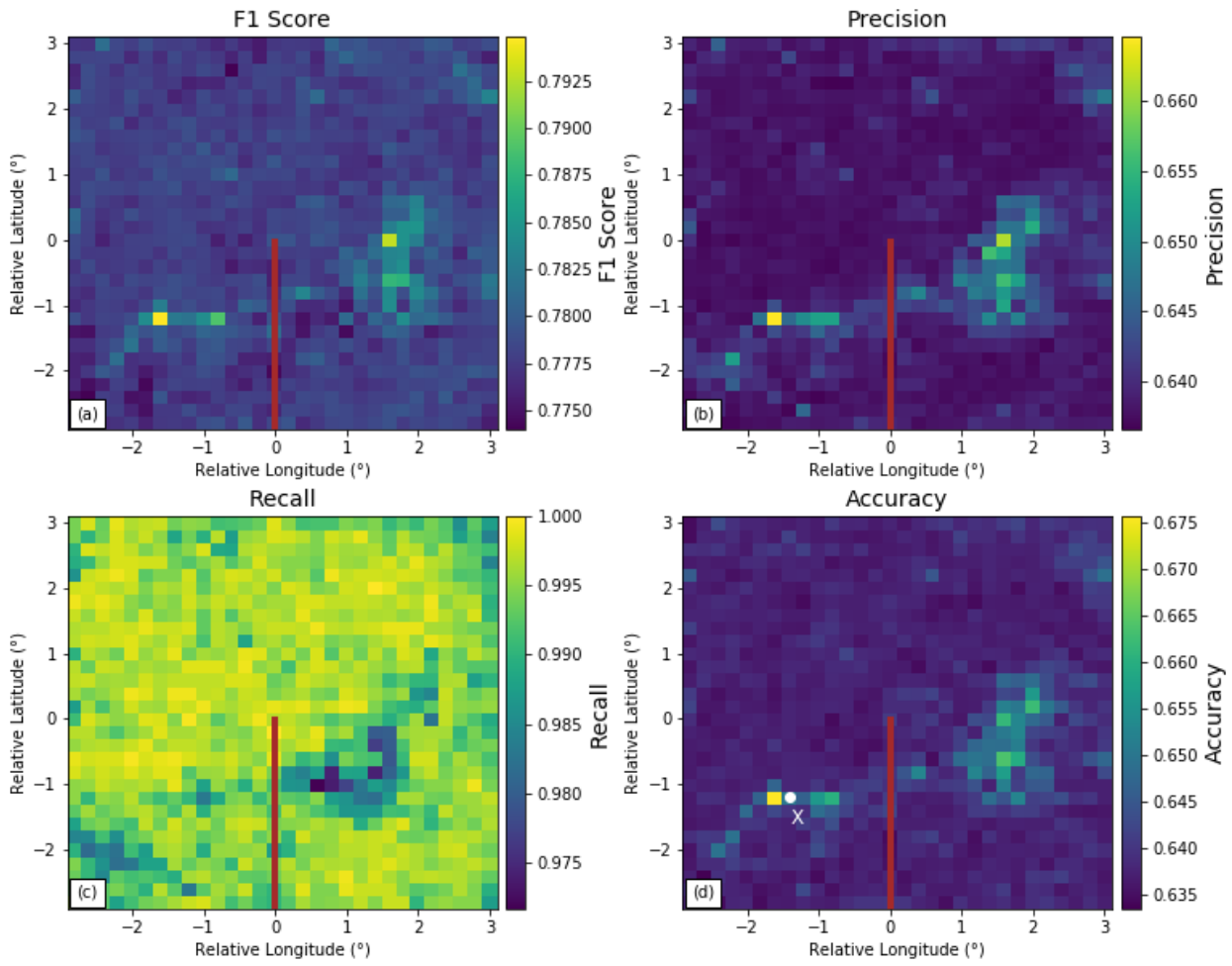
FIG. 3. Forecast skill metrics (a) F1 score, (b) precision, (c) recall and (d) accuracy using data from locations relative to the northern end of the dryline. The dryline location is indicated by the brown line. The scores for each location are calculated by using data from only that location.

29
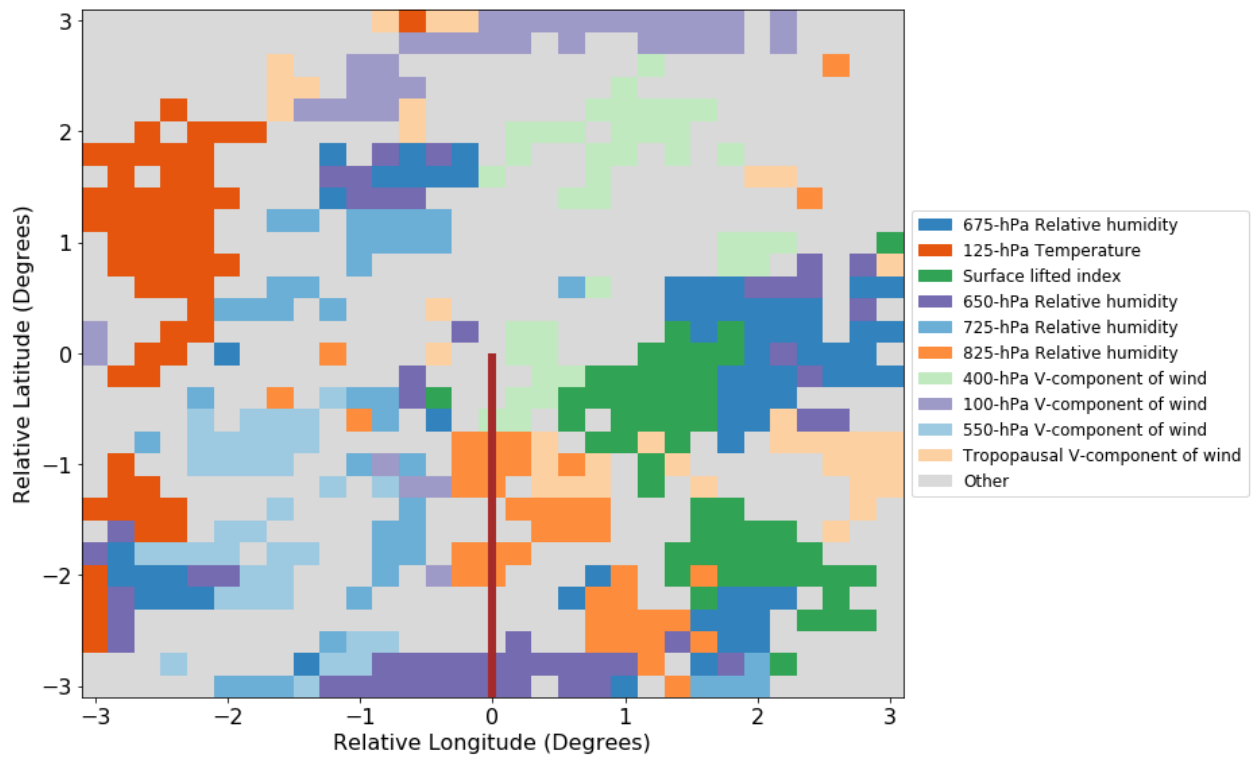
FIG. 4. The variable with highest feature importance using data from locations relative to the northern end of the dryline. The dryline is indicated by the dark brown line. The top ten variables by frequency are shown.
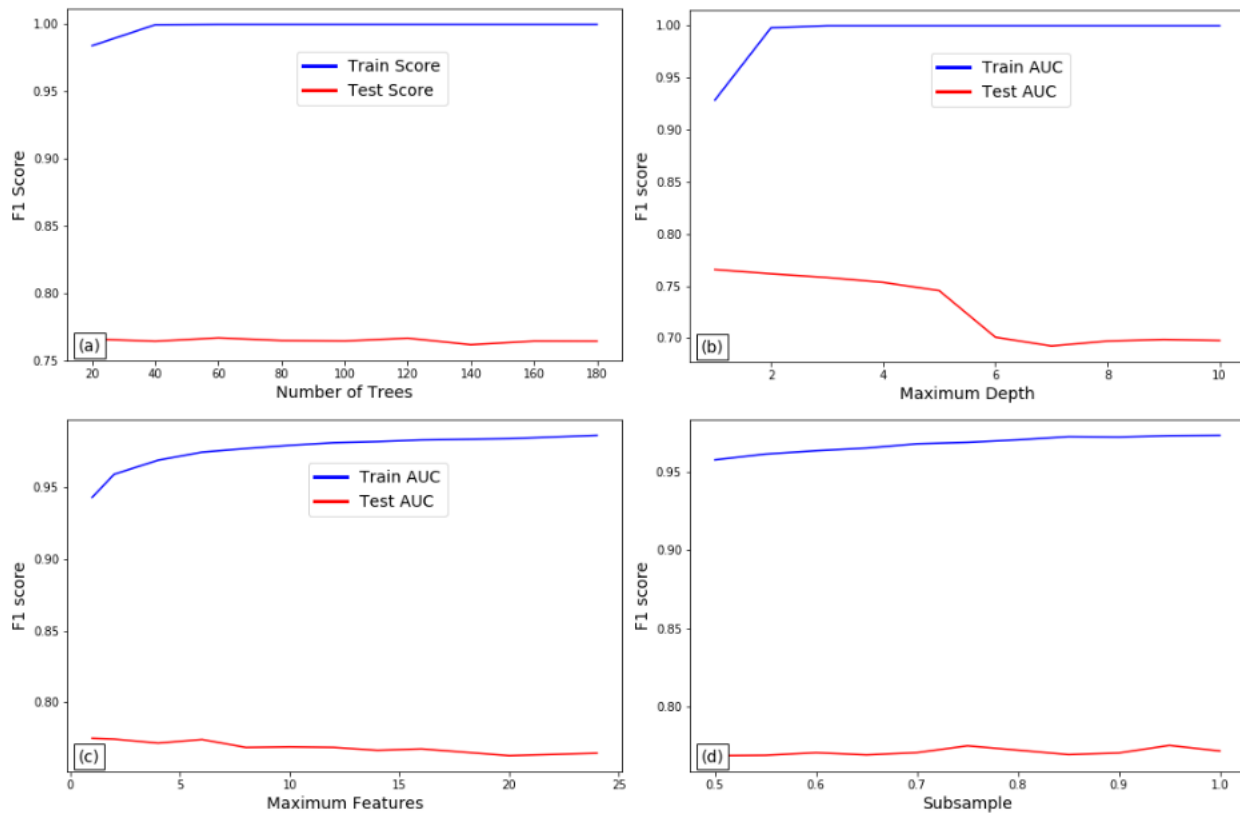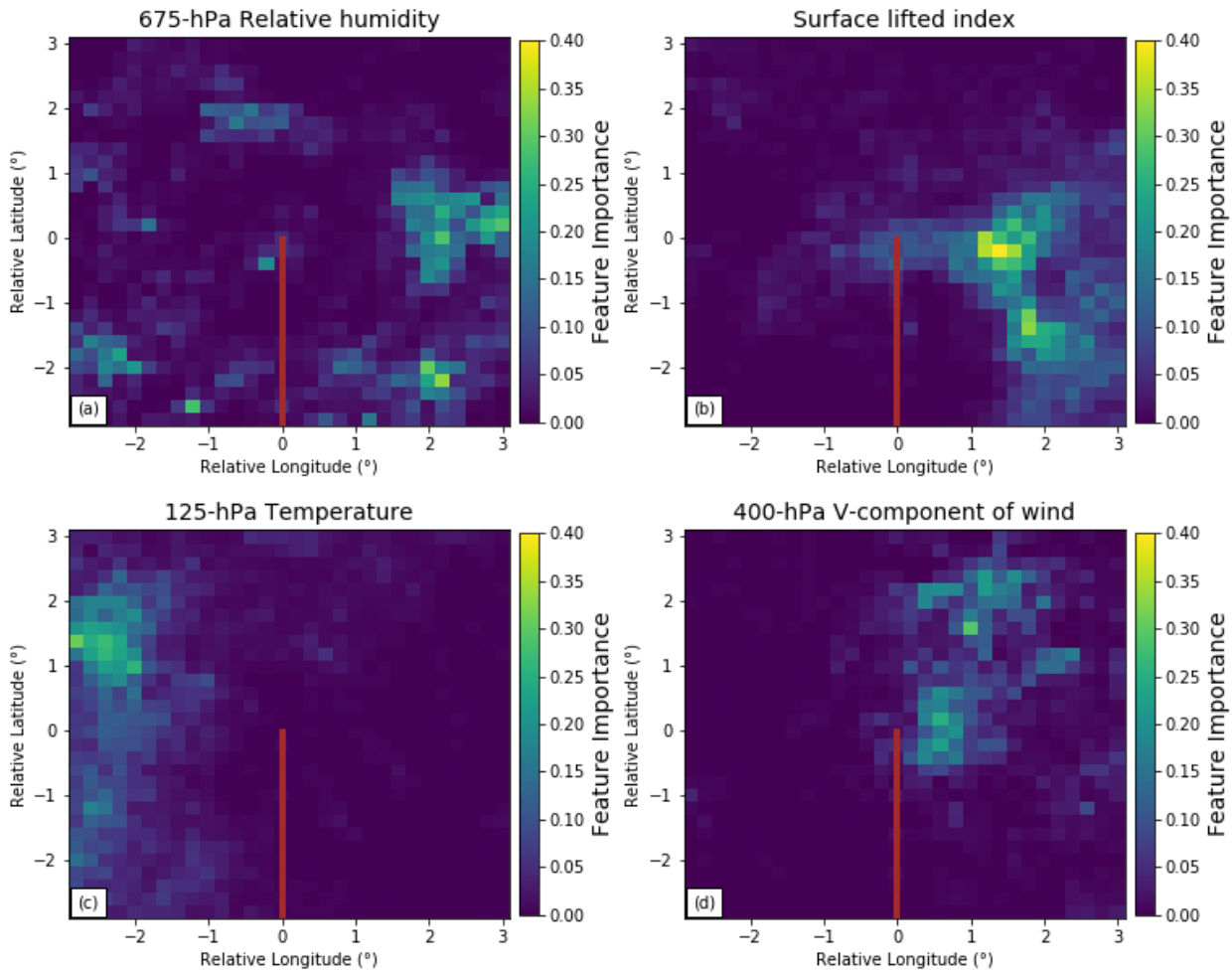
FIG. 5. Feature importance for (a) 675-hPa relative humidity, (b) surface lifted index, (c) 125-hPa temperature and (d) the 400-hPa v-component of the wind. The importance for each location is calculated by using data from only that location.
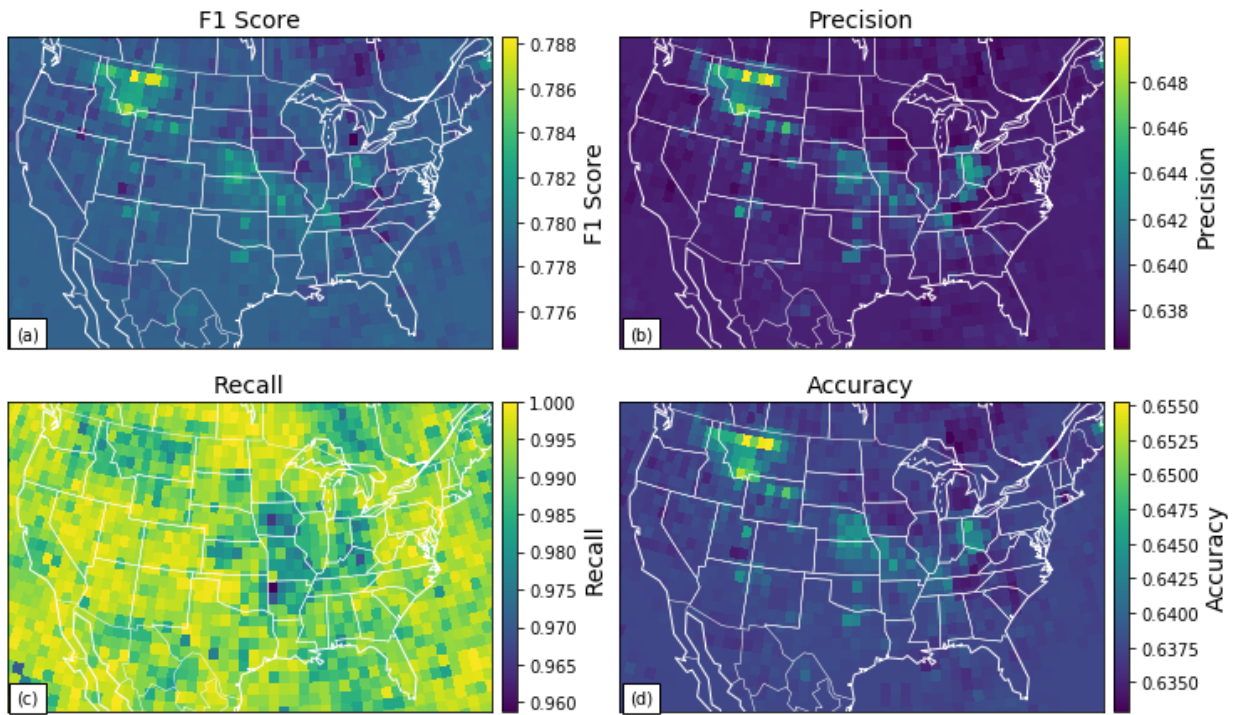
FIG. 6. Forecast skill metrics (a) F1 score, (b) precision, (c) recall and (d) accuracy using data from across the contiguous United States. The scores for each location are calculated by using data from only that location.

