UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

AN AUTOMATED, MULTI-PARAMETER DRYLINE IDENTIFICATION AND TRACKING ALGORITHM

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

ANDREW MACKENZIE
Norman, Oklahoma
2013

AN AUTOMATED, MULTI-PARAMETER DRYLINE IDENTIFICATION AND TRACKING ALGORITHM

A THESIS APPROVED FOR THE GRADUATE COLLEGE

BY

Dr. Amy McGovern, Chair

Dr. Valliappa Lakshmanan

Dr. Rodger Brown

Dr. Adam Clark

Dr. Andrew Fagg

Dr. T.H. Lee Williams

Acknowledgments

The contents of this thesis could not have been possible without significant help from my friends, colleagues, and family. I would like to first thank my advisors, Amy McGovern, Valliappa Lakshmanan, Rodger Brown, and Adam Clark. Amy and Lak have taught me a great deal, both in class and out, in addition to introducing me to a plethora of new opportunities. I have also learned a great deal from Rodger and Adam, and value their willingness to constantly contribute new ideas to this project. You have all given me a great deal of support and I cannot thank you enough. Special thanks also go to Lee Williams and Andrew Fagg for aiding in the formation of the interdisciplinary committee. This could not have happened without their participation.

I would also like to thank the National Severe Storms Laboratory, the Center for the Analysis and Prediction of Storms, and Brice Coffer, Lindsay Maudlin, Peter Veals, and Adam Clark for the use of their data.

In addition, my friends and fellow graduate students have aided me a great deal over the course of this project through contributions to my research and my sanity. In particular, I would like to thank David John Gagne, Tim Supinie, and Greg Blumberg for their great help.

Finally, I would like to thank my parents, Douglas and Rebeca MacKenzie, for their continued interest, pride, and enthusiasm and my girlfriend, Autumn, for her never ending support as well as keeping me from going too crazy.

This study was funded by NSF Grant IIS 0746816 and NOAA-OU Cooperative Agreement NA11OAR4320072.

Table of Contents

List Of Tables

List Of Figures

Abstract

An automated, multi-parameter pattern recognition algorithm is devised to locate drylines present in numerical weather prediction (NWP) model output. As model resolution and NWP ensemble sizes increase, the amount of information available increases as well. However, the identification of complex meteorological features is still performed primarily by hand. The time involved in these analyses places a limit on the amount of data that can be examined manually. Creating a system in which automated, objective analysis is possible allows the scale of various studies to be greatly expanded by increasing the number of parameter modifications attempted, cases studied, or days examined, to name only a few. The algorithm outlined in this research is designed to detect drylines, a feature common to the US High Plains (Hoch and Markowski, 2005). Drylines are a boundary between two air masses, one characterized by relatively moist air, the other by relatively dry air. This boundary is often an important feature in the initiation of storms, including those that develop into tornadic supercells (e.g., Ziegler and Rasmussen, 1998; McCarthy and Koch, 1982). The identification of drylines is achieved through the examination of multiple parameters, some of which are, to an extent, shared with similar meteorological features (e.g., wind shift along a cold front)(Bluestein, 1993; Schaefer, 1986), making objective identification non-trivial. The fully automated algorithm outlined in this thesis makes use of image processing and pattern recognition techniques adapted for spatial grids to identify drylines present in a single timestep of NWP model output.

Chapter 1

Introduction

Drylines are boundaries representing the intersection of two continental-scale air masses, one dry, the other moist. Common to the US High Plains, these meteorological features are capable of supporting the development of convection both along and ahead of the boundary (Rhea, 1966). The resulting storms are frequently strong, many with the capability of producing tornadoes, intense straight line winds, severe hail, and torrential rain (e.g., Rhea, 1966; Hane et al., 1993; Bluestein and Parker, 1993). With their potential for severe weather, a strong diurnal dependence, and a relative ambiguity in their definition, drylines are both a fascinating and important part of the atmosphere. A key example of their significance in forecasting can be seen in their impact on the 3 May 1999 tornado outbreak (Roebber et al., 2002; NWS, 1999). On this day two drylines located from northwest Texas through western Oklahoma into central Kansas aided in the initiation of 10 supercell thunderstorms that resulted in a total of 66 tornadoes.

Weather forecasts are often performed with the assistance of numerical weather prediction (NWP) models, which provide an estimate of the atmospheric state at a given time. Advances in computer technology have made possible great strides in NWP. Two developments that have resulted from these improvements are the use of higher resolution models and larger ensembles (Michalakes and Vachharajani, 2008), representing a tremendous increase in

the amount of data output. However, the speed at which traditional hand analyses can currently be performed places a limit on the amount of data that can be processed by an individual in a set amount of time. Combined with increasing data availability, this limitation suggests a need for automation. The goal of this project is not to remove subjective analysis, but provide the means to improve both the speed and accuracy of traditional analysis.

The algorithm outlined in this thesis is designed to contribute to multiple aspects of meteorology. First, case and climatological studies can be expanded in scope as the amount of time necessary to perform an initial evaluation of the data becomes drastically reduced. Second, advanced automated analysis of complex meteorological features would aid in model verification. It is relatively simple to discern displacement errors for binary values, such as rainfall. However, more complex features such as drylines, cold fronts, and cold pools have no single parameter that distinguishes them. Instead, many individual parameters must be combined to discriminate between these features and discern their locations. Finally, automation provides a significant benefit to forecasters, particularly when using longer range forecast guidance in which the presence and location of drylines is likely to vary from run to run as well as across individual ensemble members.

A wide variety of objective feature identification techniques exist, many of which focus on observational data (e.g., tornado detection, hail detection; Mitchell et al., 1998, Witt et al., 1998). Of those that are designed for use with NWP output, a large number focus on storm identification and tracking. For example, the Thunderstorm Identification, Tracking, Analysis, and Nowcasting (TITAN) (Dixon and Wiener, 1993) methodology, the Storm Cell Identification

and Tracking (SCIT) algorithm (Johnson et al., 1998), and w2segmotion (Lakshmanan and Smith, 2009; Lakshmanan et al., 2009; Lakshmanan and Smith, 2010; Lakshmanan et al., 2003), utilized by the Warning Decision Support System - Integrated Information (WDSS-II) system all provide automated detection and tracking of storms based on their respective criteria using object-based techniques and have been applied in a variety of ways (see e.g., Goudenhoofdt and Delobbe, 2012; Lakshmanan and Smith, 2009). However, at this point there are no object-based algorithms designed to objectively identify drylines present in NWP model output.

To complete this project there are two key difficulties that need to be overcome. The first is distinguishing drylines from other types of boundaries with similar characteristics. Drylines have some of the same properties as cold fronts, warm fronts, gust fronts, etc. The second is the need for objective criteria. In order to automate the analysis process, non-subjective measures that are effective for a variety of different cases are required.

The following chapter provides an overview of drylines, including a discussion of current subjective identification methods, in addition to a brief description of the computational techniques applied during the course of this research. Chapter 3 gives a brief discussion of the datasets used in the study and is followed by an overview of the developed algorithm in chapter 4. Chapter 5 describes the evaluation of the algorithm's performance in addition to preliminary results from the application of a random forest to the algorithm output. The thesis is concluded in chapter 6 with closing remarks in addition to a brief discussion of potential avenues of expansion.

Chapter 2

Background

The following provides a general overview of both drylines and the computational techniques used by the final algorithm.

2.1 Meteorological Overview of Drylines

2.1.1 General Description

A dryline is an elongated meteorological boundary of O(100km) in length denoting the separation of two air masses, one moist, the other dry. Drylines are common to the US Plains (Schaefer, 1986; Fujita, 1958), and occur most frequently from late spring to early summer. During these months, the Gulf of Mexico is dominated by a mean southerly flow, resulting in moist air being brought inland. Meanwhile, westerly flow aloft brings dry air eastward from the lee of the Rocky Mountains, or from further south in the arid regions of northern Mexico or the southwestern US into the US Plains. The resulting distribution of moisture is in part controlled by the gradual decrease in terrain elevation exhibited from the Rockies to the Mississippi River Valley. The east-west sloping inhibits east-west motion (Schaefer, 1974), enhancing the north-south extent of the moist air mass. The distribution of moisture is in line with the mean monthly dewpoint temperature field maps created by Dodd (1965), which demonstrate

a strengthened east-west moisture gradient in the Central US during the late spring to early summer, and clearly show the influx of moisture from the Gulf. The resulting intersection of the two air masses is the primary means of dryline formation, resulting in a 3D interface illustrated conceptually in Fig. 2.1. Similar to the east-west gradient in terrain height, vegetation and soil moisture increase further east in the US Plains. These features act to control the dryline, with the latter assisting in the intensification of any surface moisture gradient present, and contribute to dryline intensity, movement, and structure (Lanicci et al., 1987; Shaw et al., 1997).

Drylines are typically associated with a confluent wind field, oriented such that the mean flow is roughly 90 degrees counterclockwise of the gradient in the moisture field; however, the wind shift is not necessarily continuous. Additionally, Matteson (1969) found that streamline convergence often occurs east of an advancing dryline (or at least its region of high moisture gradient). This is in line with findings from Ogura et al. (1982) during an investigation into the role of low-level mesoscale convergence in the pre-storm environment. In certain cases, the separation of the wind shift from the primary moisture boundary is associated with a secondary moisture boundary similar to and near the dryline. The reason for these additional boundaries is not well understood, however they have been consistently observed (e.g., Ziegler et al., 1997; Buban et al., 2012).

Several studies have been performed that examine dryline frequency, most finding that drylines occur on ~40% of the days in May through July, despite having different selection parameters or locations (Rhea, 1966; Schaefer, 1973; Peterson, 1983). It should be noted that Rhea and Schaefer had fairly similar criteria, although Peterson focused only on Texas drylines. A more recent

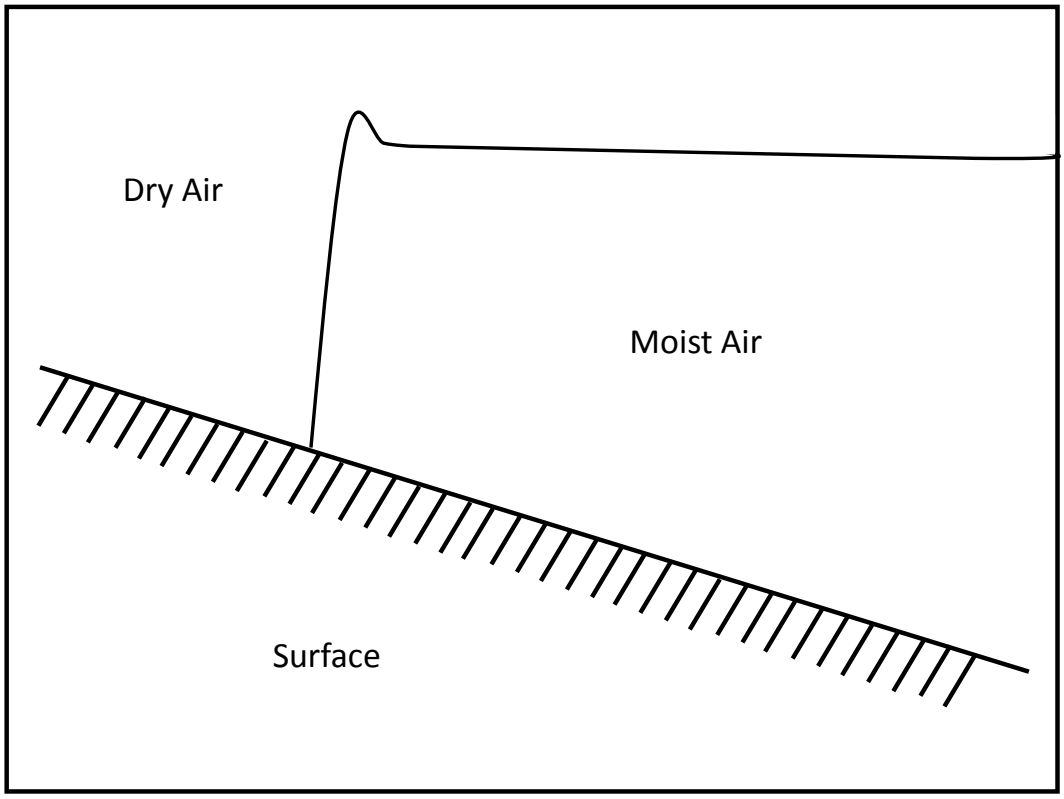


Figure 2.1: Idealized vertical cross-section of a dryline.

climatology performed by Hoch and Markowski (2005) found similar values during mid to late May; however, they also determined that dryline frequency had dropped to ~20% by late June. Coffer et al. (2013) found a similar drop off with drylines occurring on ~20% of days in June, although lower frequencies were present in earlier months (~30% for April and May).

2.1.2 Motion

Dryline motion can be broadly categorized into two cases, each with its own driving mechanisms. An "active" dryline is controlled primarily by a synoptic-scale process (e.g., a mid-latitude cyclone), and moves similar to a standard frontal boundary, associated with a line of pressure minima. While the processes that typically govern local dryline movement are presumably still in effect, the stronger forcing provided by the large scale motions override them, resulting in movement seemingly independent of these effects (described below). In contrast, "quiescent" drylines occur in the relative absence of large-scale forcings. The remaining processes have a strong diurnal dependence, resulting in different types of motion throughout the diurnal cycle.