32
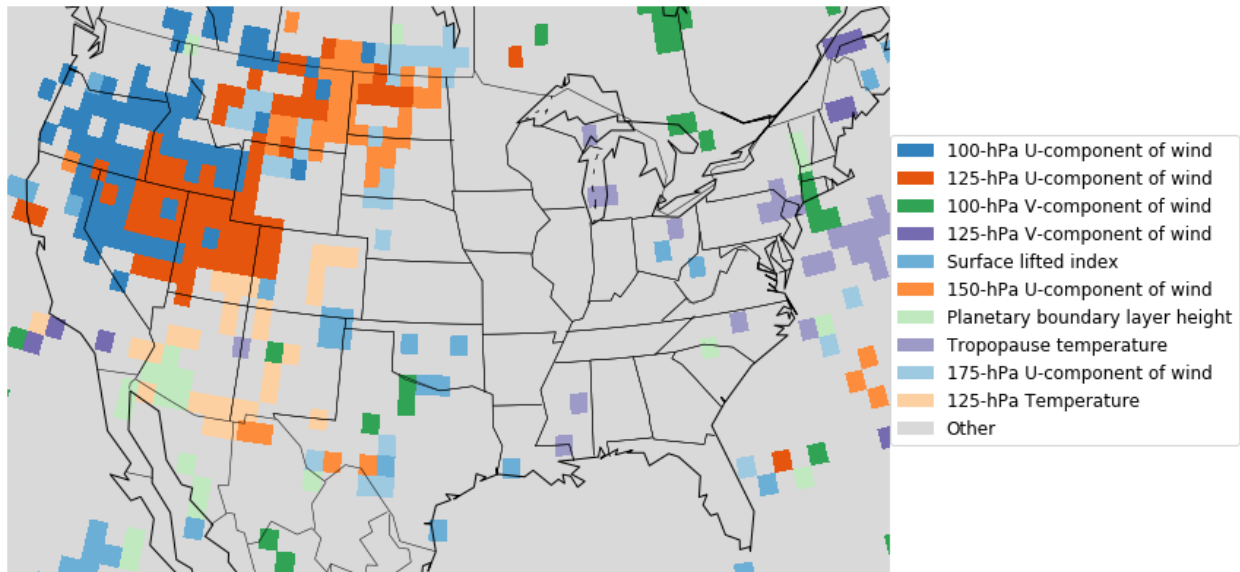
FIG. 7. The variable with highest feature importance using data from locations across the contiguous United States. The top ten variables by frequency are shown.
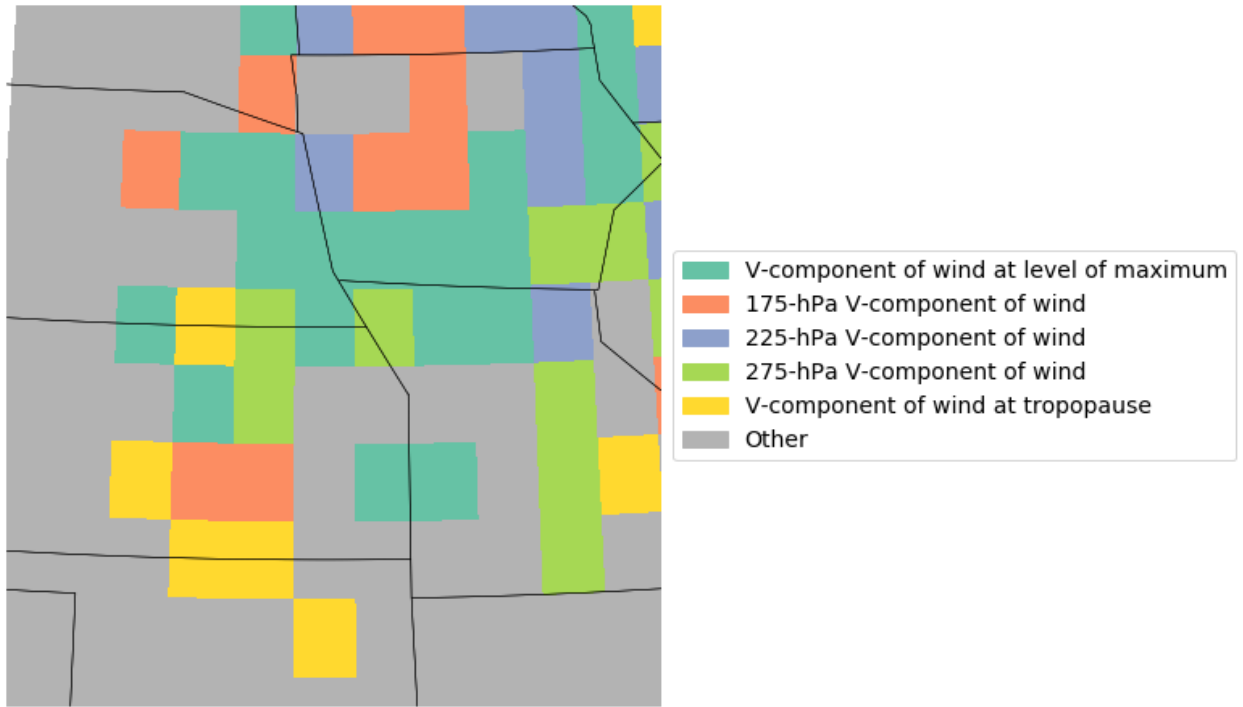
FIG. 8. The variable with highest feature importance using data from locations across across a region of the eastern Great Plains and Lower Midwest. The top five variables by frequency are shown.
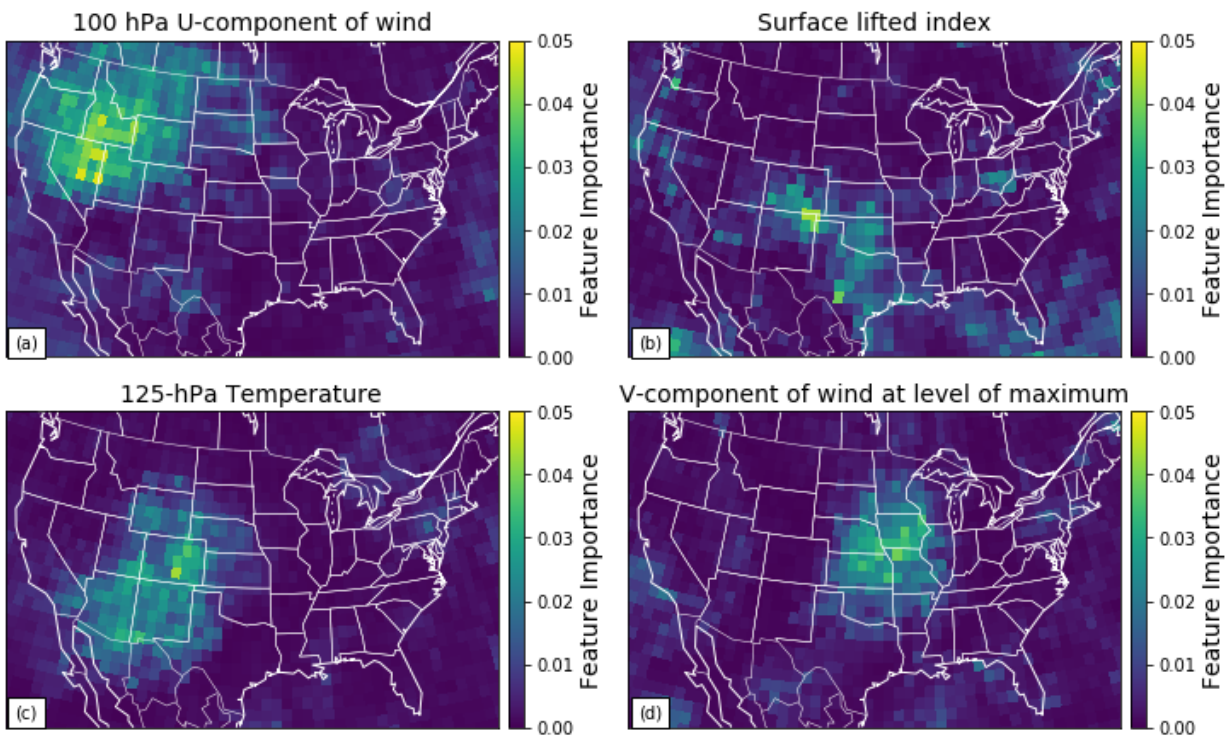
FIG. 9. Feature importance for (a) 100-hPa u-component of the wind, (b) surface lifted index, (c) 125-hPa temperature and (d) the v-component of the wind at the level of maximum. The importance for each location is calculated by using data from only that location.

# Chapter 5

# Conclusions

The primary goal of this thesis was to answer the following question. Are there large-scale differences between drylines that produce deep, moist convection and those that do not? Two different, but complementary approaches were used to achieve this objective. Firstly, a climatology of dryline convection was constructed to enable the creation of synoptic composite analyses of dryline events. The second approach built on the first, using the dryline convection dataset to answer two questions. Can machine learning be used to predict dryline convection? And what variables are best at determining whether dryline convection occurs? The primary conclusions of applying these two methods and potential avenues for further research are discussed below.

## 5.1   Conclusions and Future Work

### 5.1.1   Paper 1: A Synoptic Climatology of Dryline Convection in the Southern Great Plains

A dataset of drylines within a region of the southern Great Plains was constructed using ten years of WPC analyses for the months of April, May and June. Doppler radar, visible and infrared satellite imagery were used to identify convective drylines, where deep, moist convection was deemed to have been associated with the dryline circulation. Approximately 60% of drylines were convective, with initiation most frequently occurring between 2000 and 2100 UTC.

A relationship was found between the El Nino Southern Oscillation (ENSO) and the number of dryline days several months later. The Oceanic Nino Index

(ONI) is a running three-month average of sea surface temperature in the Nino 3.4 region compared to the climatological average. There was a negative correlation ($r$ = -0.635, $p = 0.049$) between the July-August-September ONI and the number of dryline days the following spring. As the time lag between the analyzed ONI and dryline season was reduced, the correlation became weaker. Conversely, there was a positive correlation between the the ONI in summer and early autumn and the proportion of convective drylines the following year. This correlation was strongest for June-July-August ($r = 0.810$, $p = 0.004$). These results were not deemed to fit with the overall theme of Paper 1, so were not included in the published article. However, future work could investigate these relationships further. For instance, ENSO is thought to influence the position and amplification of the upper-tropospheric jet stream. Does the position of the jet significantly affect the number of drylines in any given spring?

Using reanalysis data from 1800 UTC, synoptic composites were created of 179 convective and 104 non-convective (null) dryline days identified. Both composites feature an upper-level long-wave trough to the west of the Rockies and a ridge extending north through the east-north-central states. At the surface, the composites featured a broad surface cyclone over western Texas and southerly flow over the south-central states. However, the composites exhibit significant differences in fields of 500-hPa geopotential height, 700 and 800-hPa specific humidity, and CAPE. The convective composite is more moist at low-levels, and has higher CAPE than the null composite. These differences develop over several days preceding a dryline event. Convective drylines featured more amplified upper-level flow, associated with a deeper trough in the western US and a stronger downstream ridge than non-convective drylines in the three days preceding a dryline event. Greater poleward moisture transport occurs in stronger southerly flow, leading to higher values of low-level moisture and CAPE in the convective composite. The likely effect of higher CAPE is higher updraft velocity, meaning parcels are more likely to overcome the inversion, and deep, moist convection can sustain if they do so, because stronger updrafts are less affected by dry-air entrainment.

The fact that significant differences are observed in moisture and instability over the Great Plains three days before a dryline event emphasizes that a singular focus on the mesoscale or convective scale is not sufficient to predict dryline convection. However, some of the effects of synoptic-scale forcing were not observed

in the composites. Despite the fact that convective drylines were associated with stronger troughing and greater forcing for ascent, no significant differences were found in mid-level temperature over the dryline region. Greater ascent would also be expected to result in lower inhibition, but no significant differences were observed between the composites.

Perhaps the biggest unanswered question with regards to synoptic-scale processes is whether convective drylines have stronger convergence. Schultz et al. (2007) identified a relationship between short-wave troughing and dryline confluence. Given the increased troughing in the convective composite, it follows that the convective composites should have stronger dryline convergence. However, it was not possible to establish the strength of the dryline using the techniques employed when creating the composites. The width of the dryline circulation is usually smaller than the NARR is capable of resolving. Furthermore, both convergence and vertical motion fields are noisy, even when there is a well-defined dryline. Future work could investigate the relationship between dryline convergence and the likelihood of convection by identifying drylines using model or reanalysis data, rather than the WPC analyses used here. Although reanalysis data were obtained to create composites, differences in the dryline location between the surface analyses and NARR data meant that measuring dryline convergence was subject to large errors.

Another benefit of an alternative identification method is the potential to increase the size of the dryline dataset. The WPC surface analysis archive is incomplete before 2006. An objective method of dryline identification such as used by Hoch and Markowski (2005) could avoid the restriction of sample size. They created a 30-year climatology of drylines, using Barnes interpolation to account for variations in the density of surface observations. Recent advances in dryline identification, such as the use of machine learning (Clark et al. 2015) mean that using model analyses or reanalyses data to identify drylines are becoming an increasingly viable option. However, one of the difficulties of using an objective method to identify drylines is the inability to distinguish between the dryline and intersecting boundaries such as fronts and convective outflow (Clark et al. 2015). The ability to correctly label drylines justifies the choice to manually identify them, despite the fact that it is labour intensive.

Objective identification techniques are also a possibility for identifying dryline convection. The process of determining whether radar echoes exceed the required

threshold could be automated, saving time and enabling a larger dataset to be examined more quickly. An alternative approach would be to document lightning strikes. However, the dryline convection problem would have to be re-framed, because on occasion deep, moist convection can occur without lightning.

After identifying convective drylines, the variables for which synoptic composites were created relied upon current knowledge of deep, moist convection. Composites were produced for variables thought to be representative of the ingredients of deep, moist convection. However, the composites were not comprehensive. It is possible that variables that were not examined also differ between convective and non-convective drylines. This potential drawback was a motivating factor in trialing a new method of investigating dryline convection, machine learning.

## 5.1.2   Paper 2: Using Machine Learning to Predict Dryline Convection

The creation of the climatology provided an opportunity to use an alternative approach to identifying key variables. A machine learning algorithm was trained and tested using the newly created dryline dataset, and short-range model analyses. An important benefit of this approach was to be able to examine variables in an objective manner, and provide some measure of the importance of each variable in terms of how useful it is in distinguishing between convective and non-convective drylines.

A gradient boosting model was trained using RAP and RUC analyses to achieve two objectives. Firstly, to establish the skill of gradient boosting in predicting whether a dryline will produce deep, moist convection. Secondly, to identify the variables that are most important when making those predictions. The gradient boosting method had a high probability of detection of convective drylines (mean recall = 0.993). However, the high rate of detection occurred because the model labelled too many drylines as convective. As a result, the model was not good at predicting non-convective drylines, which resulted in relatively low precision (mean = 0.638) and accuracy (mean = 0.638). The modest values of precision and accuracy mean that the gradient boosting model is not viable as a method of directly forecasting dryline convection.

There are several potential reasons why gradient boosting did not have high forecast skill. The most likely explanations are that either the parameters used

were not optimal, the dataset was not large enough, or a combination of the two. Parameter tuning attempted to maximize the harmonic mean of precision and recall, but initially overfit the training data. The parameters were adjusted to reduce overfitting, but as a consequence reduced performance. Future work would benefit from using a larger sample of drylines to increase the signal-to-noise ratio, and provide more data for training. It could then be determined whether more training data would improve forecast skill, and what is the upper limit to the ability of gradient boosting.