Quiescent daytime motion is attributed primarily to vertical mixing. Terrain slope results in drylines having the 3D structure shown in Fig. 2.1. Note that at the surface intersection of the two air masses, the depth of the moist region is relatively shallow. The relative lack of depth means that less surface heating is required before the interface becomes less defined and the two air masses begin to mix along the boundary. Additionally, the dry air mass is typically fairly deep during the day (i.e., when this heating would occur) and is well mixed. That is to say that there is little difference between two air parcels outside of dissimilarities occurring due to their vertical distribution. Most importantly, this allows for

the quasi-moist region created during the degradation of the original boundary to be thoroughly mixed with the larger dry air mass. The mixing results in the old boundary region taking on the properties of the dry air, shifting the interface between the two air masses. The well mixed state of the dry region also allows for easy transport of upper-level momentum to the surface, which, given the mean westerly flow aloft in the US mid-latitudes, may also assist in the eastward propagation of the dryline during the day. It is important to note that terrain slope is key to this transition. When encountering a region of flat terrain, the dryline may appear to suddenly jump, as the process described above takes place over the entire moist region at the same time. Similarly, an extreme downward slope can drastically inhibit the dryline's progression as the "erosion" process is dramatically slowed. (Schaefer, 1986; Markowski and Richardson, 2010; Bluestein, 1993).

Beginning around dusk, a combination of effects start to reverse the eastward progression of the dryline (Jones and Bannon, 2002). The relative lack of surface vegetation on the dry side of the boundary results in large surface radiation fluxes. Combined with the dry air's lower specific heat, the two contribute to rapid cooling of the dry air mass at the surface, forming a nocturnal temperature inversion. The inversion decreases the upper-level westerly momentum previously being transported to the surface in the day, inhibiting eastward progression. The differing conditions in the moist region impedes the development of similar processes. Furthermore, clouds that may have formed later in the day over the moist region may still be present, providing an additional measure of protection from nocturnal radiational cooling. All of these aspects come together to create an ageostrophic easterly flow that begins to dominate local

winds, moving the dryline westward. (Schaefer, 1986; Markowski and Richardson, 2010)

2.1.3 Role in Convective Initiation

Localized convergence along the dryline, combined with sufficient low-level moisture, provides a region ideal for convective development. In addition to smaller or less intense cloud formation, drylines are commonly associated with thunderstorm development. Rhea (1966) performed an analysis examining this association over the months of April through June from 1959 to 1962. Utilizing surface and upper air analysis charts in combination with sounding and hourly radar summaries, he found that drylines were a preferred region for new storm growth. In addition, Rhea determined that storms originating near the dryline became better organized more frequently than those forming further away. Studies (e.g., Ziegler et al., 1997) have also indicated that a strengthening dryline is often associated with environmental conditions favorable for convective initiation. However, Schultz et al. (2007) point out that this is not always the case (e.g., 03 May 1999 – see Thompson and Edwards, 2000 for an in-depth study of the event).

In their dryline modeling study, Ziegler et al. (1997) suggest that idealized cloud models may be inadequate when describing convective development along the dryline. In particular, they found that moisture, convective instability, and shear all exhibited regions of strong horizontal inhomogeneity, in contrast to the uniform environment commonly assumed in idealized models. Additionally, their results suggest that deep convection (i.e., that typically related with intense storms) is only possible with the presence of a mesoscale updraft resulting from dryline induced vertical circulations (as opposed to a thermal bubble

formed through perturbations of the local thermal / moisture fields). Atkins et al. (1998) investigate the hypothesis that the radar thin lines found along drylines are manifestations of updrafts associated with horizontal convective rolls (HCR). They found that the HCR interaction with drylines was similar to those of HCRs with sea-breeze fronts formed during offshore flow conditions. The relationship was believed to contribute to conditions favorable to cloud initiation along the frontal (dryline) boundary. Xue and Martin (2006) found corroborating evidence in their modeling study of the 24 May 2002 dryline case, in addition to suggesting that the HCRs also contribute to misocyclone intensifications. Other studies have also found examples of HCRs intersecting dryline circulations (e.g., Wakimoto et al., 2006). Gravity-wave-like effects have been shown to also be a potential cause for convective initiation in near-dryline regions (Koch, 1979).

2.1.4 Dryline Identification

One major difficulty in the development of objective analysis techniques for dryline identification is the disagreement on the subjective analysis criteria. Despite intense examination, subjectively analyzed frontal boundary locations vary greatly between individuals (Sanders and Doswell, 1995). Figure 2.2 from Uccellini et al. (1992) contains an overlay of multiple subjective analyses of the same event. The variability it demonstrates carries through to the analysis of drylines. To provide a definitive location, a defining trait must be selected. Schultz et al. (2007) maintain that the dryline should be located along the line of maximum moisture gradient. In contrast, others such as Rhea (1966) and Bluestein (1993) use the associated wind shift as the defining characteristic of position. While the resulting discrepancies in location are typically minor, the

aforementioned separation of wind shift and moisture gradient occasionally seen in drylines can cause much larger differences. For the purposes of this study, we will identify a dryline's position based primarily on its moisture gradient. The selection follows the criteria set by Coffer et al. (2013), who developed the truth dataset used in this project. In addition, storms developing along the dryline typically form in the moist sector. As wind shifts that do not follow the moisture gradient usually occur in the moist region, the use of moisture gradient as the defining feature places a bound that is likely to ensure storm interference is limited to one side of the boundary.

The next issue involves determining the appropriate moisture gradient characteristics. Schaefer (1973) defines the dryline as a persistent, sharp moisture gradient with at least a 10°F difference in dewpoint temperatures between surrounding surface stations. A minimum dewpoint of 50°F and maximum virtual temperature change of 5°F are also required. However, Schaefer later suggests examining the 9 g/kg isohume or 55°F isodrosotherm as the first guess location of the dryline (Schaefer, 1986). Moreover, these values were chosen for being approximate minimum moisture thresholds associated with tornadic storm formation (Williams, 1976). Crawford and Bluestein (1997), using data from the Cooperative Oklahoma Profiler Studies May 1991 (COPS-91) field experiment, incorporate a temporal aspect for detecting dryline passage. They subjectively determined that an eastward (westward) dryline passage was best typified by a 3°C (2°C) decrease (increase) in dewpoint temperature at a rate of at least 0.5°C per hour over a series of neighboring stations. More recently, Hoch and Markowski (2005) present a spatial mixing ratio gradient threshold of 3.0 g/kg/100km. This threshold was also used in Coffer et al. (2013).

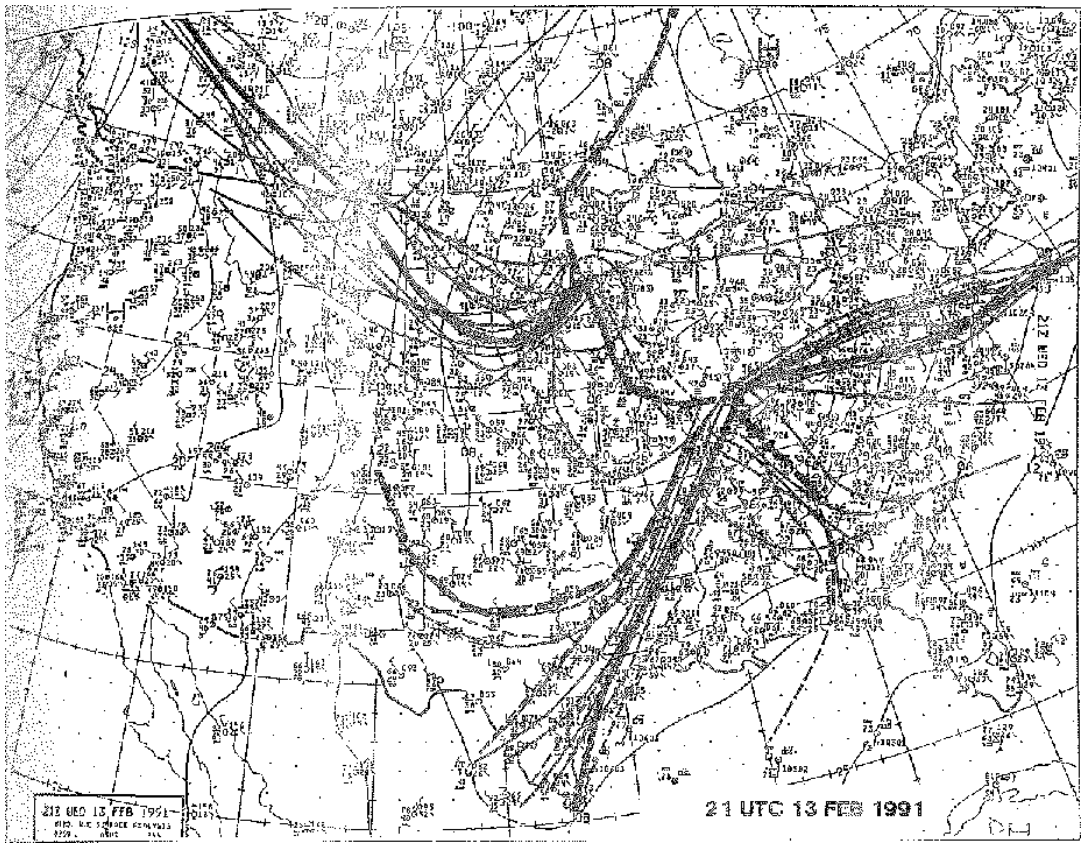


Figure 2.2: An overlay of analyzed boundaries by workshop participants from Uccellini et al. (1992). Thick black lines indicate fronts while troughs and other features are represented with dashed lines. The thin solid lines in the background are objectively analyzed isobars in 4 mb intervals.

The primary reason for dewpoint being commonly used in dryline identification is convenience, as the majority of US moisture observations are provided in terms of either relative humidity or dewpoint (e.g., the Automated Surface Observing System; NOAA, 1998). However, dewpoint is an inconsistent variable. A change in dewpoint from 70°F to 75°F involves an increase in water vapor 12 times larger than a change from 0°F to 5°F (Dodd, 1965). In addition, dewpoint is not conserved across pressure levels, resulting in a sensitivity to terrain elevation. Bearing in mind that the change in moisture content is important, this variability becomes a key factor in maintaining objective consistency. A more accurate measurement is that of the mixing ratio (or specific humidity), which is a direct measure of the amount of moisture present in the atmosphere and is conserved with respect to pressure.

Visible and infrared satellite imagery can assist in dryline location by identifying the cloudy regions that commonly form on the moist side of the dryline as well as identifying the clear regions on the dry side (e.g., Hane et al., 2002). Radar can act in a similar fashion. Additionally, drylines commonly exhibit what is known as a "fine-line" or "thin-line" clear air radar echo near the surface boundary (e.g., Murphey et al., 2006). The thin-line is commonly thought to be the result of insects being caught in confluent winds along the boundary (e.g., Russell and Wilson, 1997; Martin and Shapiro, 2007). Recently, the International H2O Project (IHOP) experiment has provided highly detailed radar imagery that has since been used in many case studies (e.g., Buban et al., 2007; Miao and Geerts, 2007; Sipprell and Geerts, 2007; Wakimoto and Murphey, 2010).

2.2 Numerical Methods

While the majority of techniques used in this project are common in the field of computer science, a brief overview has been provided for readers unfamiliar with the content.

2.2.1 Smoothing

In addition to noise removal, smoothing controls the scale at which data are examined by removing features of a certain spatial frequency, making careful control of filter parameters key to later processing. The type of filter used is dependent upon the type of noise present in the data and the features being examined. The primary concerns of dryline identification are high frequency, small scale variations interfering with gradient calculations and otherwise drawing attention away from the desired boundaries. For this study, we have elected to use a Gaussian filter.

Gaussian filters smooth an image by providing a weighted sum of the values neighboring a given point. The weights are based on a Gaussian distribution described by:

$$f(\boldsymbol{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})) \tag{2.1}$$

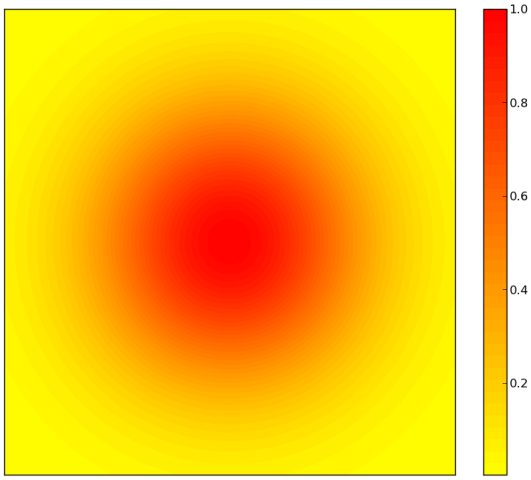
where n is the dimensionality of the Gaussian, $\boldsymbol{\Sigma}$ is the covariance matrix of the Gaussian, and $\boldsymbol{\mu}$ is the center of the Gaussian. The $\frac{1}{(2\pi)^{\frac{n}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}}$ term acts to normalize the distribution such that all Gaussians of a dimensionality equal one when integrated over their domain. The $\boldsymbol{\Sigma}$ parameter controls the extent and orientation of the distribution, determining the scale at which features are

removed. The scalability allows adjustment of the filter's frequency response, effectively modifying the frequency cutoff.

Application of the Gaussian distribution to gridded data requires a discrete representation of the continuous function. The discretization is typically performed by sampling values from the continuous function and placing them at corresponding locations on a grid, or kernel (Fig. 2.3). The kernel is applied to an unsmoothed grid via convolution. The continuous space Gaussian distribution's infinite extent results in truncation errors in discrete space due to the kernel's size; however, the errors can be reduced by increasing the size of the kernel at the cost of increased computational time. Ninety nine percent of the Gaussian distribution's influence is contained within 3σ (standard deviations) of the mean, so kernel widths set to at least twice the grid space representation of 3σ are commonly used (Lakshmanan, 2012).