Despite the modest accuracy of gradient boosting, examination of variable importance revealed large spatial variations, and produced many results that were consistent with the synoptic composites. The importance of moisture and instability peaked east of the dryline, although in the case of instability, CAPE was outperformed by another measure of instability, the lifted index. Over the larger, more coarse domain of the contiguous US, the u-component of the wind at 100-hPa and 125-hPa were the two variables that had highest feature importance at the greatest number of locations. Most of these locations were concentrated in a region over the central Rockies. The u-component of the wind in this region is likely important because the strength of flow across the Rockies regulates the strength of lee troughing, and hence dryline confluence. Whether a stronger dryline is more likely to produce deep, moist convection is yet to be confirmed.

The importance of the large-scale pattern was also demonstrated by the 150-hPa temperature and the v-component of the wind around the level of the upper-tropospheric jet. The former was the leading variable in a cluster of locations in the southwest US, the latter in the eastern Great Plains and lower Midwest. The temperature at 125-hPa is related to the height of the tropopause, a proxy for the strength of upper-level troughing. However, the importance of the v-component of the wind at high altitude indicates that not only is the strength of the jet important, but also the amplification of the flow.

It is important to note that feature importance is a measure of how useful a variable was when making predictions, but not necessarily how good a variable was in distinguishing between convective and non-convective drylines. This is an especially pertinent point given the large differences observed between recall and precision. The leading variables may only be good at detecting convective drylines, rather than distinguishing between them. However, the importance of the large-scale pattern is consistent with the results of the synoptic compositing, and

does not contradict any current knowledge of atmospheric processes. Nonetheless, the modest accuracy of gradient boosting results in lower confidence in the conclusions made than would otherwise be the case.

Although this thesis has identified key differences between drylines that produce deep, moist convection and those that do not, the question arises as to how these results can be used to help forecasters? Is there a way to identify in advance if a dryline will produce deep, moist convection? Two alternate paths are offered to try and answer this question. Firstly, the role of synoptic-scale forcing could be investigated by way of testing the sensitivity of dryline convection to the strength of an upper-level trough. This could be achieved by performing model simulations of a convective dryline. The second path would attempt to improve the machine learning performance by accounting for the lack of variable independence. Data could be reduced by principal component analysis. Removal of variables that contribute similar information to the model would increase the signal-to-noise ratio of the training data, and potentially increase model performance. The model performance might also be enhanced by the inclusion of data not typically produced by mesoscale weather models, such as soil moisture or vegetation. Following one, or both of these research paths is recommended to improve predictability of deep, moist convection along the dryline.

# Bibliography

AMS, 2016: Dryline - AMS Glossary. Accessed 13 June 2016, http://glossary.ametsoc.org/wiki/Dryline.

Anagnostou, E. N., 2004: A convective/stratiform precipitation classification algorithm for volume scanning weather radar observations. *Meteorological Applications*, **11**, 291–300, doi:10.1017/S1350482704001409.

Arnup, S. J. and M. J. Reeder, 2007: The diurnal and seasonal variation of the Northern Australian dryline. *Mon. Weather Rev.*, **135**, 2995–3008, doi: doi:10.1175/MWR3455.1.

Atkins, N. T., R. M. Wakimoto, and C. L. Ziegler, 1998: Observations of the finescale structure of a dryline during VORTEX 95. *Mon. Weather Rev.*, **126**, 525–550, doi:10.1175/1520-0493(1998)126⟨0525:OOTFSO⟩2.0.CO;2.

Bankert, R. L., 1994: Cloud classification of AVHRR imagery in maritime regions using a probabilistic neural network. *Journal of Applied Meteorology*, **33**, 909–918, doi:10.1175/1520-0450(1994)033⟨0909:CCOAII⟩2.0.CO;2.

Beebe, R. G., 1958: An instability line development as observed by the tornado research airplane. *J. Meteorol.*, **15**, 278–282.

Benjamin, S. G. and T. N. Carlson, 1986: Some effects of surface heating and topography on the regional severe storm environment. Part I: Three-dimensional simulations. *Mon. Weather Rev.*, **114**, 307–329, doi:doi:10.1175/1520-0493(1986)114⟨0307:SEOSHA⟩2.0.CO;2.

Benjamin, S. G., et al., 2004: An Hourly AssimilationForecast Cycle: The RUC. *Mon. Weather Rev.*, **132**, 495–518, doi:10.1175/1520-0493(2004)132⟨0495:AHACTR⟩2.0.CO;2.

Benjamin, S. G., et al., 2016: A North American Hourly Assimilation and Model Forecast Cycle: The Rapid Refresh. *Mon. Weather Rev.*, **144**, 1669–1694, doi: 10.1175/MWR-D-15-0242.1.

Billet, J., M. DeLisi, B. G. Smith, and C. Gates, 1997: Use of regression techniques to predict hail size and the probability of large hail. *Weather Forecast.*, **12**, 154–164, doi:10.1175/1520-0434(1997)012⟨0154:UORTTP⟩2.0.CO;2.

Blanchard, D. O., 1998: Assessing the vertical distribution of convective available potential energy. *Weather Forecast.*, **13**, 870–877, doi:doi:10.1175/ 1520-0434(1998)013⟨0870:ATVDOC⟩2.0.CO;2.

Bluestein, H. B., 1992: *Synoptic-Dynamic Meteorology in Midlatitudes: Volume 1, Principles of Kinematics and Dynamics.* Oxford University Press, 431 pp.

Bluestein, H. B. and S. S. Parker, 1993: Modes of isolated, severe convective storm formation along the dryline. *Mon. Weather Rev.*, **121**, 1354–1372, doi: doi:10.1175/1520-0493(1993)121⟨1354:MOISCS⟩2.0.CO;2.

Breiman, L., 2001: Random forests. *Mach. Learn.*, **45**, 5–32, doi:10.1023/A: 1010933404324.

Breiman, L., J. Friedman, R. Olshen, and C. Stone, 1984: *Classification and regression trees.* Wadsworth International Group, 368 pp.

Buban, M. S., C. L. Ziegler, E. R. Mansell, and Y. P. Richardson, 2012: Simulation of dryline misovortex dynamics and cumulus formation. *Mon. Weather Rev.*, **140**, 3525–3551, doi:10.1175/mwr-d-11-00189.1.

Buban, M. S., C. L. Ziegler, E. N. Rasmussen, and Y. P. Richardson, 2007: The dryline on 22 May 2002 during IHOP: Ground-radar and in situ data analyses of the dryline and boundary layer evolution. *Mon. Weather Rev.*, **135**, 2473– 2505, doi:doi:10.1175/MWR3453.1.

Campbell, P. C., B. Geerts, and P. T. Bergmaier, 2014: A dryline in Southeast Wyoming. Part I: Multiscale analysis using observations and modeling on 22 June 2010. *Mon. Weather Rev.*, **142**, 268–289, doi:doi:10.1175/ MWR-D-13-00049.1.

Caruana, R. and A. Niculescu-Mizil, 2006: An empirical comparison of supervised learning algorithms. *Proceedings of the 23rd international conference on Machine learning*, 161–168.

Clark, A. J., A. MacKenzie, A. McGovern, V. Lakshmanan, and R. A. Brown, 2015: An automated, multiparameter dryline identification algorithm. *Weather Forecast.*, **30**, 1781–1794, doi:10.1175/waf-d-15-0070.1.

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. *Weather Forecast.*, **28**, 746–761.

Colby, F. P., 1980: The role of convective instability in an Oklahoma squall line. *J. Atmos. Sci.*, **37**, 2113–2119, doi:doi:10.1175/1520-0469(1980)037⟨2113:TROCII⟩2.0.CO;2.

Colby, F. P., 1984: Convective inhibition as a predictor of convection during AVE-SESAME II. *Mon. Weather Rev.*, **112 (11)**, 2239–2252, doi:doi:10.1175/1520-0493(1984)112⟨2239:CIAAPO⟩2.0.CO;2.

Crawford, T. M. and H. B. Bluestein, 1997: Characteristics of dryline passage during COPS-91. *Mon. Weather Rev.*, **125**, 463–477, doi:10.1175/1520-0493(1997)125⟨0463:codpdc⟩2.0.co;2.

Crook, N. A. and M. W. Moncrieff, 1988: The effect of large-scale convergence on the generation and maintenance of deep moist convection. *J. Atmos. Sci.*, **45**, 3606–3624.

Demoz, B., et al., 2006: The dryline on 22 May 2002 during IHOP_2002: Convective-scale Measurements at the profiling site. *Mon. Weather Rev.*, **134**, 294–310, doi:10.1175/MWR3054.1.

Dial, G. L., J. P. Racy, and R. L. Thompson, 2010: Short-term convective mode evolution along synoptic boundaries. *Weather Forecast.*, **25**, 1430–1446, doi:10.1175/2010waf2222315.1.

Dodd, A. V. and A. V. Dodd, 1965: Dew point distribution in the contiguous United States. *Mon. Wea. Rev*, **93**, 113–122.

Doswell, C. A., 1987: The distinction between large-scale and mesoscale contri-
bution to severe convection: A case study example. *Weather Forecast.*, **2**, 3–16,
doi:10.1175/1520-0434(1987)002⟨0003:TDBLSA⟩2.0.CO;2.

Doswell, C. A. and L. F. Bosart, 2001: Extratropical synoptic-scale processes and
severe convection. *Meteorol. Monogr.*, **28**, 27–70, doi:10.1175/0065-9401-28.50.
27.

Duell, R. S. and M. S. Van Den Broeke, 2016: Climatology, synoptic conditions,
and misanalyses of Mississippi River Valley drylines. *Mon. Weather Rev.*, **144**,
927–943, doi:doi:10.1175/MWR-D-15-0108.1.

Dunn, L. B., 1991: Evaluation of vertical motion: Past, present, and fu-
ture. *Weather Forecast.*, **6**, 65–75, doi:doi:10.1175/1520-0434(1991)006⟨0065:
EOVMPP⟩2.0.CO;2.

Durran, D. R. and L. W. Snellman, 1987: The diagnosis of synoptic-scale vertical
motion in an operational environment. *Weather Forecast.*, **2**, 17–31, doi:doi:
10.1175/1520-0434(1987)002⟨0017:TDOSSV⟩2.0.CO;2.

Friedman, J. H., 2001: Greedy function approximation: A gradient boosting
machine. *Ann. Stat.*, 1189–1232.

Fujita, T. T., 1958: Structure and movement of a dry front. *Bull. Amer. Meteor.
Soc*, **39**, 574–582.

Gagne, D. J., A. McGovern, and J. Brotzge, 2009: Classification of convective
areas using decision trees. *Journal of Atmospheric and Oceanic Technology*, **26**,
1341–1353, doi:10.1175/2008JTECHA1205.1.

Gagne, D. J., A. McGovern, S. E. Haupt, R. A. Sobash, J. K. Williams, and
M. Xue, 2017: Storm-based probabilistic hail forecasting with machine learning
applied to convection-allowing ensembles. *Weather Forecast.*, **32**, 1819–1840,
doi:10.1175/WAF-D-17-0010.1.

Geerts, B., 2008: Dryline characteristics near Lubbock, Texas, based on radar
and West Texas Mesonet data for May 2005 and May 2006. *Weather Forecast.*,
**23**, 392–406, doi:10.1175/2007waf2007044.1.

Golden, J. H., 1980: SESAME News: Forecasting and research on severe storms in China: A summary of two seminars. *Bull. Am. Meteorol. Soc.*, **61**, 7.