In most cases, the n dimensional (nD) Gaussian kernel is isotropic and can be decomposed into n identical 1 dimensional (1-D) kernels, where n is the dimensionality of the Gaussian. Each of the smaller 1-D kernels is applied in a single direction, greatly decreasing the computational time required for convolution. While anisotropic Gaussians can be decomposed, the process is much more involved (Geusebroek et al., 2002).

Gaussian distributions are also self-similar in that the convolution of two Gaussians results in another Gaussian. The self-similar property allows m Gaussians to be convolved with each other and another function with only one convolution with the full non-Gaussian grid instead of m. This is of particular use in filter banks (see appendix B.1), where more complex smoothing functions are used.



(a) Continuous 2D Gaussian.

0.09	0.21	0.29	0.21	0.09
0.21	0.54	0.74	0.54	0.21
0.29	0.74	1.00	0.74	0.29
0.21	0.54	0.74	0.54	0.21
0.09	0.21	0.29	0.21	0.09

(b) Discrete 2D Gaussian (kernel).

Figure 2.3: Continuous and discrete representations of a 2D Gaussian.

2.2.2 Gradient Calculation

The detection of moisture change is critical to the identification of drylines, making numerical differentiation a key aspect of this project. Ideally, the use of finite difference methods (e.g., central differencing) combined with Richardson extrapolation produce can produce excellent results. However, all finite difference methods are highly sensitive to noise (e.g., Chartrand, 2011). Smoothing can ameliorate some of the issues that arise from noisy data; however, the effects will still be present to some degree.

2.2.2.1 Sobel Operator

The Sobel operator is used in image processing for edge detection (Weeks, 1996). It is somewhat insensitive to low frequency noise, while also retaining the gradient orientation during calculation. The method is implemented in vector format, calculating the gradient in component directions and later combining them via the square root of the sum of their squares. The process removes the sign of the gradient; however, it can be inferred from the gradient's orientation.

The Sobel operator is most commonly applied using a 3x3 matrix. While larger kernels may be used, there is little guidance for their implementation (Nixon and Aguado, 2012). Nixon and Aguado (2012) describe a method designed to combine smoothing and differencing in the same operator. To smooth, the method uses Pascal's triangle due to the resulting coefficients' similarity to a Gaussian kernel, if given infinite extent. The differentiation is similarly provided by Pascal's triangle for subtraction. The resulting gradient operator is formed by performing element-wise multiplication of the smoothing and differentiation matrices.

2.2.2.2 Wind Direction

It should be noted that taking the gradient of the wind direction requires manipulation of standard techniques. The Sobel operator is designed to work with data on a linear scale. However, wind direction is an inherently cyclic variable (e.g., a change 7° to 9° has the same magnitude and direction as 359° to 1°). To apply the technique, we introduce the assumption that no change in wind direction will exceed 180° in magnitude. While not necessarily true, it is both likely and the greatest detectable change. The assumption is implemented by setting the value of each gridpoint being examined during the gradient calculation to 180°. The surrounding values are then shifted such that their relative positions remain the same. Finally, all values are adjusted so that they remain within [0° - 360°) (e.g., a reading shifted from 345° to 375° is changed to 15°). This results in all values being at most 180° from the central value on a linear scale. From here, the gradient calculations continue as normal.

Alternatively, the sine(wind direction) would perform the same function; however the sign of the gradient would not be retained.

2.2.3 Binary Erosion, Dilation

Binary erosion and dilation are morphological techniques used, in their base forms, to shrink or enlarge objects in a binary set of pixels, respectively (e.g., Weeks (1996)). To describe these methods, we first examine dilation. Consider two sets, A and B. A contains the pixels of the objects to be dilated while B contains the pixels of the object that the objects in A are to be dilated by. In most cases, B is a square centered at (0,0) (e.g., B={(-1,-1), (-1,0), (-1,1), (0,-1), (0,0), (0,1), (1,-1), (1,0), (1,1)} as seen in Fig. 2.4b), although more

complex shapes are permitted. The resulting set is determined by Minkowski set addition of A and B,

$$C = A \oplus B \tag{2.2}$$

where C is the set of pixels belonging to the dilated objects. C contains all possible vector additions of A and B. To demonstrate, consider the case where A is a T shaped object and B is a 3x3 square, A = {(0,-1), (0,0), (-1,1), (0,1), (1,1)}, B = {(-1,-1), (-1,0), (-1,1), (0,-1), (0,0), (0,1), (1,-1), (1,0), (1,1)}. The resulting set of pixels contains {(-1, -2), (-1, -1), (-1, 0), (0, -2), (0, -1), (0, 0), (1, -2), (1, -1), (1, 0), (-1, 1), (0, 1), (1, 1), (-2, 0), (-2, 1), (-2, 2), (-1, 2), (0, 2), (1, 2), (2, 0), (2, 1), (2, 2)}. This process is illustrated in Fig. 2.4.

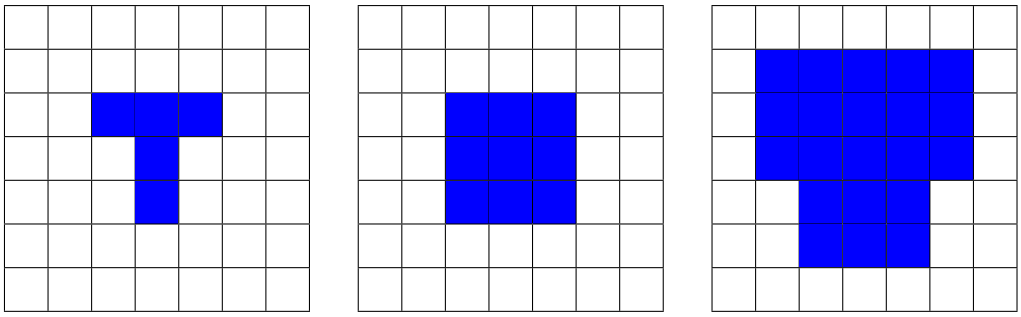
Binary erosion is very similar to binary dilation. It can be thought of as the compliment of the binary dilation of the compliment of A (i.e., all pixels not part of the objects). This results in C being calculated as

$$C = (A^c \oplus -B)^c \tag{2.3}$$

where ^{c} denotes the compliment of a set and the negative sign indicates a rotation of 180°. It should be noted that an implementation based on this description may result in large computational time if the size of A is small relative to A^c. Therefore, binary dilation and erosion techniques may be written differently in practice.

By combining both erosion and dilation, or selecting particular types of B masks, these techniques can have wide ranging effects. For example, if B consists of a single non-origin pixel (i.e., not (0,0)), the resulting dilation will actually be a translation. Similarly, alternating the methods generally results in the objects retaining their approximate original size and location. Binary erosion followed

by binary dilation, called binary opening, removes relatively small features (e.g., noise), smooths contours, and enlarges uneven holes. Binary opening necessarily results in output objects of equal or lesser size than their input counterparts. Binary closing is the reverse of binary opening, resulting in filled holes (e.g., Fig. 2.5). It will either enlarge objects or leave them the same.

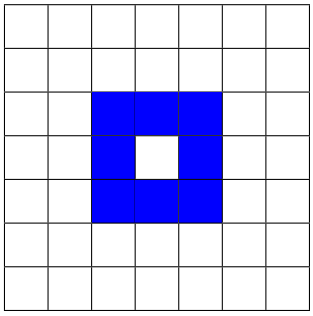


(a) Pixel set A. (b) Pixel set B. (c) Pixel set C.

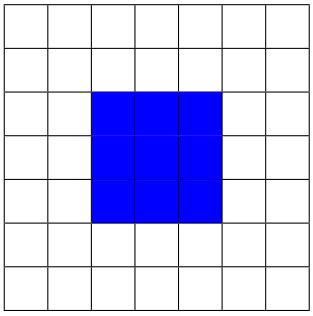
Figure 2.4: Binary dilation applied to a T shape using a 3x3 square.

2.2.4 Non-max suppression (NMS)

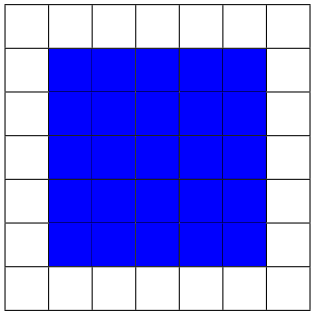
Non-max suppression (NMS) is a technique designed to find continuous lines of maximum (or minimum) gradient (Sun and Vallotton, 2009). The method requires that the magnitude and direction of the gradient be known, but otherwise needs no additional input. NMS is applied by performing a local max only on the examined gridpoint and its neighbors that are perpendicular to the direction of the gradient. The result is (ideally) a continuous line of maxima in the gradient field. While the algorithm also identifies regions of noise, these are limited to thin lines, with more spurious regions being extremely fragmented. The primary drawback of the method (of the implementation used here) is its inability to branch, allowing only one path to be followed per line. The issue can



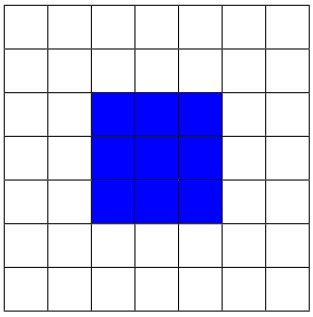
(a) Pixel set A.



(b) Pixel set B.



(c) Pixel set after dilation.



(d) Pixel set after erosion applied to C.

Figure 2.5: Binary closing applied to an empty box using a 3x3 square.

be somewhat solved by means of binary dilation, although spurious connections become more likely with increasing iterations.

Chapter 3

Dataset Description

The dryline dataset used by Coffer et al. (2013) was developed using the 0000 UTC 4 km Weather Research and Forecasting (WRF) model initialization 24 hour forecasts for the months of April, May, and June from 2007–2011 (they also developed similar datasets using the Rapid Update Cycle and North American Mesoscale models). The described time frame was utilized due to the high frequency of dryline events in the US High Plains during these months (see chapter 2.1) and to maintain consistency with previous studies (e.g., Dodd, 1965; Hoch and Markowski, 2005). The WRF model is a Numerical Weather Prediction (NWP) model designed through a collaborative effort of multiple agencies. The WRF's primary utility is in its modular design, enabling various boundary, physics, and other schemes to be tested in differing combinations with ease (Skamarock et al., 2008). The particular settings used in the dataset's origin study are summarized in Table 3.1. It should be noted that data occurring after 9 June 2009 were taken from WRF 3.1.1 runs as opposed to WRF 2.2. The transition resulted in a slight expansion of the domain; however, the other model settings remained unchanged. As the majority identified drylines present in the data occur near the center of the domain, it is believed that the version change has little to no impact on our results.

Coffer et al. (2013) used the Grid Analysis and Display System (GrADS[1]) for gradient calculations and visualization during their research. Following Hoch and Markowski (2005), the primary field they examined was that of the gradient of specific humidity. Drylines were defined as elongated boundaries on the order of 100 km containing some point with a minimum specific humidity gradient magnitude of 3 g/kg/100km. A wind shift consistent with daytime dryline kinematics was also required. Coffer et al. performed several other subjective evaluations in addition to the objective criteria mentioned above (e.g., no strong thermal gradient, not primarily the result of storm outflow). The resulting dataset comprises individual points that are the result of a series of lines connecting subjectively drawn dryline segments. An example of this output can be seen in Fig. 3.1. The data points are represented by black dots and overlay a plot of the specific humidity gradient magnitude. This notation is used throughout the rest of the thesis.

initialization time	00 UTC
grid resolution	4 km
domain	~ CONUS (small expansion from WRF 2.2 to WRF 3.1.1)
microphysics	WRF single-moment six-class (Hong and Lim, 2006)
boundary layer	Mellor-Yamada-Janjić (Mellor and Yamada, 1982)
land surface model	Noah (Chen and Dudhia, 2001)
shortwave radiation	Dudhia (Dudhia, 1989)
longwave radiation	Rapid Radiative Transfer Model (Mlawer et al., 1997)

Table 3.1: Coffer et al. (2013) WRF model settings.

[1]See http://www.iges.org/grads/

The existence of "truth" points allows for the above dataset to be used during algorithm development and distance evaluation as well as during the application of machine learning (see chapter 4 and sections 5.1.1 and 5.2, respectively). However, the statistical evaluation performed in section 5.1.2 requires the use of an independent dataset. Therefore, 2012 WRF data using the same settings as Coffer et al. (2013) was subjectively analyzed by the author and visually compared to algorithm output.

NWP model output contains estimates of the atmospheric state at a given time for each gridpoint present in the model. The variables it provides (and the locations, numbers, etc. of the gridpoints) vary depending on the model and the settings used. Of key importance in this research are: temperature, dewpoint, specific humidity, vapor pressure, u (latitudinal) and v (longitudinal) winds, and surface pressure. Additionally, a number of other variables were derived from these base outputs. For more information, see appendix A.

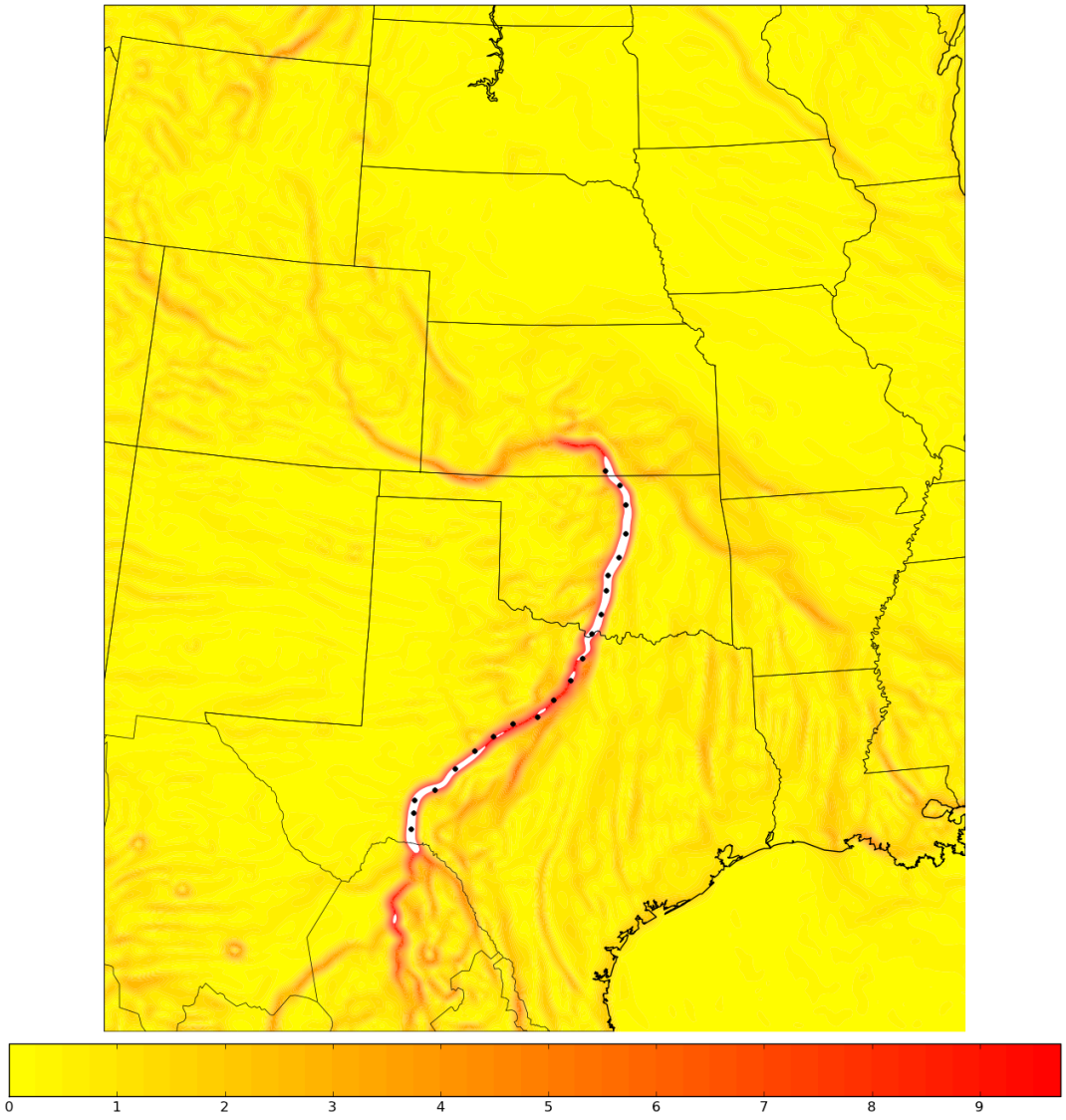


Figure 3.1: Specific humidity gradient magnitude (g/kg/64km) calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 11 May 2010. The black dots indicate the subjectively analyzed dryline location found by Coffer et al. (2013).

Chapter 4

Algorithm Formulation

The following provides a step by step overview of the developed algorithm as it is applied to a 24 hour forecast of 26 April 2012 in which a well defined dryline with a secondary moisture boundary is present. This day has been selected as it provides a clear demonstration of the impact of nearly each step in the algorithm in addition to having several non-dryline boundaries present. Additional information regarding several other implementations examined and further discussion on the techniques mentioned in this section can be found in appendix A and section 2.2, respectively. The overall structure of the algorithm is diagrammed in Fig. 4.1.

The algorithm can be described by three basic steps. First, all regions exhibiting dryline-like moisture values are identified. Next, lines of moisture gradient magnitude maxima are found, providing positions along which objectively analyzed drylines can be located. The dryline-like environment and moisture gradient magnitude ridge masks are then combined and reexamined. The remaining regions form the algorithm's final output. The following provides a more in depth examination of each process.

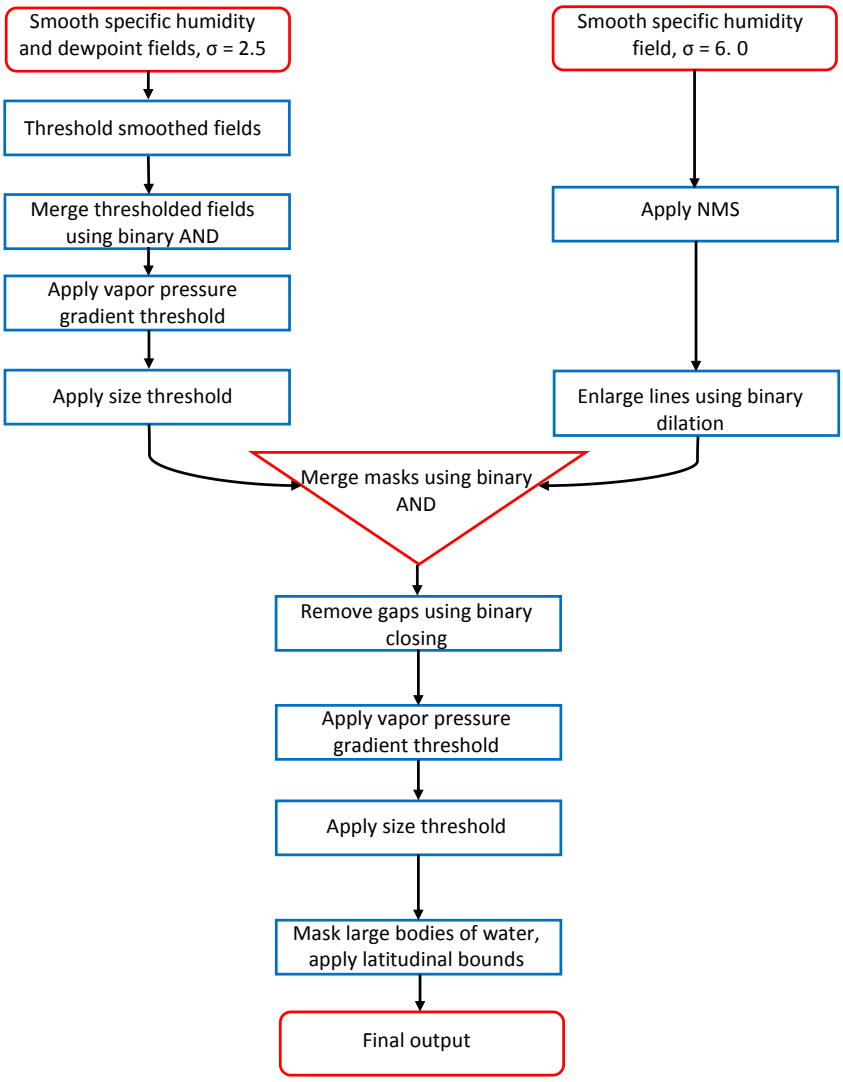


Figure 4.1: Algorithm flow chart.

4.1 26 April 2012 synoptic environment: Key features

The specific humidity field (Fig. 4.3) shows the presence of a strong moisture boundary running from Mexico into southwest Texas, which turns northeast to merge with another boundary near the southern Oklahoma-Texas border. The second boundary runs from central Texas to the merger, and continues along the Red River until turning northeast at the western Oklahoma-Texas border. The structure of the moisture field becomes less organized in this region, although a faint boundary can be seen following the line of pressure minima connecting the low pressure systems in north-central Texas and northeast Illinois (Fig. 4.2). None of the three features have any significant thermal gradient (Fig. 4.5), ruling out cold and warm fronts, as well as storm outflow when considered in conjunction with the wind field (Fig. 4.6). However, the source region of the dry side of the northern boundary is atypical from a standard dryline, with air coming from the northern US and Canada. Combined with the relative weakness of the moisture gradient, we conclude that the northern boundary is not a dryline. In contrast, all indications point to the southwest boundary being a dryline, with a strong moisture gradient, confluent wind shift, weak thermal gradient, and appropriate air mass source regions. The feature present in central Texas is less certain as the dry side winds originate from the Gulf of Mexico. However, given the strength of the moisture gradient, it appears that the air has dried sufficiently while over land to satisfy our criteria. Therefore, the algorithm should identify two dryline boundaries in Texas that merge at the Red River, and continue towards the low pressure region in north-central Oklahoma.

4.2 Thresholding

The first section of the algorithm is designed to identify regions of dryline-like environments. It begins by smoothing the raw specific humidity and dewpoint fields with a Gaussian filter (section 2.2.1) using a σ of 2.5 grid points, which, at the resolution of our data, is approximately 10 km. The process is designed to retain those features roughly the size of the dryline, simplifying subsequent identification in addition to reducing noise.

Next, the gradients of the smoothed moisture fields are calculated using a 3x3 Sobel operator (section 2.2.2.1). As with all finite difference methods, the technique is particularly sensitive to high frequency noise, making the previous smoothing key to obtaining reasonable values. The gradient process results in Figs. 4.7 and 4.8, showing the two strong moisture boundaries present in Texas and Oklahoma as well as the third, weaker boundary to the northeast. Dewpoint's dependence on elevation and non-linear relation with atmospheric water content results in the latter boundary not being as clearly defined in the dewpoint gradient magnitude field as opposed its greater visibility in the specific humidity gradient magnitude data. In addition, a larger number of spurious features are visible in the western portion of the image (e.g., along the pressure minima connecting the lows in Texas and Colorado).

Once calculated, the magnitudes of the two moisture gradient fields are thresholded to provide a preliminary estimate of dryline locations. The initial thresholding is performed using optimistic (i.e., lower) thresholds designed reduce the risk of breaking apart desired features early on at the cost of including spurious values. The dewpoint and specific humidity gradient fields generally differ along changes in terrain elevation (discussed in section 2.1.4 and above).

However, these differences are small enough that strong gradient regions are captured in both thresholded fields (Figs. 4.9, 4.10). This feature can be exploited through the application of a binary AND to the two thresholded grids. The process identifies those pixels containing values that meet both the specific humidity and dewpoint gradient magnitude thresholds while ignoring all others. This results in the retention of true dryline regions while using the variation of the two fields to remove or break up undesired features that were incorrectly identified by one or the other grids. In this case (Fig. 4.11), the weaker capture of the northeastern boundary (i.e., less of the structure is retained) by the thresholded dewpoint gradient magnitude acts to break up the better capture by the thresholded specific humidity gradient magnitude. Likewise, the better defined features seen in the northwest portion of the thresholded dewpoint gradient magnitude field are removed by the thresholded specific humidity gradient magnitude.

The remaining regions are then required to contain at least one pixel above a vapor pressure gradient threshold and extend beyond a 180km (height) x 220km (width) bounding box. The former criterion removes a large number of non-dryline features in addition to smaller disconnected regions (Fig. 4.12); however, it is unable to remove those sections connected to a true dryline including a large region in southwest KS/northwest OK and a number of smaller protrusions further south. The latter prevents the inclusion of small, noise-like features

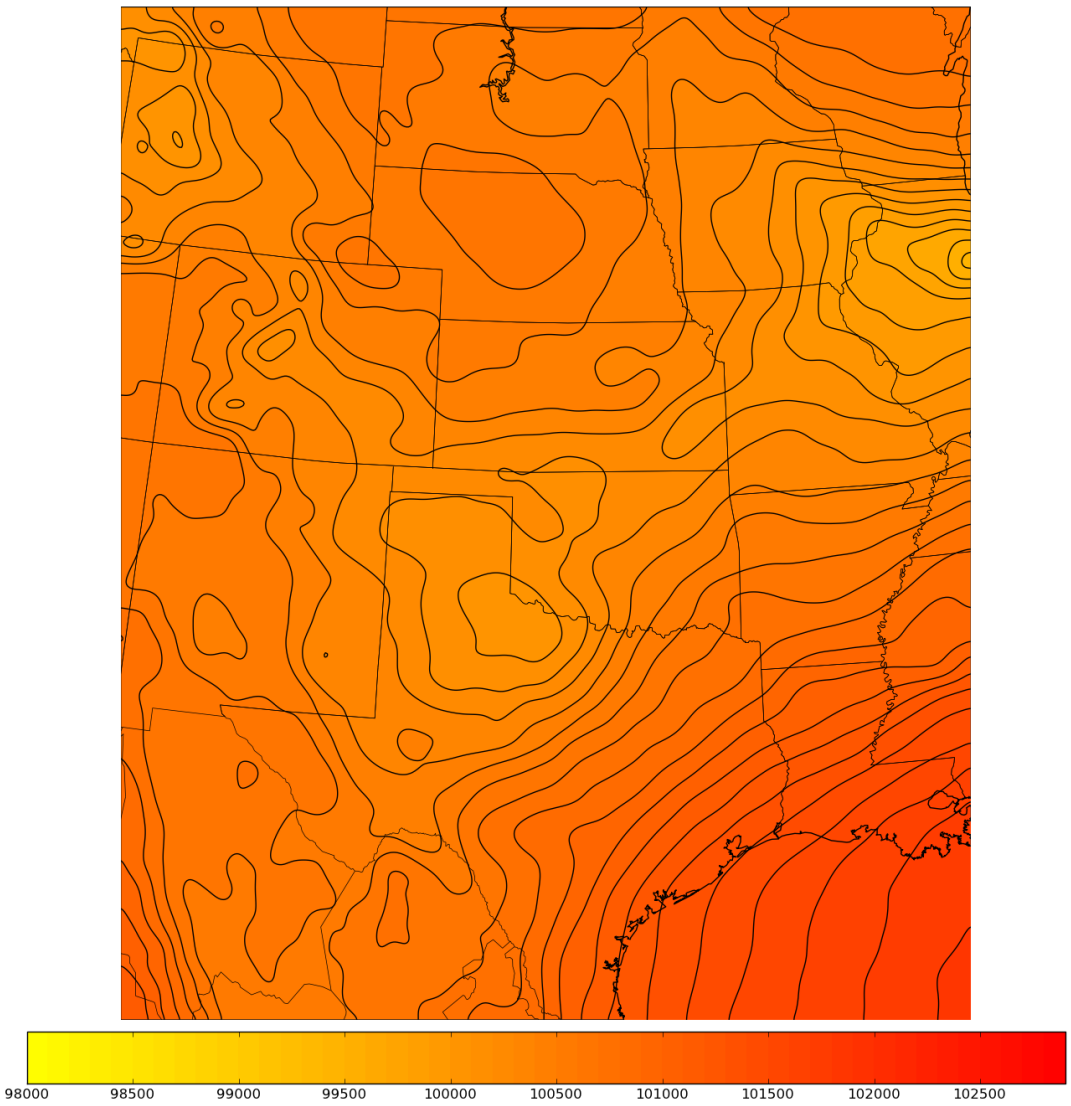


Figure 4.2: 24 h forecast of surface pressure (Pa), contoured every 100 Pa (1 mb), from the NSSL-WRF valid 0000 UTC 26 April 2012.