Hane, C. E., H. B. Bluestein, T. M. Crawford, M. E. Baldwin, and R. M. Rabin, 1997: Severe thunderstorm development in relation to along-dryline variability: A case study. *Mon. Weather Rev.*, **125**, 231–251, doi:doi:10.1175/1520-0493(1997)125⟨0231:STDIRT⟩2.0.CO;2.

Hane, E. C., 2004: Quiescent and synoptically-active drylines: A comparison based upon case studies. *Meteorol. Atmos. Phys.*, **86**, 195–211, doi:10.1007/s00703-003-0026-y.

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. Am. Meteorol. Soc.*, **77**, 1169–1178, doi:doi: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 US Great Plains region. *J. Clim.*, **18**, 2132–2137, doi:10.1175/jcli3392.1.

Hoskins, B. J., I. Draghici, and H. C. Davies, 1978: A new look at the $\omega$-equation. *Q. J. R. Meteorol. Soc.*, **104**, 31–38, doi:10.1002/qj.49710443903.

Hunter, J. D., 2007: Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, **9 (3)**, 90–95, doi:10.1109/MCSE.2007.55.

Johnson, Z. F. and N. M. Hitchens, 2018: Effects of soil moisture on the longitudinal dryline position in the southern great plain. *Journal of Hydrometeorology*, **19**, 273–287, doi:10.1175/JHM-D-17-0091.1.

Jones, P. A. and P. R. Bannon, 2002: A mixed-layer model of the diurnal dryline. *J. Atmos. Sci.*, **59**, 2582–2593, doi:10.1175/1520-0469(2002)059⟨2582:amlmot⟩2.0.co;2.

Kitzmiller, D. H., W. E. McGovern, and R. E. Saffle, 1995: The WSR-88D severe weather potential algorithm. *Weather Forecast.*, **10**, 141–159, doi:10.1175/1520-0434(1995)010⟨0141:TWSWPA⟩2.0.CO;2.

Kotsiantis, S. B., I. Zaharakis, and P. Pintelas, 2007: Supervised machine learning: A review of classification techniques. *Emerg. Artif. Intell. Appl. Comput. Eng.*, **160**, 3–24.

Lakshmanan, V., A. Fritz, T. Smith, K. Hondl, and G. Stumpf, 2007: An automated technique to quality control radar reflectivity data. *Journal of Applied Meteorology and Climatology*, **46**, 288–305, doi:10.1175/JAM2460.1.

Lakshmanan, V., C. Karstens, J. Krause, and L. Tang, 2014: Quality control of weather radar data using polarimetric variables. *Journal of Atmospheric and Oceanic Technology*, **31**, 1234–1249, doi:10.1175/JTECH-D-13-00073.1.

Lanicci, J. and T. Warner, 1991: A synoptic climatology of the elevated mixed-layer inversion over the Southern Great Plains in Spring. Part I: Structure, dynamics, and seasonal evolution. *Weather Forecast.*, **6**, 181–197, doi:10.1175/1520-0434(1991)006⟨0181:ASCOTE⟩2.0.CO;2.

Lanicci, J. M., T. N. Carlson, and T. T. Warner, 1987: Sensitivity of the Great Plains severe-storm environment to soil-moisture distribution. *Mon. Weather Rev.*, **115**, 2660–2673, doi:10.1175/1520-0493(1987)115⟨2660:sotgps⟩2.0.co;2.

Lucas, C., E. J. Zipser, and M. A. LeMone, 1994a: Convective available potential energy in the environment of oceanic and continental Clouds: Correction and comments. *J. Atmos. Sci.*, **51**, 3829–3830, doi:doi:10.1175/1520-0469(1994)051⟨3829:CAPEIT⟩2.0.CO;2.

Lucas, C., E. J. Zipser, and M. A. Lemone, 1994b: Vertical velocity in oceanic convection off tropical Australia. *J. Atmos. Sci.*, **51**, 3183–3193, doi:doi:10.1175/1520-0469(1994)051⟨3183:VVIOCO⟩2.0.CO;2.

Ma, L., G. Zhang, and E. Lu, 2018: Using the gradient boosting decision tree to improve the delineation of hourly rain areas during the summer from advanced Himawari imager data. *Journal of Hydrometeorology*, **19**, 761–776, doi:10.1175/JHM-D-17-0109.1.

Manzato, A., 2013: Hail in Northeast Italy: A neural network ensemble forecast using sounding-derived indices. *Weather Forecast.*, **28**, 3–28, doi:10.1175/WAF-D-12-00034.1.

Markowski, P., C. Hannon, E. Rasmussen, P. Markowski, C. Hannon, and E. Rasmussen, 2006: Observations of convection initiation failure from the 12 June 2002 IHOP deployment. *Mon. Weather Rev.*, **134**, 375–405.

Marzban, C. and G. J. Stumpf, 1996: A neural network for tornado prediction based on Doppler radar-derived attributes. *Journal of Applied Meteorology*, **35**, 617–626, doi:10.1175/1520-0450(1996)035⟨0617:ANNFTP⟩2.0.CO;2.

Marzban, C. and G. J. Stumpf, 1998: A neural network for damaging wind prediction. *Weather Forecast.*, **13**, 151–163, doi:10.1175/1520-0434(1998)013⟨0151:ANNFDW⟩2.0.CO;2.

Marzban, C. and A. Witt, 2001: A Bayesian neural network for severe-hail size prediction. *Weather and Forecasting*, **16**, 600–610, doi:10.1175/1520-0434(2001)016⟨0600:ABNNFS⟩2.0.CO;2.

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, doi:doi:10.1175/1520-0469(1982)039⟨0225:TEOAOD⟩2.0.CO;2.

McCaul, E. W. and C. Cohen, 2002: The impact on simulated storm structure and intensity of variations in the mixed layer and moist layer depths. *Mon. Weather Rev.*, **130**, 1722–1748, doi:10.1175/1520-0493(2002)130⟨1722:tiosss⟩2.0.co;2.

McGovern, A., D. J. Gagne, J. Basara, T. M. Hamill, and D. Margolin, 2015: Solar energy prediction : An international contest to initiate interdisciplinary research on compelling meteorological problems. *Bulletin of the American Meteorological Society*, **96**, 1388–1393, doi:10.1175/BAMS-D-14-00006.1.

McGovern, A., R. Lagerquist, D. J. Gagne, G. E. Jergensen, K. L. Elmore, C. R. Homeyer, and T. Smith, 2019: Making the black box more transparent: Understanding the physical implications of machine learning. *Bulletin of the American Meteorological Society*, **100**, 2175–2199, doi:10.1175/BAMS-D-18-0195.1.

McGuire, E. L., 1962: The vertical structure of three dry lines as revealed by aircraft traverses. Natl. Sev. Storms Proj. Rep. 77, 11 pp. [Available from NCAR, P.O. Box 3000, Boulder, CO 80307.]

Mecikalski, J. R., J. K. Williams, C. P. Jewett, D. Ahijevych, A. LeRoy, and J. R. Walker, 2015: Probabilistic 0-1-h convective initiation nowcasts that combine geostationary satellite observations and numerical weather prediction model data. *Journal of Applied Meteorology and Climatology*, **54**, 1039–1059, doi:10.1175/JAMC-D-14-0129.1.

Mesinger, F., et al., 2006: North American regional reanalysis. *Bulletin of the American Meteorological Society*, **87**, 343–360, doi:10.1175/BAMS-87-3-343.

Mueller, C. K., J. W. Wilson, and N. A. Crook, 1993: The utility of sounding and mesonet data to nowcast thunderstorm initiation. *Weather Forecast.*, **8**, 132–146, doi:10.1175/1520-0434(1993)008⟨0132:TUOSAM⟩2.0.CO;2.

Murphey, H. V., R. M. Wakimoto, C. Flamant, and D. E. Kingsmill, 2006: Dryline on 19 June 2002 during IHOP. Part I: Airborne Doppler and LEANDRE II analyses of the thin line structure and convection initiation. *Mon. Weather Rev.*, **134**, 406–430, doi:10.1175/mwr3063.1.

Natekin, A. and A. Knoll, 2013: Gradient boosting machines, a tutorial. *Front. Neurorobot.*, **7**, 21, doi:10.3389/fnbot.2013.00021.

NCEI, 2016: Accessed 13 June 2016. Interactive Map Tool — National Centers for Environmental Information (NCEI) formerly known as National Climatic Data Center (NCDC). https://www.ncdc.noaa.gov/data-access/radar-data/radar-map-tool.

Parsons, D. B., et al., 1991: The finescale structure of a West Texas dryline. *Mon. Weather Rev.*, **119**, 1242–1258, doi:10.1175/1520-0493(1991)119⟨1242:TFSOAW⟩2.0.CO;2.

Peckham, S. E. and L. J. Wicker, 2000: The influence of topography and lower-tropospheric winds on dryline morphology. *Mon. Weather Rev.*, **128**, 2165–2189, doi:10.1175/1520-0493(2000)128⟨2165:tiotal⟩2.0.co;2.

Peckham, S. E., R. B. Wilhelmson, L. J. Wicker, and C. L. Ziegler, 2004: Numerical simulation of the interaction between the dryline and horizontal convective rolls. *Mon. Weather Rev.*, **132**, 1792–1812, doi:10.1175/1520-0493(2004)132⟨1792:nsotib⟩2.0.co;2.

Peterson, R. E., 1983: The west Texas dryline: Occurrence and behavior, Preprints,. *13th Conf. on Severe Local Storms*, Tulsa, OK, Amer. Meteor. Soc., J9–J11.

Pietrycha, A. E. and E. N. Rasmussen, 2004: Finescale surface observations of the dryline: A mobile mesonet perspective. *Weather Forecast.*, **19**, 1075–1088, doi:doi:10.1175/819.1.

Rhea, J. O., 1966: A study of thunderstorm formation along dry lines. *J. Appl. Meteorol.*, **5**, 58–63, doi:doi:10.1175/1520-0450(1966)005⟨0058:ASOTFA⟩2.0. CO;2.

Rogers, E., et al., 2009: The ncep north american mesoscale modeling system: Recent changes and future plans. *Preprints, 23rd Conference on Weather Analysis and Forecasting/19th Conference on Numerical Weather Prediction.*

Rumelhart, D. E. and J. L. McClelland, 1986: Learning internal representations by error propagation. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations*, MIT Press, 318–362.

Sanders, F. and B. J. Hoskins, 1990: An easy method for estimation of Q-vectors from weather maps. *Weather Forecast.*, **5**, 346–353, doi:doi:10.1175/1520-0434(1990)005⟨0346:AEMFEO⟩2.0.CO;2.

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

Schaefer, J. T., 1974: The Life Cycle of the Dryline. *J. Appl. Meteorol.*, **13**, 444–449, doi:10.1175/1520-0450(1974)013⟨0444:TLCOTD⟩2.0.CO;2.

Schaefer, J. T., 1986: The dryline. *Mesoscale Meteorol. Forecast.*, P. S. Ray, Ed., American Meteorological Society, 549–572.

Schaffer, C., 1993: Overfitting avoidance as bias. *Mach. Learn.*, **10**, 153–178, doi:10.1023/A:1022653209073.

Schultz, D. M., C. C. Weiss, and P. M. Hoffman, 2007: The synoptic regulation of dryline intensity. *Mon. Weather Rev.*, **135**, 1699–1709, doi:10.1175/mwr3376.1.