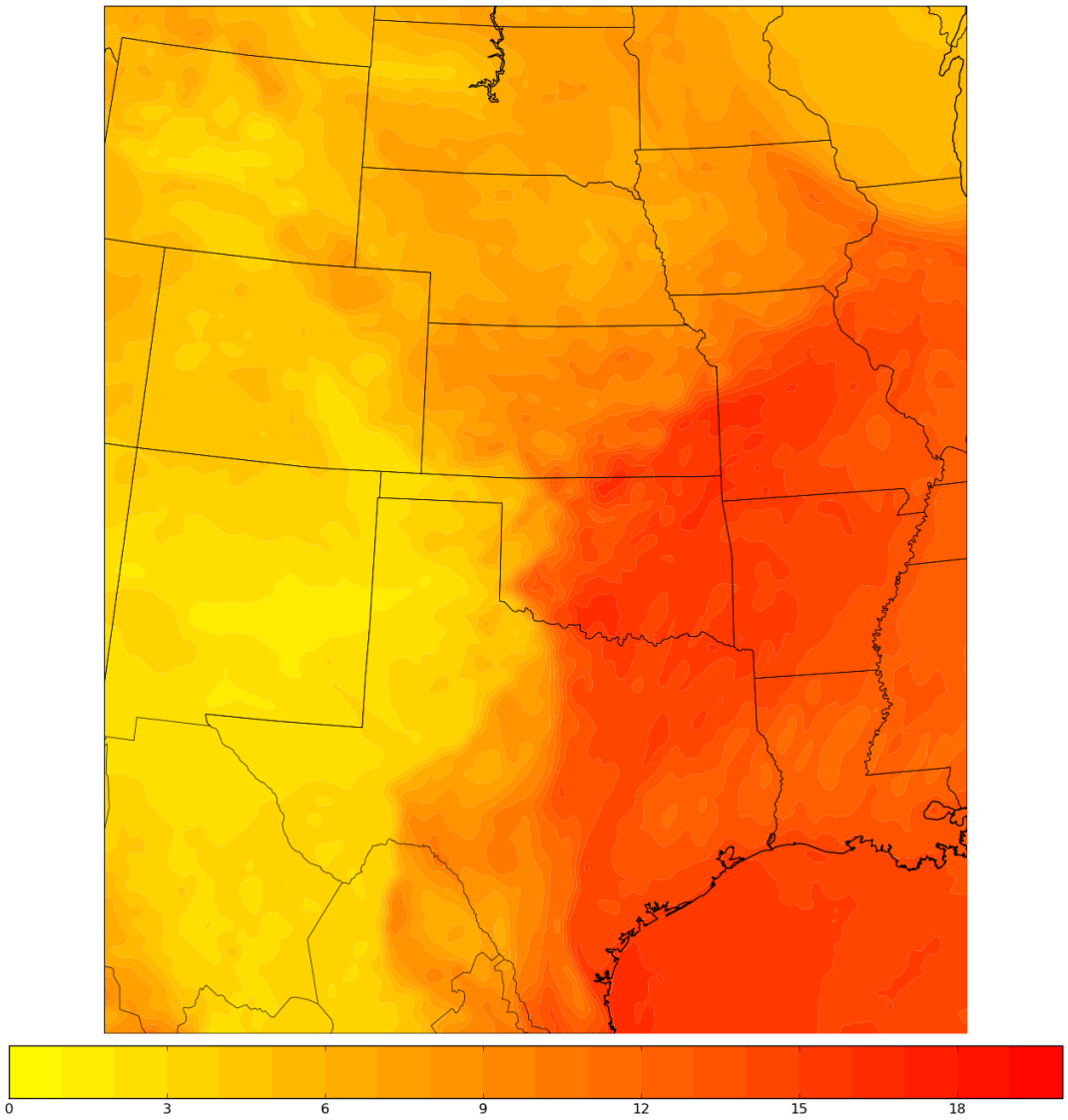


Figure 4.3: 24 h forecast of specific humidity (g/kg) from the NSSL-WRF valid 0000 UTC 26 April 2012.

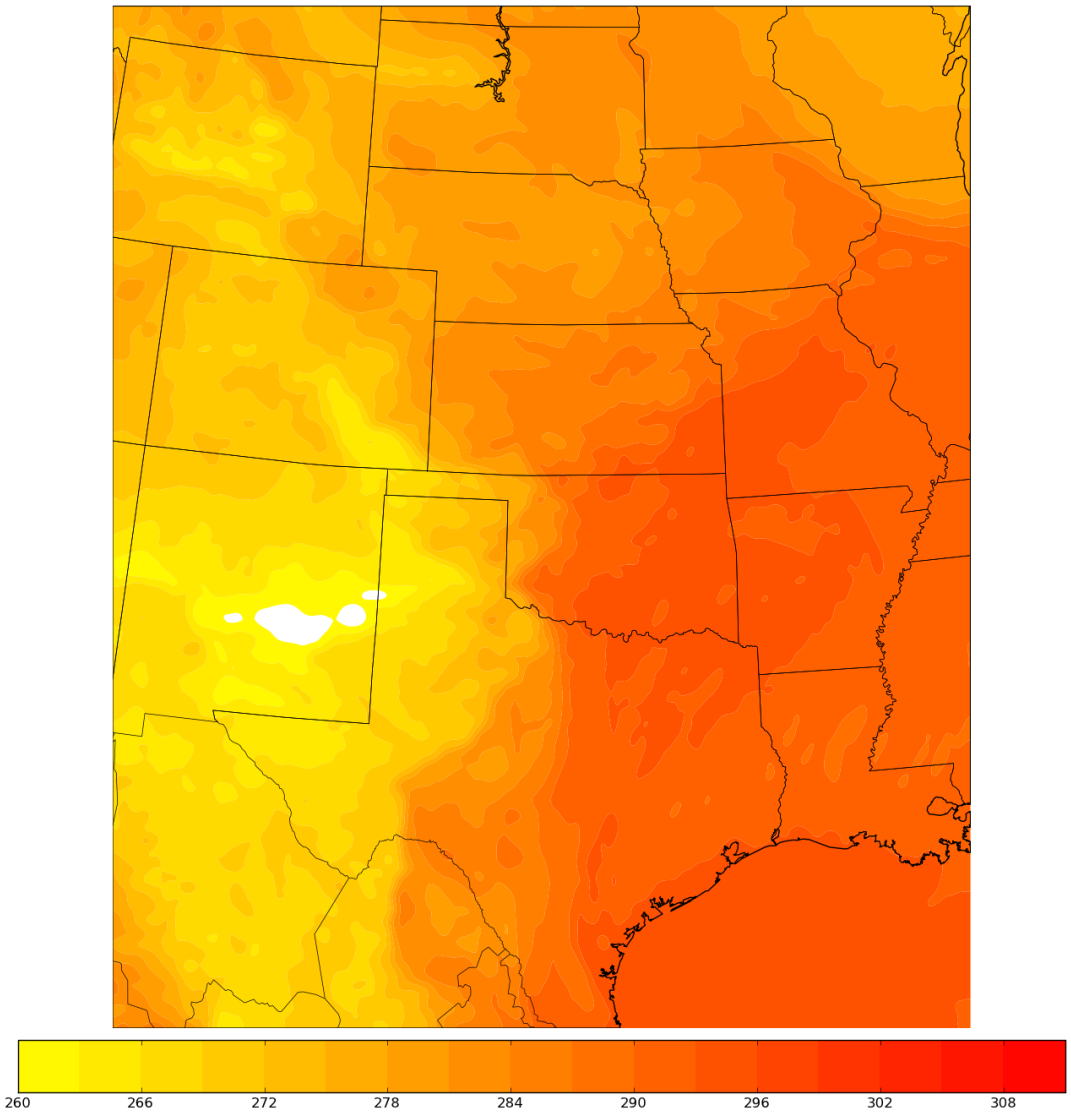


Figure 4.4: 24 h forecast of dewpoint (K) from the NSSL-WRF valid 0000 UTC 26 April 2012.

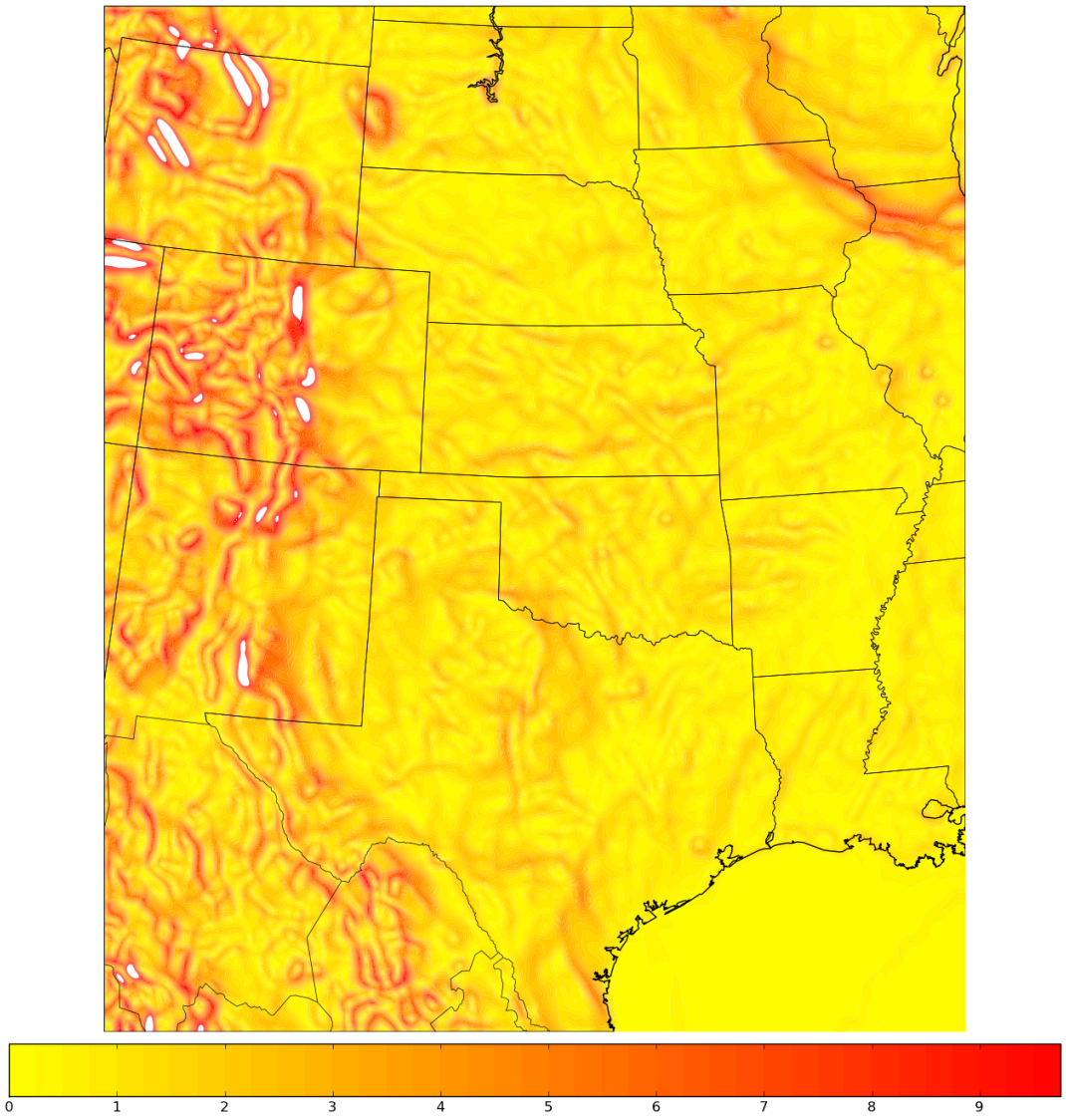


Figure 4.5: Virtual potential temperature gradient magnitude (K/64km) calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 26 April 2012.

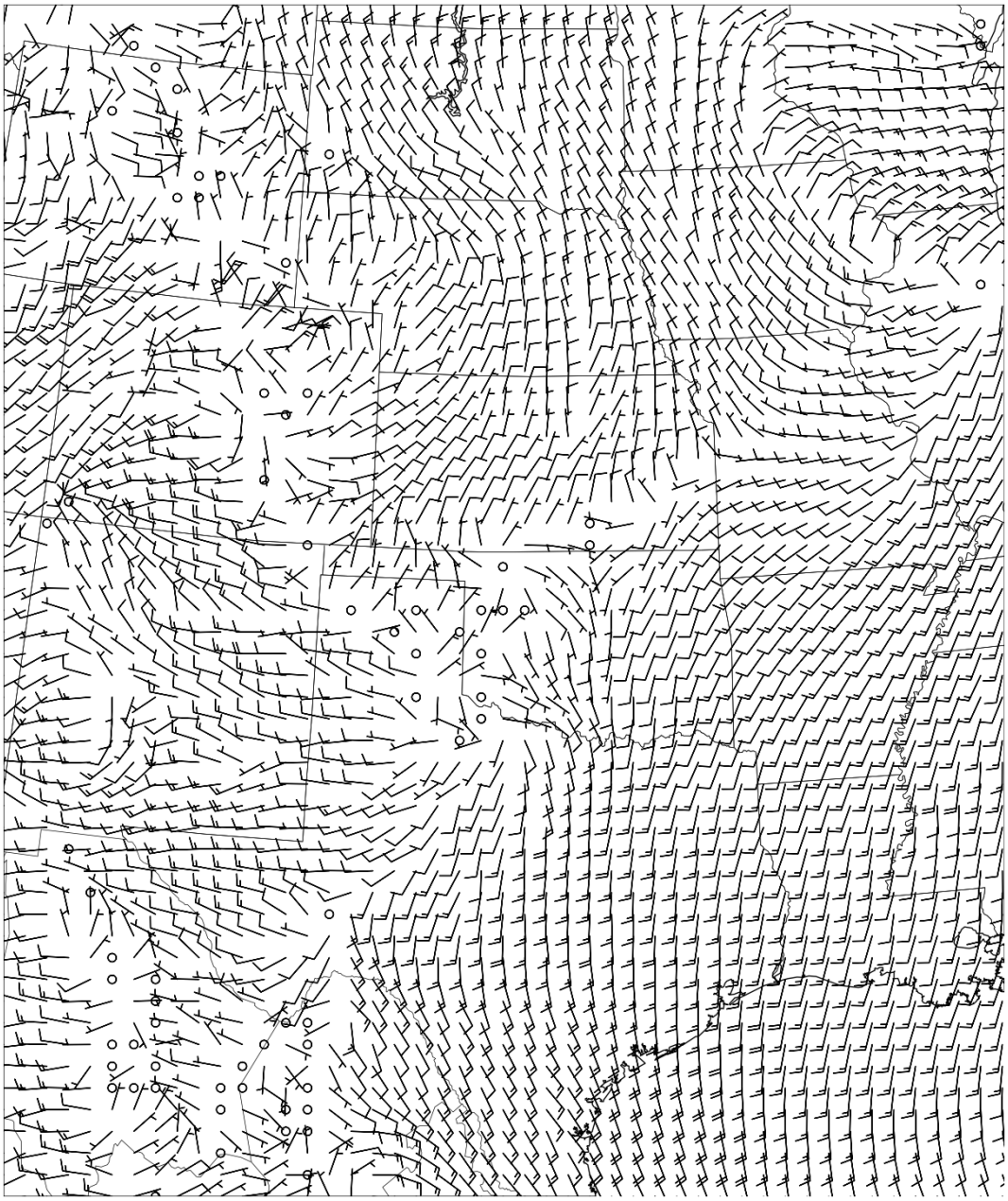


Figure 4.6: 24 h forecast of 10 m wind speed and direction from the NSSL-WRF valid 0000 UTC 26 April 2012. Small flags indicate wind speeds of 5 kts, and larger flags 10 kts. Circles indicate calm winds.

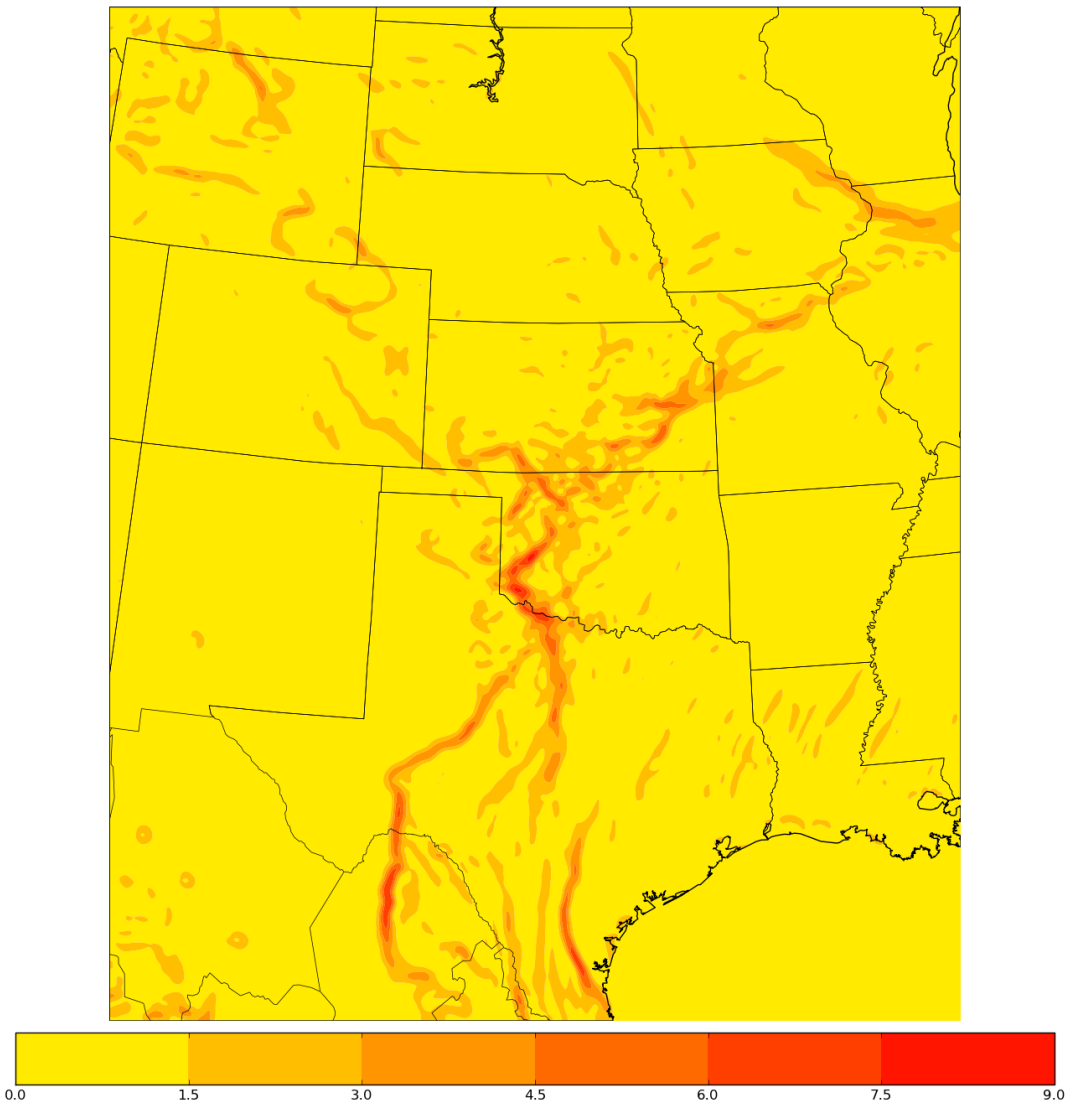


Figure 4.7: Specific humidity gradient magnitude (g/kg/64km) calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 26 April 2012.

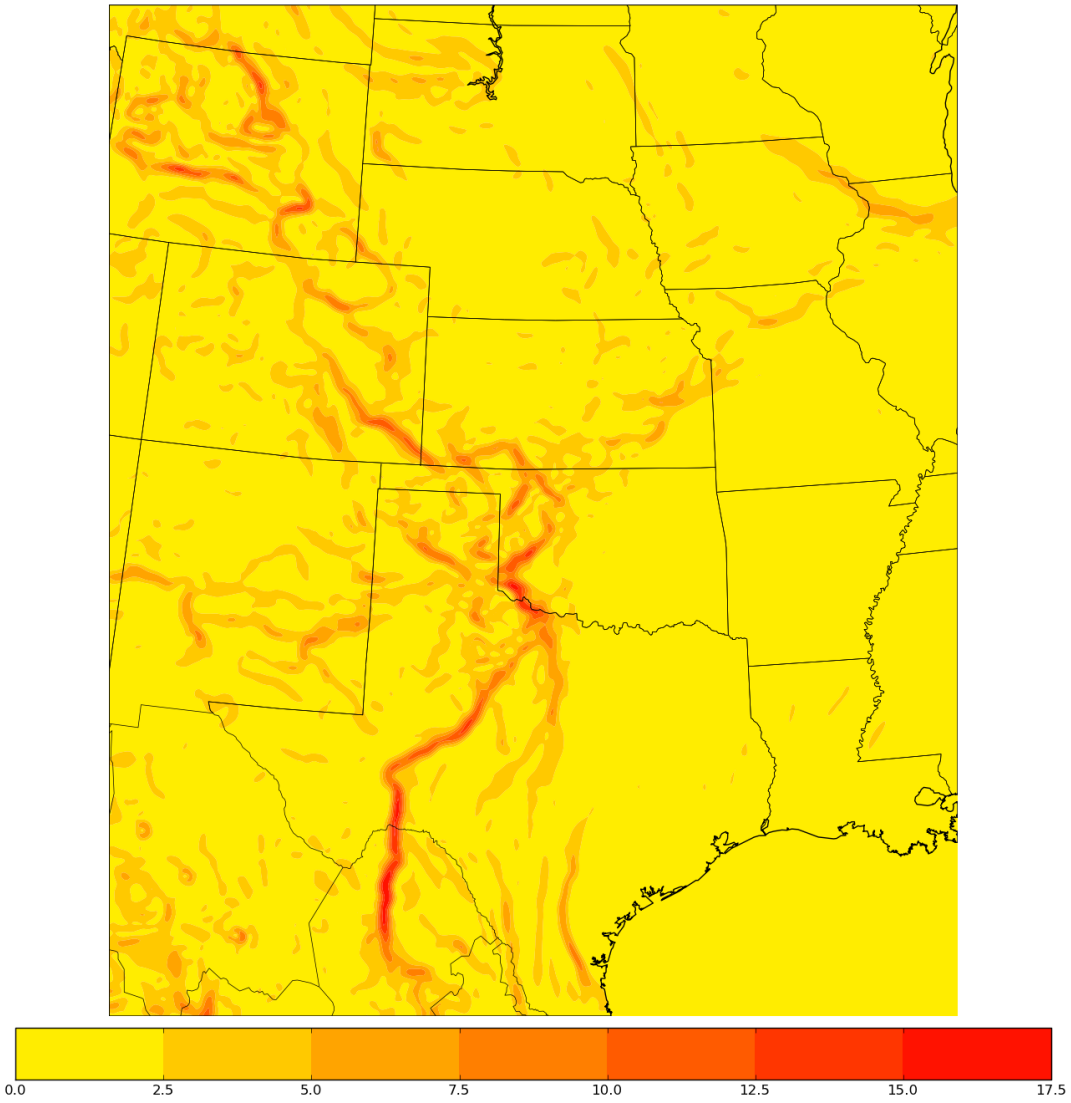


Figure 4.8: Dewpoint gradient magnitude (K/64km) calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 26 April 2012.