Schumann, M. R. and P. J. Roebber, 2010: The influence of upper-tropospheric potential vorticity on convective morphology. *Mon. Weather Rev.*, **138**, 463–474, doi:10.1175/2009mwr3091.1.

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, doi: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.

Svozil, D., V. Kvasnička, and J. Pospíchal, 1997: Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems*, Elsevier, Vol. 39, 43–62, doi:10.1016/S0169-7439(97)00061-0.

Trenberth, K. E., 1978: On the interpretation of the diagnostic quasi-geostrophic omega equation. *Mon. Weather Rev.*, **106**, 131–137, doi:doi:10.1175/1520-0493(1978)106⟨0131:OTIOTD⟩2.0.CO;2.

Trier, S. B., et al., 2015: Mesoscale thermodynamic influences on convection initiation near a surface dryline in a convection-permitting ensemble. *Mon. Weather Rev.*, **143**, 3726–3753, doi:10.1175/MWR-D-15-0133.1.

Wakimoto, R. M. and H. V. Murphey, 2009: Analysis of a dryline during IHOP: Implications for convection initiation. *Mon. Weather Rev.*, **137**, 912–936, doi:10.1175/2008mwr2584.1.

Weckwerth, T. M., J. W. Wilson, R. M. Wakimoto, and N. A. Crook, 1997: Horizontal convective rolls: Determining the environmental conditions supporting their existence and characteristics. *Mon. Weather Rev.*, **125**, 505–526, doi:doi:10.1175/1520-0493(1997)125⟨0505:HCRDTE⟩2.0.CO;2.

Weckwerth, T. M., et al., 2004: An overview of the International $H_2O$ Project (IHOP_2002) and some preliminary highlights. *Bull. Am. Meteorol. Soc.*, **85**, doi:10.1175/BAMS-85-2-253.

Weiss, C. C., H. B. Bluestein, and A. L. Pazmany, 2006: Finescale radar observations of the 22 May 2002 dryline during the International $H_2O$ Project (IHOP). *Mon. Weather Rev.*, **134**, 273–293, doi:10.1175/MWR3068.1.

Weldegaber, M. H., B. B. Demoz, L. C. Sparling, R. Hoff, and S. Chiao, 2011: Observational analysis of moisture evolution and variability in the boundary layer during the dryline on 22 May 2002. *Meteorol. Atmos. Phys.*, **110**, 87–102, doi:10.1007/s00703-010-0109-5.

Weston, K. J., 1972: The dry-line of Northern India and its role in cumulonimbus convection. *Q. J. R. Meteorol. Soc.*, **98**, 519–531, doi:10.1002/qj.49709841704.

Whiton, R. C., P. L. Smith, S. G. Bigler, K. E. Wilk, and A. C. Harbuck, 1998: History of operational use of weather radar by U.S. weather services. Part II:

Development of operational Doppler weather radars. *Weather Forecast.*, **13**, 244–252, doi:doi:10.1175/1520-0434(1998)013⟨0244:HOOUOW⟩2.0.CO;2.

Williams, J. K., 2014: Using random forests to diagnose aviation turbulence. *Machine learning*, **95**, 51–70.

Williams, J. K., D. Ahijevych, S. Dettling, and M. Steiner, 2008: Combining observations and model data for short-term storm forecasting. *Remote sensing applications for aviation weather hazard detection and decision support*, International Society for Optics and Photonics, Vol. 7088, 708805.

Xu, W., L. Ning, and Y. Luo, 2020: Wind Speed Forecast Based on Post-Processing of Numerical Weather Predictions Using a Gradient Boosting Decision Tree Algorithm. *Atmosphere*, **11**, 738, doi:10.3390/atmos11070738.

Xue, M. and W. J. Martin, 2006a: A high-resolution modeling study of the 24 May 2002 dryline case during IHOP. Part I: Numerical simulation and general evolution of the dryline and convection. *Mon. Weather Rev.*, **134**, 149–171, doi:10.1175/mwr3071.1.

Xue, M. and W. J. Martin, 2006b: A high-resolution modeling study of the 24 May 2002 dryline case during IHOP. Part II: Horizontal convective rolls and convective initiation. *Mon. Weather Rev.*, **134**, 172–191, doi:10.1175/mwr3072.1.

Ziegler, C. L. and C. E. Hane, 1993: An observational study of the dryline. *Mon. Weather Rev.*, **121**, 1134–1151, doi:10.1175/1520-0493(1993)121⟨1134:AOSOTD⟩2.0.CO;2.

Ziegler, C. L., T. J. Lee, and R. A. Pielke, 1997: Convective initiation at the dryline: A modeling study. *Mon. Weather Rev.*, **125**, 1001–1026, doi:10.1175/1520-0493(1997)125⟨1001:ciatda⟩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, doi:doi: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. *Weather Forecast.*, **13**, 1106–1131, doi:10.1175/1520-0434(1998)013⟨1106:tiomca⟩2.0.co;2.

# Appendix A

# RAP Model Output Inventory

Name, Level Type, Height

1, Geopotential Height, isobaricInhPa, 1000
2, Geopotential Height, isobaricInhPa, 975
3, Geopotential Height, isobaricInhPa, 950
4, Geopotential Height, isobaricInhPa, 925
5, Geopotential Height, isobaricInhPa, 900
6, Geopotential Height, isobaricInhPa, 875
7, Geopotential Height, isobaricInhPa, 850
8, Geopotential Height, isobaricInhPa, 825
9, Geopotential Height, isobaricInhPa, 800
10, Geopotential Height, isobaricInhPa, 775
11, Geopotential Height, isobaricInhPa, 750
12, Geopotential Height, isobaricInhPa, 725
13, Geopotential Height, isobaricInhPa, 700
14, Geopotential Height, isobaricInhPa, 675
15, Geopotential Height, isobaricInhPa, 650
16, Geopotential Height, isobaricInhPa, 625
17, Geopotential Height, isobaricInhPa, 600
18, Geopotential Height, isobaricInhPa, 575
19, Geopotential Height, isobaricInhPa, 550
20, Geopotential Height, isobaricInhPa, 525
21, Geopotential Height, isobaricInhPa, 500
22, Geopotential Height, isobaricInhPa, 475
23, Geopotential Height, isobaricInhPa, 450
24, Geopotential Height, isobaricInhPa, 425

```
25, Geopotential Height, isobaricInhPa, 400
26, Geopotential Height, isobaricInhPa, 375
27, Geopotential Height, isobaricInhPa, 350
28, Geopotential Height, isobaricInhPa, 325
29, Geopotential Height, isobaricInhPa, 300
30, Geopotential Height, isobaricInhPa, 275
31, Geopotential Height, isobaricInhPa, 250
32, Geopotential Height, isobaricInhPa, 225
33, Geopotential Height, isobaricInhPa, 200
34, Geopotential Height, isobaricInhPa, 175
35, Geopotential Height, isobaricInhPa, 150
36, Geopotential Height, isobaricInhPa, 125
37, Geopotential Height, isobaricInhPa, 100
38, Temperature, isobaricInhPa, 1000
39, Temperature, isobaricInhPa, 975
40, Temperature, isobaricInhPa, 950
41, Temperature, isobaricInhPa, 925
42, Temperature, isobaricInhPa, 900
43, Temperature, isobaricInhPa, 875
44, Temperature, isobaricInhPa, 850
45, Temperature, isobaricInhPa, 825
46, Temperature, isobaricInhPa, 800
47, Temperature, isobaricInhPa, 775
48, Temperature, isobaricInhPa, 750
49, Temperature, isobaricInhPa, 725
50, Temperature, isobaricInhPa, 700
51, Temperature, isobaricInhPa, 675
52, Temperature, isobaricInhPa, 650
53, Temperature, isobaricInhPa, 625
54, Temperature, isobaricInhPa, 600
55, Temperature, isobaricInhPa, 575
56, Temperature, isobaricInhPa, 550
57, Temperature, isobaricInhPa, 525
58, Temperature, isobaricInhPa, 500
59, Temperature, isobaricInhPa, 475
60, Temperature, isobaricInhPa, 450
61, Temperature, isobaricInhPa, 425
62, Temperature, isobaricInhPa, 400
```

```
63, Temperature, isobaricInhPa, 375
64, Temperature, isobaricInhPa, 350
65, Temperature, isobaricInhPa, 325
66, Temperature, isobaricInhPa, 300
67, Temperature, isobaricInhPa, 275
68, Temperature, isobaricInhPa, 250
69, Temperature, isobaricInhPa, 225
70, Temperature, isobaricInhPa, 200
71, Temperature, isobaricInhPa, 175
72, Temperature, isobaricInhPa, 150
73, Temperature, isobaricInhPa, 125
74, Temperature, isobaricInhPa, 100
75, Relative humidity, isobaricInhPa, 1000
76, Relative humidity, isobaricInhPa, 975
77, Relative humidity, isobaricInhPa, 950
78, Relative humidity, isobaricInhPa, 925
79, Relative humidity, isobaricInhPa, 900
80, Relative humidity, isobaricInhPa, 875
81, Relative humidity, isobaricInhPa, 850
82, Relative humidity, isobaricInhPa, 825
83, Relative humidity, isobaricInhPa, 800
84, Relative humidity, isobaricInhPa, 775
85, Relative humidity, isobaricInhPa, 750
86, Relative humidity, isobaricInhPa, 725
87, Relative humidity, isobaricInhPa, 700
88, Relative humidity, isobaricInhPa, 675
89, Relative humidity, isobaricInhPa, 650
90, Relative humidity, isobaricInhPa, 625
91, Relative humidity, isobaricInhPa, 600
92, Relative humidity, isobaricInhPa, 575
93, Relative humidity, isobaricInhPa, 550
94, Relative humidity, isobaricInhPa, 525
95, Relative humidity, isobaricInhPa, 500
96, Relative humidity, isobaricInhPa, 475
97, Relative humidity, isobaricInhPa, 450
98, Relative humidity, isobaricInhPa, 425
99, Relative humidity, isobaricInhPa, 400
100, Relative humidity, isobaricInhPa, 375
```

```
101, Relative humidity, isobaricInhPa, 350
102, Relative humidity, isobaricInhPa, 325
103, Relative humidity, isobaricInhPa, 300
104, Relative humidity, isobaricInhPa, 275
105, Relative humidity, isobaricInhPa, 250
106, Relative humidity, isobaricInhPa, 225
107, Relative humidity, isobaricInhPa, 200
108, Relative humidity, isobaricInhPa, 175
109, Relative humidity, isobaricInhPa, 150
110, Relative humidity, isobaricInhPa, 125
111, Relative humidity, isobaricInhPa, 100
112, U component of wind, isobaricInhPa, 1000
113, V component of wind, isobaricInhPa, 1000
114, U component of wind, isobaricInhPa, 975
115, V component of wind, isobaricInhPa, 975
116, U component of wind, isobaricInhPa, 950
117, V component of wind, isobaricInhPa, 950
118, U component of wind, isobaricInhPa, 925
119, V component of wind, isobaricInhPa, 925
120, U component of wind, isobaricInhPa, 900
121, V component of wind, isobaricInhPa, 900
122, U component of wind, isobaricInhPa, 875
123, V component of wind, isobaricInhPa, 875
124, U component of wind, isobaricInhPa, 850
125, V component of wind, isobaricInhPa, 850
126, U component of wind, isobaricInhPa, 825
127, V component of wind, isobaricInhPa, 825
128, U component of wind, isobaricInhPa, 800
129, V component of wind, isobaricInhPa, 800
130, U component of wind, isobaricInhPa, 775
131, V component of wind, isobaricInhPa, 775
132, U component of wind, isobaricInhPa, 750
133, V component of wind, isobaricInhPa, 750
134, U component of wind, isobaricInhPa, 725
135, V component of wind, isobaricInhPa, 725
136, U component of wind, isobaricInhPa, 700
137, V component of wind, isobaricInhPa, 700
138, U component of wind, isobaricInhPa, 675
```

```
139, V component of wind, isobaricInhPa, 675
140, U component of wind, isobaricInhPa, 650
141, V component of wind, isobaricInhPa, 650
142, U component of wind, isobaricInhPa, 625
143, V component of wind, isobaricInhPa, 625
144, U component of wind, isobaricInhPa, 600
145, V component of wind, isobaricInhPa, 600
146, U component of wind, isobaricInhPa, 575
147, V component of wind, isobaricInhPa, 575
148, U component of wind, isobaricInhPa, 550
149, V component of wind, isobaricInhPa, 550
150, U component of wind, isobaricInhPa, 525
151, V component of wind, isobaricInhPa, 525
152, U component of wind, isobaricInhPa, 500
153, V component of wind, isobaricInhPa, 500
154, U component of wind, isobaricInhPa, 475
155, V component of wind, isobaricInhPa, 475
156, U component of wind, isobaricInhPa, 450
157, V component of wind, isobaricInhPa, 450
158, U component of wind, isobaricInhPa, 425
159, V component of wind, isobaricInhPa, 425
160, U component of wind, isobaricInhPa, 400
161, V component of wind, isobaricInhPa, 400
162, U component of wind, isobaricInhPa, 375
163, V component of wind, isobaricInhPa, 375
164, U component of wind, isobaricInhPa, 350
165, V component of wind, isobaricInhPa, 350
166, U component of wind, isobaricInhPa, 325
167, V component of wind, isobaricInhPa, 325
168, U component of wind, isobaricInhPa, 300
169, V component of wind, isobaricInhPa, 300
170, U component of wind, isobaricInhPa, 275
171, V component of wind, isobaricInhPa, 275
172, U component of wind, isobaricInhPa, 250
173, V component of wind, isobaricInhPa, 250
174, U component of wind, isobaricInhPa, 225
175, V component of wind, isobaricInhPa, 225
176, U component of wind, isobaricInhPa, 200
```

```
177, V component of wind, isobaricInhPa, 200
178, U component of wind, isobaricInhPa, 175
179, V component of wind, isobaricInhPa, 175
180, U component of wind, isobaricInhPa, 150
181, V component of wind, isobaricInhPa, 150
182, U component of wind, isobaricInhPa, 125
183, V component of wind, isobaricInhPa, 125
184, U component of wind, isobaricInhPa, 100
185, V component of wind, isobaricInhPa, 100
186, Vertical velocity, isobaricInhPa, 1000
187, Vertical velocity, isobaricInhPa, 975
188, Vertical velocity, isobaricInhPa, 950
189, Vertical velocity, isobaricInhPa, 925
190, Vertical velocity, isobaricInhPa, 900
191, Vertical velocity, isobaricInhPa, 875
192, Vertical velocity, isobaricInhPa, 850
193, Vertical velocity, isobaricInhPa, 825
194, Vertical velocity, isobaricInhPa, 800
195, Vertical velocity, isobaricInhPa, 775
196, Vertical velocity, isobaricInhPa, 750
197, Vertical velocity, isobaricInhPa, 725
198, Vertical velocity, isobaricInhPa, 700
199, Vertical velocity, isobaricInhPa, 675
200, Vertical velocity, isobaricInhPa, 650
201, Vertical velocity, isobaricInhPa, 625
202, Vertical velocity, isobaricInhPa, 600
203, Vertical velocity, isobaricInhPa, 575
204, Vertical velocity, isobaricInhPa, 550
205, Vertical velocity, isobaricInhPa, 525
206, Vertical velocity, isobaricInhPa, 500
207, Vertical velocity, isobaricInhPa, 475
208, Vertical velocity, isobaricInhPa, 450
209, Vertical velocity, isobaricInhPa, 425
210, Vertical velocity, isobaricInhPa, 400
211, Vertical velocity, isobaricInhPa, 375
212, Vertical velocity, isobaricInhPa, 350
213, Vertical velocity, isobaricInhPa, 325
214, Vertical velocity, isobaricInhPa, 300
```