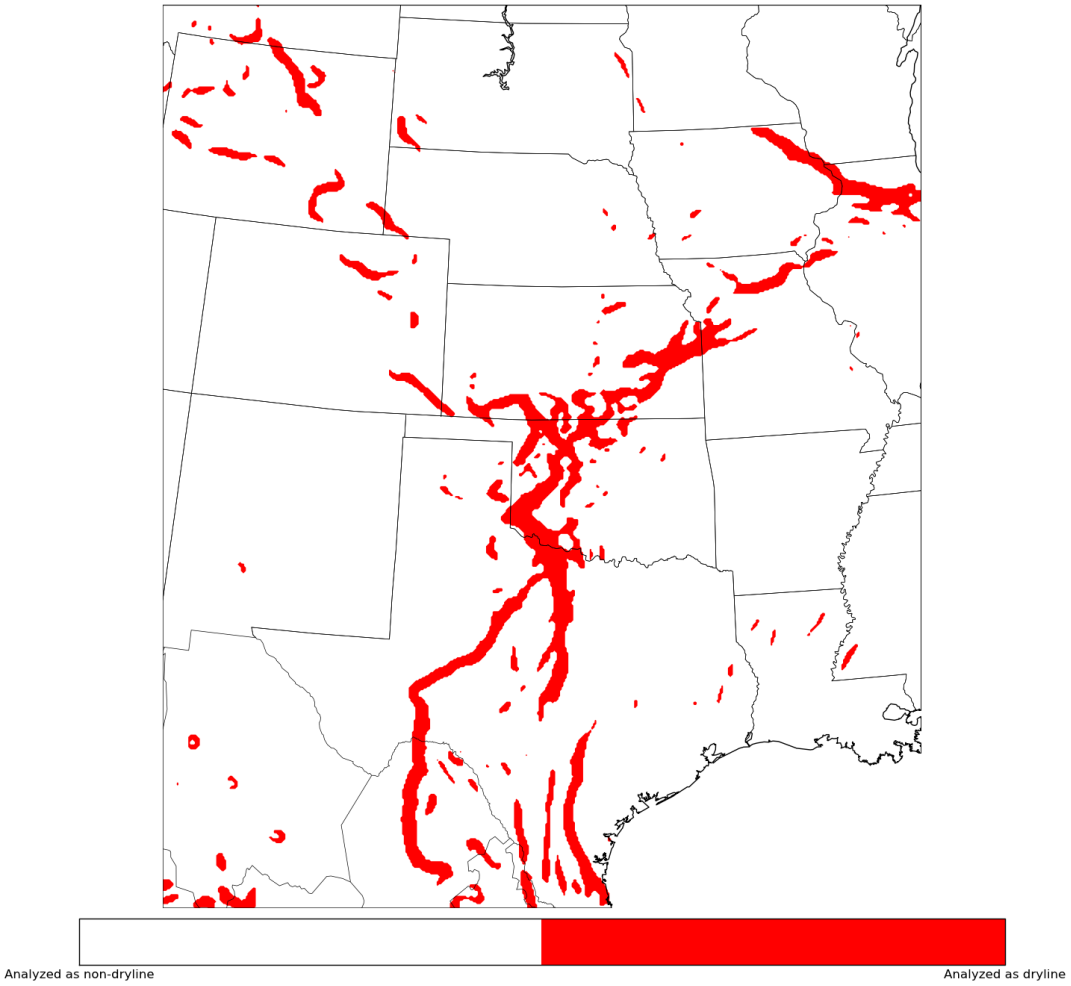


Figure 4.9: Thresholded [>2.0 g/kg/64km] specific humidity gradient magnitude (g/kg/64km) calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 26 April 2012.

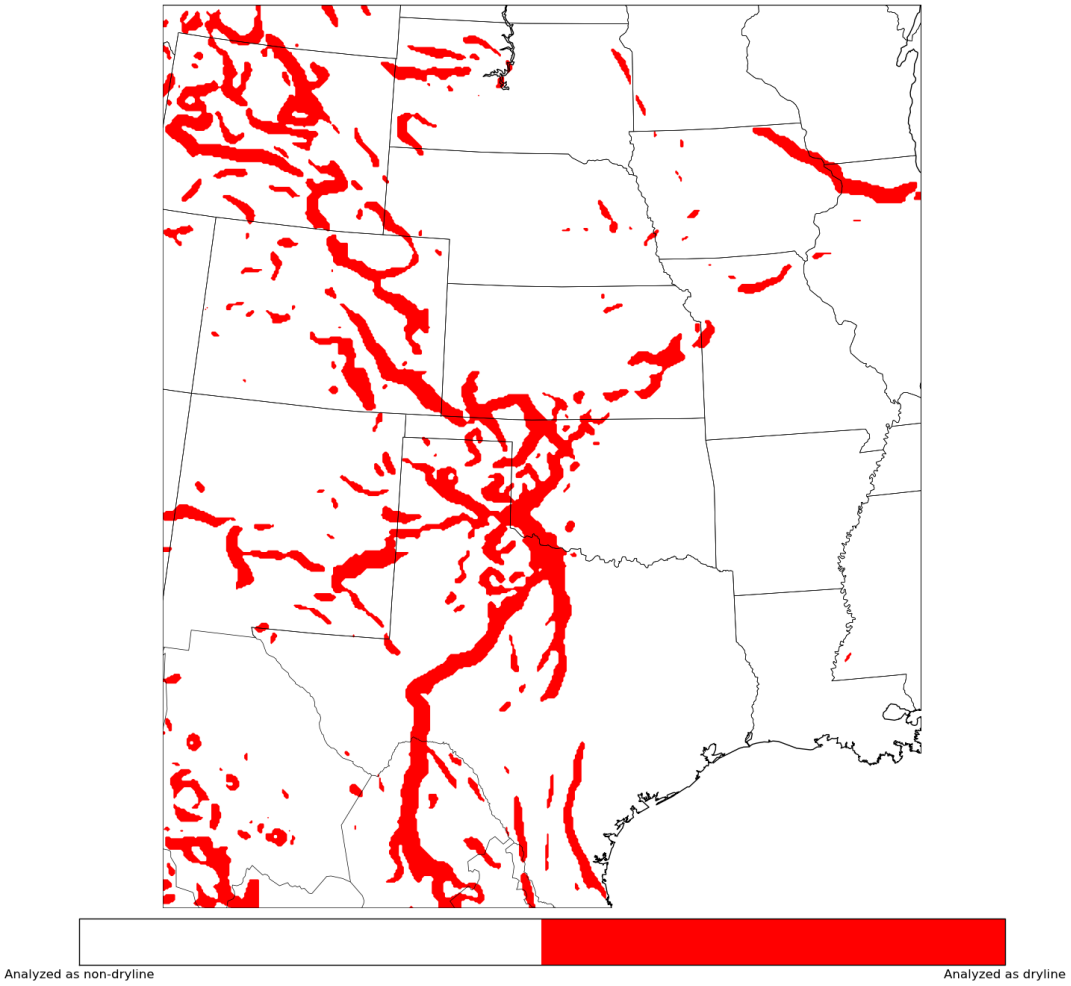


Figure 4.10: Thresholded [>3.5 K/64km] dewpoint gradient magnitude (K/64km) calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 26 April 2012.

regardless of their moisture gradient; however, it has no effect on this particular case (Fig. 4.13). The output forms the final mask of this section of the algorithm.

4.3 Non-max suppression (NMS)

The next section of the algorithm begins independently of the previous step and is designed to identify the lines of strongest moisture gradient along which any potential drylines could occur. The process is based on the application of non-max suppression (NMS, section 2.2.4) to the specific humidity gradient. However, the method does not allow for branching (i.e., two lines of maxima merging into or splitting from one). This creates the potential for identified lines to follow smaller scale, spurious features instead of the true dryline. In particular, more realistic branches may have a weaker specific humidity gradient magnitude near the branch point that reintensifies further along the identified region. Two steps are taken to prevent these deviations. First, the specific humidity field used for this process is smoothed using a $\sigma = 6$ grid points (approximately 24 km, for this study) Gaussian filter, resulting in the specific humidity gradient magnitude field shown by Fig. 4.14. The increased smoothing results in expanded regions of high gradient magnitude as well as the removal of finer-scale variability (compare to Fig. 4.7). These changes improve the NMS output by removing some of the smaller potential branches while also increasing the spacing between identified lines. In addition, the smoothing greatly reduces the number of artifacts caused by the smaller variations in gradient intensity. Second, a number of binary dilations (i.e., dilation of the identified regions through the addition of a one pixel border; section 2.2.3) equal to half the size

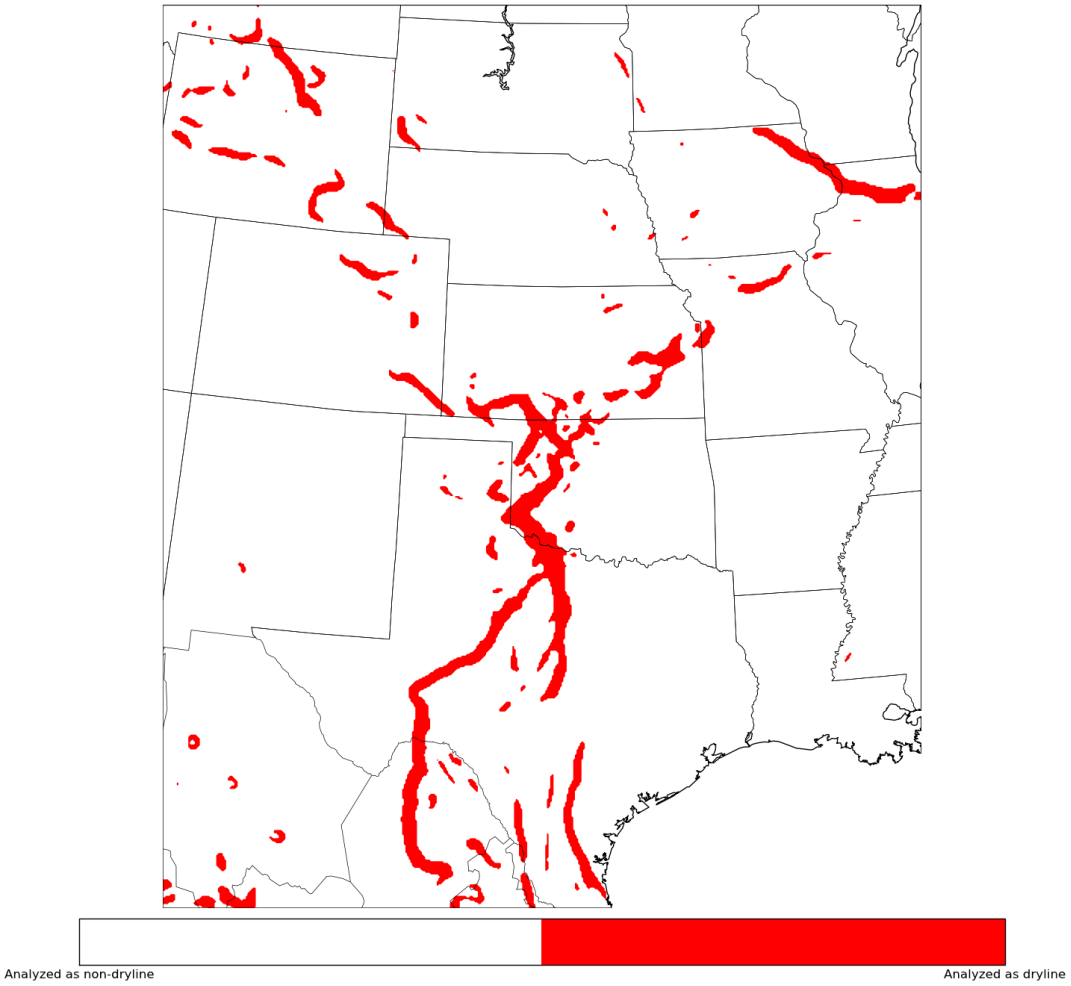


Figure 4.11: Binary AND of thresholded specific humidity gradient magnitude (g/kg/64km) and dewpoint gradient magnitude (g/kg/64km) calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 26 April 2012.

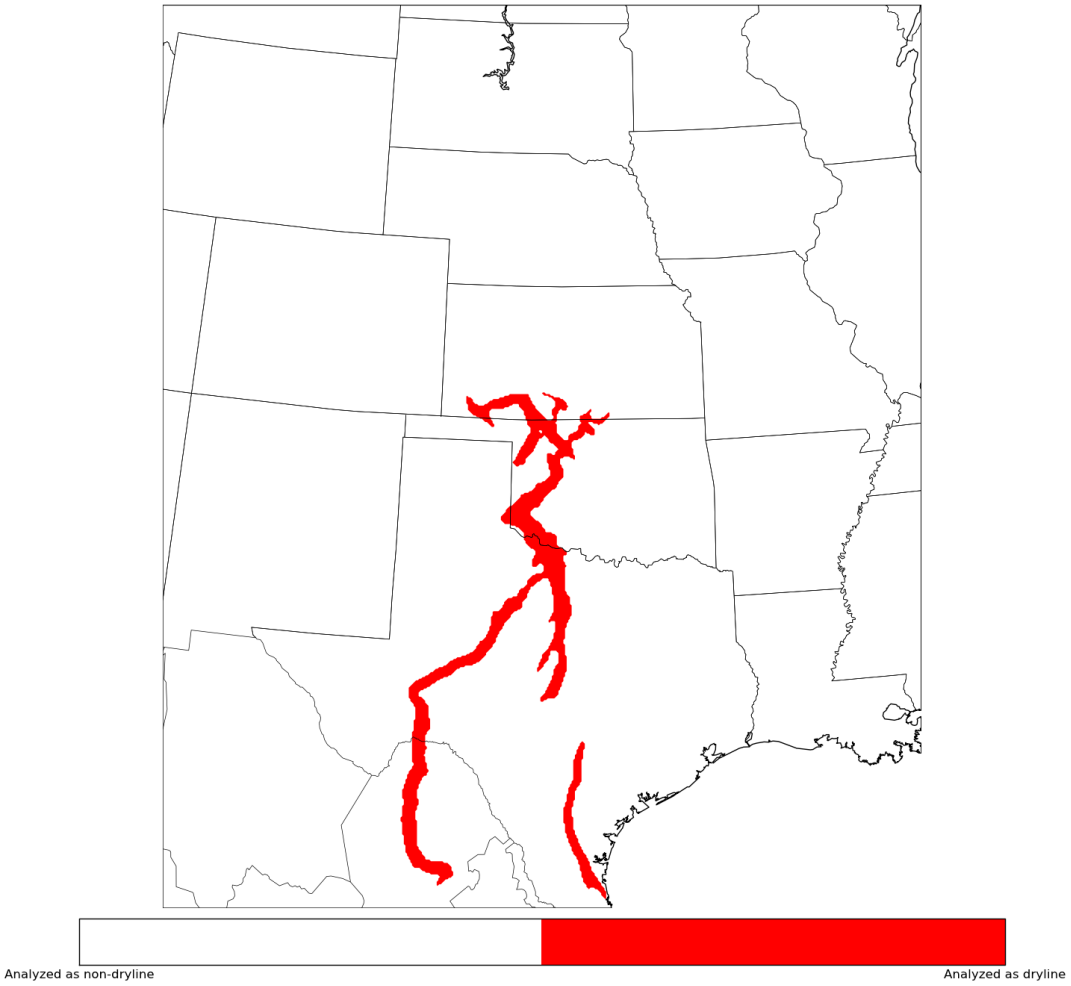


Figure 4.12: Vapor pressure gradient magnitude threshold [>1000 Pa/64km] applied to binary AND of threshold masks using 24 h forecast data from the NSSL-WRF valid 0000 UTC 26 April 2012.