215, Vertical velocity, isobaricInhPa, 275
216, Vertical velocity, isobaricInhPa, 250
217, Vertical velocity, isobaricInhPa, 225
218, Vertical velocity, isobaricInhPa, 200
219, Vertical velocity, isobaricInhPa, 175
220, Vertical velocity, isobaricInhPa, 150
221, Vertical velocity, isobaricInhPa, 125
222, Vertical velocity, isobaricInhPa, 100
223, MSLP MAPS System Reduction, meanSea, 0
224, Surface pressure, surface, 0
225, Pressure tendency, surface, 0
226, Potential temperature, heightAboveGround, 2
227, 2 metre dewpoint temperature, heightAboveGround, 2
228, Dew point depression or deficit, heightAboveGround, 2
229, 10 metre U wind component, heightAboveGround, 10
230, 10 metre V wind component, heightAboveGround, 10
231, 2 metre temperature, heightAboveGround, 2
232, Specific humidity, heightAboveGround, 2
233, Pseudo-adiabatic potential temperature, surface, 0
234, Convective available potential energy, surface, 0
235, Convective inhibition, surface, 0
236, Surface lifted index, surface, 0
237, Best lifted index to 500 hPa, surface, 0
238, Large scale precipitation non-convective, surface, 0
239, Convective precipitation water, surface, 0
240, Categorical rain, surface, 0
241, Categorical freezing rain, surface, 0
242, Categorical ice pellets, surface, 0
243, Categorical snow, surface, 0
244, Water equivalent of accumulated snow depth, surface, 0
245, Storm relative helicity, surface, 0
246, Pressure, isothermZero, 0
247, Geopotential Height, isothermZero, 0
248, Relative humidity, isothermZero, 0
249, Pressure, tropopause, 0
250, Potential temperature, tropopause, 0
251, U component of wind, tropopause, 0
252, V component of wind, tropopause, 0

```
253, Pressure, maxWind, 0
254, U component of wind, maxWind, 0
255, V component of wind, maxWind, 0
256, Precipitation rate, surface, 0
257, Moisture availability, surface, 0
258, Storm surface runoff, surface, 0
259, Temperature, pressureFromGroundLayer, 3000
260, Relative humidity, pressureFromGroundLayer, 3000
261, U component of wind, pressureFromGroundLayer, 3000
262, V component of wind, pressureFromGroundLayer, 3000
263, Vertical velocity, pressureFromGroundLayer, 3000
264, Temperature, pressureFromGroundLayer, 6000
265, Relative humidity, pressureFromGroundLayer, 6000
266, U component of wind, pressureFromGroundLayer, 6000
267, V component of wind, pressureFromGroundLayer, 6000
268, Vertical velocity, pressureFromGroundLayer, 6000
269, Temperature, pressureFromGroundLayer, 9000
270, Relative humidity, pressureFromGroundLayer, 9000
271, U component of wind, pressureFromGroundLayer, 9000
272, V component of wind, pressureFromGroundLayer, 9000
273, Vertical velocity, pressureFromGroundLayer, 9000
274, Temperature, pressureFromGroundLayer, 12000
275, Relative humidity, pressureFromGroundLayer, 12000
276, U component of wind, pressureFromGroundLayer, 12000
277, V component of wind, pressureFromGroundLayer, 12000
278, Vertical velocity, pressureFromGroundLayer, 12000
279, Temperature, pressureFromGroundLayer, 15000
280, Relative humidity, pressureFromGroundLayer, 15000
281, U component of wind, pressureFromGroundLayer, 15000
282, V component of wind, pressureFromGroundLayer, 15000
283, Vertical velocity, pressureFromGroundLayer, 15000
284, Temperature, pressureFromGroundLayer, 18000
285, Relative humidity, pressureFromGroundLayer, 18000
286, U component of wind, pressureFromGroundLayer, 18000
287, V component of wind, pressureFromGroundLayer, 18000
288, Vertical velocity, pressureFromGroundLayer, 18000
289, Storm relative helicity, heightAboveGroundLayer, 1000
290, Orography, surface, 0
```

291, Pressure of highest freezing level, unknown, 0

292, Height of highest freezing level, unknown, 0

293, Relative humidity, unknown, 0

294, Snow depth, surface, 0

295, Precipitable water, entireAtmosphere, 0

296, 2 metre relative humidity, heightAboveGround, 2

297, Baseflow-groundwater runoff, depthBelowLand, 3

298, U-component storm motion, surface, 0

299, V-component storm motion, surface, 0

300, Planetary boundary layer height, surface, 0

301, Wind speed gust, surface, 0

302, Geometrical height, cloudBase, 0

303, Geometrical height, cloudTop, 0

304, Visibility, surface, 0

305, Maximum equivalent potential temperature level, unknown, 0

306, Height of convective cloud top, unknown, 0

307, Height of equilibrium level, unknown, 0

308, Temperature, tropopause, 0

309, Convective available potential energy, pressureFromGroundLayer, 25500

310, Convective inhibition, pressureFromGroundLayer, 25500

311, Temperature, surface, 0

312, Absolute vorticity, isobaricInhPa, 500

313, Geopotential Height, unknown, 0

314, Maximum/Composite radar reflectivity, entireAtmosphere, 0

315, Derived radar reflectivity, heightAboveGround, 1000

316, Derived radar reflectivity, heightAboveGround, 4000

317, Image data, entireAtmosphere, 0

# Appendix B

# RAP Model Input to Gradient Boosting Model

```
Name, Level Type, Height

Geopotential Height isobaricInhPa 1000
Geopotential Height isobaricInhPa 975
Geopotential Height isobaricInhPa 950
Geopotential Height isobaricInhPa 925
Geopotential Height isobaricInhPa 900
Geopotential Height isobaricInhPa 875
Geopotential Height isobaricInhPa 850
Geopotential Height isobaricInhPa 825
Geopotential Height isobaricInhPa 800
Geopotential Height isobaricInhPa 775
Geopotential Height isobaricInhPa 750
Geopotential Height isobaricInhPa 725
Geopotential Height isobaricInhPa 700
Geopotential Height isobaricInhPa 675
Geopotential Height isobaricInhPa 650
Geopotential Height isobaricInhPa 625
Geopotential Height isobaricInhPa 600
Geopotential Height isobaricInhPa 575
Geopotential Height isobaricInhPa 550
Geopotential Height isobaricInhPa 525
Geopotential Height isobaricInhPa 500
```

```
Geopotential Height isobaricInhPa 475
Geopotential Height isobaricInhPa 450
Geopotential Height isobaricInhPa 425
Geopotential Height isobaricInhPa 400
Geopotential Height isobaricInhPa 375
Geopotential Height isobaricInhPa 350
Geopotential Height isobaricInhPa 325
Geopotential Height isobaricInhPa 300
Geopotential Height isobaricInhPa 275
Geopotential Height isobaricInhPa 250
Geopotential Height isobaricInhPa 225
Geopotential Height isobaricInhPa 200
Geopotential Height isobaricInhPa 175
Geopotential Height isobaricInhPa 150
Geopotential Height isobaricInhPa 125
Geopotential Height isobaricInhPa 100
Temperature isobaricInhPa 1000
Temperature isobaricInhPa 975
Temperature isobaricInhPa 950
Temperature isobaricInhPa 925
Temperature isobaricInhPa 900
Temperature isobaricInhPa 875
Temperature isobaricInhPa 850
Temperature isobaricInhPa 825
Temperature isobaricInhPa 800
Temperature isobaricInhPa 775
Temperature isobaricInhPa 750
Temperature isobaricInhPa 725
Temperature isobaricInhPa 700
Temperature isobaricInhPa 675
Temperature isobaricInhPa 650
Temperature isobaricInhPa 625
Temperature isobaricInhPa 600
Temperature isobaricInhPa 575
Temperature isobaricInhPa 550
Temperature isobaricInhPa 525
Temperature isobaricInhPa 500
Temperature isobaricInhPa 475
```

```
Temperature isobaricInhPa 450
Temperature isobaricInhPa 425
Temperature isobaricInhPa 400
Temperature isobaricInhPa 375
Temperature isobaricInhPa 350
Temperature isobaricInhPa 325
Temperature isobaricInhPa 300
Temperature isobaricInhPa 275
Temperature isobaricInhPa 250
Temperature isobaricInhPa 225
Temperature isobaricInhPa 200
Temperature isobaricInhPa 175
Temperature isobaricInhPa 150
Temperature isobaricInhPa 125
Temperature isobaricInhPa 100
Relative humidity isobaricInhPa 1000
Relative humidity isobaricInhPa 975
Relative humidity isobaricInhPa 950
Relative humidity isobaricInhPa 925
Relative humidity isobaricInhPa 900
Relative humidity isobaricInhPa 875
Relative humidity isobaricInhPa 850
Relative humidity isobaricInhPa 825
Relative humidity isobaricInhPa 800
Relative humidity isobaricInhPa 775
Relative humidity isobaricInhPa 750
Relative humidity isobaricInhPa 725
Relative humidity isobaricInhPa 700
Relative humidity isobaricInhPa 675
Relative humidity isobaricInhPa 650
Relative humidity isobaricInhPa 625
Relative humidity isobaricInhPa 600
Relative humidity isobaricInhPa 575
Relative humidity isobaricInhPa 550
Relative humidity isobaricInhPa 525
Relative humidity isobaricInhPa 500
Relative humidity isobaricInhPa 475
Relative humidity isobaricInhPa 450
```

```
Relative humidity isobaricInhPa 425
Relative humidity isobaricInhPa 400
Relative humidity isobaricInhPa 375
Relative humidity isobaricInhPa 350
Relative humidity isobaricInhPa 325
Relative humidity isobaricInhPa 300
Relative humidity isobaricInhPa 275
Relative humidity isobaricInhPa 250
Relative humidity isobaricInhPa 225
Relative humidity isobaricInhPa 200
Relative humidity isobaricInhPa 175
Relative humidity isobaricInhPa 150
Relative humidity isobaricInhPa 125
Relative humidity isobaricInhPa 100
U component of wind isobaricInhPa 1000
U component of wind isobaricInhPa 975
U component of wind isobaricInhPa 950
U component of wind isobaricInhPa 925
U component of wind isobaricInhPa 900
U component of wind isobaricInhPa 875
U component of wind isobaricInhPa 850
U component of wind isobaricInhPa 825
U component of wind isobaricInhPa 800
U component of wind isobaricInhPa 775
U component of wind isobaricInhPa 750
U component of wind isobaricInhPa 725
U component of wind isobaricInhPa 700
U component of wind isobaricInhPa 675
U component of wind isobaricInhPa 650
U component of wind isobaricInhPa 625
U component of wind isobaricInhPa 600
U component of wind isobaricInhPa 575
U component of wind isobaricInhPa 550
U component of wind isobaricInhPa 525
U component of wind isobaricInhPa 500
U component of wind isobaricInhPa 475
U component of wind isobaricInhPa 450
U component of wind isobaricInhPa 425
```

```
U component of wind isobaricInhPa 400
U component of wind isobaricInhPa 375
U component of wind isobaricInhPa 350
U component of wind isobaricInhPa 325
U component of wind isobaricInhPa 300
U component of wind isobaricInhPa 275
U component of wind isobaricInhPa 250
U component of wind isobaricInhPa 225
U component of wind isobaricInhPa 200
U component of wind isobaricInhPa 175
U component of wind isobaricInhPa 150
U component of wind isobaricInhPa 125
U component of wind isobaricInhPa 100
V component of wind isobaricInhPa 1000
V component of wind isobaricInhPa 975
V component of wind isobaricInhPa 950
V component of wind isobaricInhPa 925
V component of wind isobaricInhPa 900
V component of wind isobaricInhPa 875
V component of wind isobaricInhPa 850
V component of wind isobaricInhPa 825
V component of wind isobaricInhPa 800
V component of wind isobaricInhPa 775
V component of wind isobaricInhPa 750
V component of wind isobaricInhPa 725
V component of wind isobaricInhPa 700
V component of wind isobaricInhPa 675
V component of wind isobaricInhPa 650
V component of wind isobaricInhPa 625
V component of wind isobaricInhPa 600
V component of wind isobaricInhPa 575
V component of wind isobaricInhPa 550
V component of wind isobaricInhPa 525
V component of wind isobaricInhPa 500
V component of wind isobaricInhPa 475
V component of wind isobaricInhPa 450
V component of wind isobaricInhPa 425
V component of wind isobaricInhPa 400
```

```
V component of wind isobaricInhPa 375
V component of wind isobaricInhPa 350
V component of wind isobaricInhPa 325
V component of wind isobaricInhPa 300
V component of wind isobaricInhPa 275
V component of wind isobaricInhPa 250
V component of wind isobaricInhPa 225
V component of wind isobaricInhPa 200
V component of wind isobaricInhPa 175
V component of wind isobaricInhPa 150
V component of wind isobaricInhPa 125
V component of wind isobaricInhPa 100
Vertical velocity isobaricInhPa 1000
Vertical velocity isobaricInhPa 975
Vertical velocity isobaricInhPa 950
Vertical velocity isobaricInhPa 925
Vertical velocity isobaricInhPa 900
Vertical velocity isobaricInhPa 875
Vertical velocity isobaricInhPa 850
Vertical velocity isobaricInhPa 825
Vertical velocity isobaricInhPa 800
Vertical velocity isobaricInhPa 775
Vertical velocity isobaricInhPa 750
Vertical velocity isobaricInhPa 725
Vertical velocity isobaricInhPa 700
Vertical velocity isobaricInhPa 675
Vertical velocity isobaricInhPa 650
Vertical velocity isobaricInhPa 625
Vertical velocity isobaricInhPa 600
Vertical velocity isobaricInhPa 575
Vertical velocity isobaricInhPa 550
Vertical velocity isobaricInhPa 525
Vertical velocity isobaricInhPa 500
Vertical velocity isobaricInhPa 475
Vertical velocity isobaricInhPa 450
Vertical velocity isobaricInhPa 425
Vertical velocity isobaricInhPa 400
Vertical velocity isobaricInhPa 375
```

```
Vertical velocity isobaricInhPa 350
Vertical velocity isobaricInhPa 325
Vertical velocity isobaricInhPa 300
Vertical velocity isobaricInhPa 275
Vertical velocity isobaricInhPa 250
Vertical velocity isobaricInhPa 225
Vertical velocity isobaricInhPa 200
Vertical velocity isobaricInhPa 175
Vertical velocity isobaricInhPa 150
Vertical velocity isobaricInhPa 125
Vertical velocity isobaricInhPa 100
MSLP (MAPS System Reduction) meanSea 0
Surface pressure surface 0
Pressure tendency surface 0
Potential temperature heightAboveGround 2
2 metre dewpoint temperature heightAboveGround 2
Dew point depression (or deficit) heightAboveG...
10 metre U wind component heightAboveGround 10
10 metre V wind component heightAboveGround 10
2 metre temperature heightAboveGround 2
Specific humidity heightAboveGround 2
Pseudo-adiabatic potential temperature surface 0
Convective available potential energy surface 0
Convective inhibition surface 0
Surface lifted index surface 0
Large scale precipitation (non-convective) sur...
Convective precipitation (water) surface 0
Categorical rain surface 0
Categorical freezing rain surface 0
Categorical ice pellets surface 0
Categorical snow surface 0
Storm relative helicity heightAboveGroundLayer 30
Pressure isothermZero 0
Geopotential Height isothermZero 0
Relative humidity isothermZero 0
Pressure tropopause 0
Potential temperature tropopause 0
U component of wind tropopause 0
```

```
V component of wind tropopause 0
Pressure maxWind 0
U component of wind maxWind 0
V component of wind maxWind 0
Precipitation rate surface 0
Moisture availability surface 0
Storm surface runoff surface 0
Temperature pressureFromGroundLayer 30
Relative humidity pressureFromGroundLayer 30
U component of wind pressureFromGroundLayer 30
V component of wind pressureFromGroundLayer 30
Vertical velocity pressureFromGroundLayer 30
Temperature pressureFromGroundLayer 60
Relative humidity pressureFromGroundLayer 60
U component of wind pressureFromGroundLayer 60
V component of wind pressureFromGroundLayer 60
Vertical velocity pressureFromGroundLayer 60
Temperature pressureFromGroundLayer 90
Relative humidity pressureFromGroundLayer 90
U component of wind pressureFromGroundLayer 90
V component of wind pressureFromGroundLayer 90
Vertical velocity pressureFromGroundLayer 90
Temperature pressureFromGroundLayer 12
Relative humidity pressureFromGroundLayer 12
U component of wind pressureFromGroundLayer 12
V component of wind pressureFromGroundLayer 12
Vertical velocity pressureFromGroundLayer 12
Temperature pressureFromGroundLayer 15
Relative humidity pressureFromGroundLayer 15
U component of wind pressureFromGroundLayer 15
V component of wind pressureFromGroundLayer 15
Vertical velocity pressureFromGroundLayer 15
Temperature pressureFromGroundLayer 18
Relative humidity pressureFromGroundLayer 18
U component of wind pressureFromGroundLayer 18
V component of wind pressureFromGroundLayer 18
Vertical velocity pressureFromGroundLayer 18
Storm relative helicity heightAboveGroundLayer 10
```

```
Orography surface 0
Pressure of highest freezing level unknown 0
Height of highest freezing level unknown 0
Relative humidity unknown 0
Snow depth surface 0
Precipitable water entireAtmosphere 0
2 metre relative humidity heightAboveGround 2
U-component storm motion surface 0
V-component storm motion surface 0
Planetary boundary layer height surface 0
Wind speed (gust) surface 0
Visibility surface 0
Height of convective cloud top unknown 0
Height of equilibrium level unknown 0
Temperature tropopause 0
Convective available potential energy pressure...
Convective inhibition pressureFromGroundLayer ...
Temperature surface 0
```