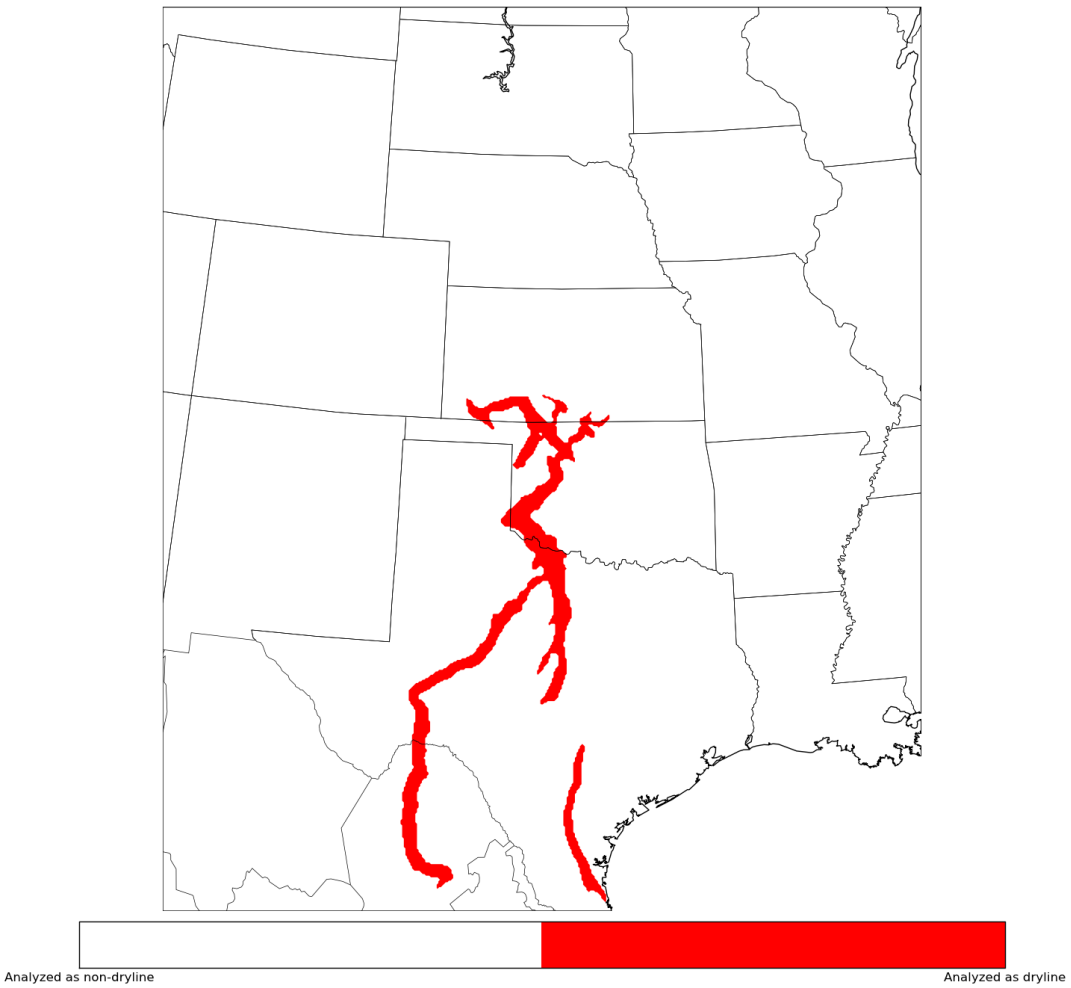


Figure 4.13: Size criteria [> 180km (height) x 220km (width) bounding box] applied to binary AND of threshold masks (post-vapor pressure gradient magnitude threshold) using 24 h forecast data from the NSSL-WRF valid 0000 UTC 26 April 2012.

of the mask used by NMS expand the identified lines, which helps to reconnect any broken branches. The procedure has the added benefit of creating a buffered mask that can be compared with the dryline-like region found in section 4.2. The resulting mask (pre-binary dilation) can be seen in Fig. 4.15. The lines on which the drylines lie are longer than the threshold mask (Fig. 4.13) would suggest. These extensions, in addition to the large number of other lines, are due to the lack of stopping criteria present in NMS. The lines are only required to follow ridges in gradient magnitude, regardless of a location's value.

4.4 Combination and final processing

The final portion of the algorithm begins by combining the binary masks created in the previous two steps using a binary AND, resulting in Fig. 4.16. The combination allows for identified lines of strong gradient to be given bounds, shortening the southward extension of the two boundaries from NMS while also diminishing the northernmost extension of the eastern boundary. In addition, the threshold mask removes the majority of the noise present in the NMS output, while NMS removes many of the spurious extensions seen in the threshold mask. The process has the additional benefit of further breaking apart the undesired features in northern Oklahoma/southern Kansas (see Fig. 4.18). In essence, the thresholding mask removes the spurious lines found by NMS, restricting the output to regions where a dryline-like environment is present. Similarly, the NMS mask acts to limit the spurious deviations that are identified during the initial thresholding process by removing all regions that do not fall along a line of maximum moisture gradient. However, the output mask often has small gaps

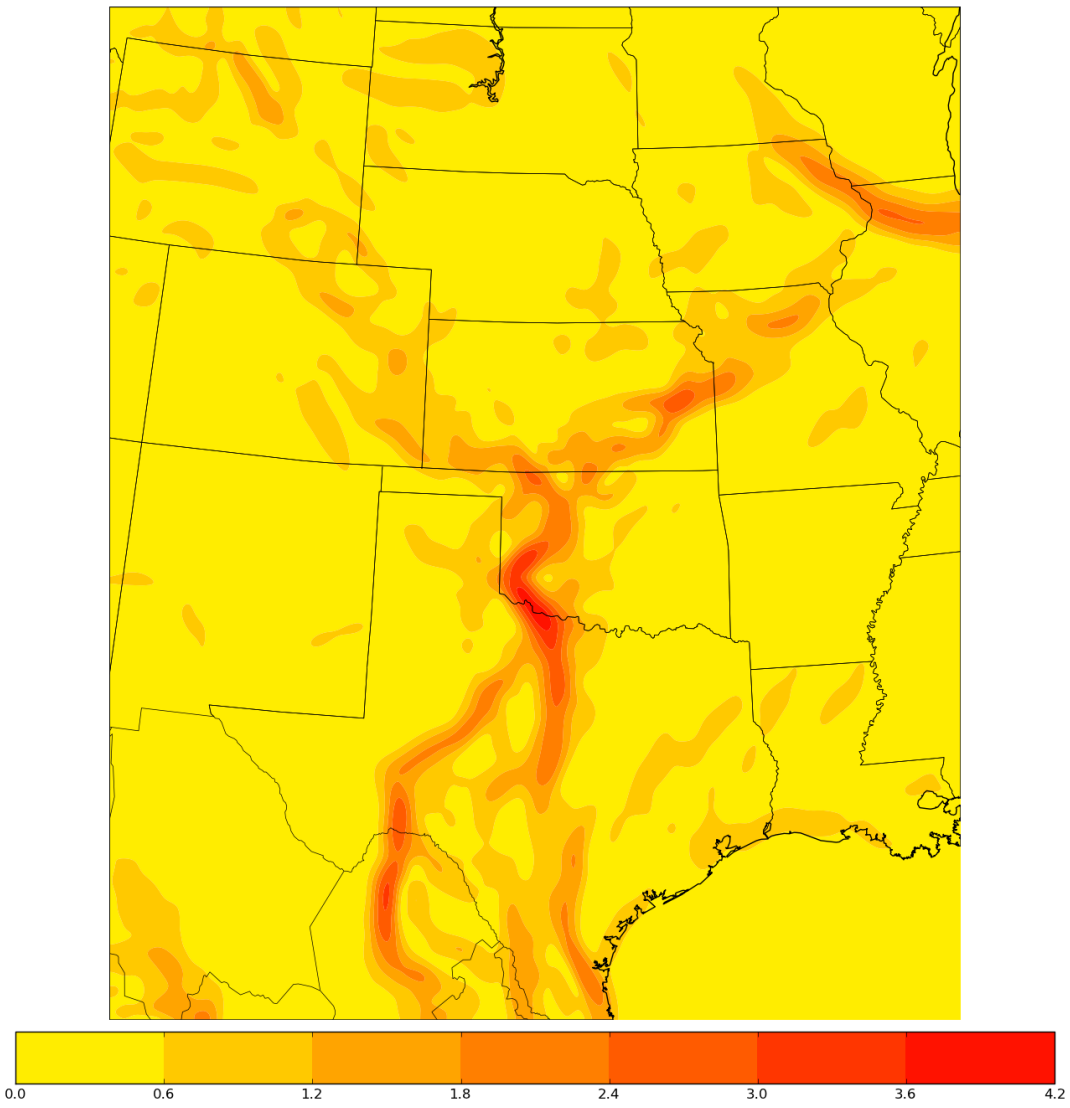


Figure 4.14: Specific humidity gradient magnitude (g/kg/64km) ($\sigma = 6$) calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 26 April 2012.

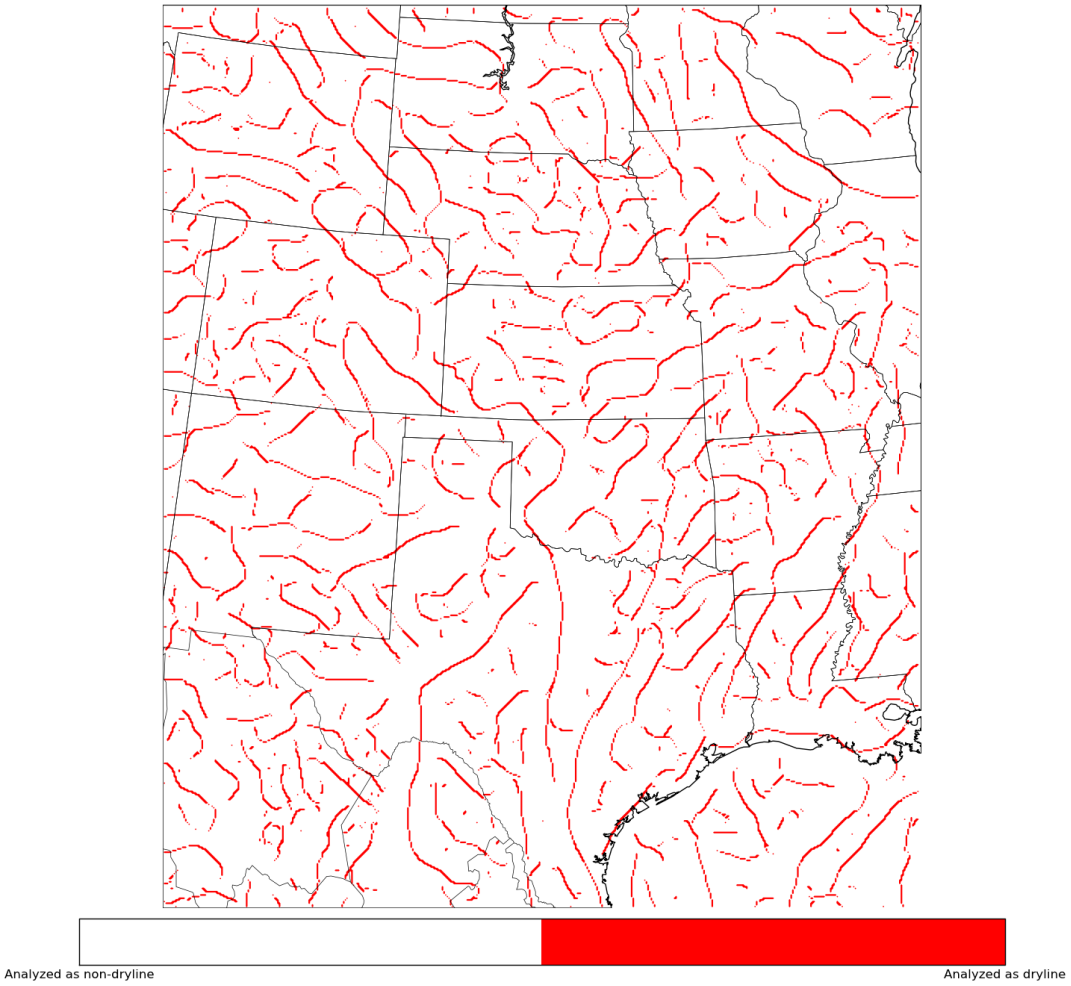


Figure 4.15: Application of NMS to specific humidity gradient magnitude ($\sigma = 6$) calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 26 April 2012.

present due to minor discrepancies in the two input grids. The holes are removed using binary closing (section 2.2.3), which recursively expands and contracts the identified regions to fill in the empty spaces, and results in Fig. 4.17.

The vapor pressure gradient magnitude check is reapplied to remove any disconnected noisy regions resulting from the above process. In this case, the algorithm is able to remove the non-dryline features in northern Oklahoma / southern Kansas that were previously connected to the larger eastern feature (Fig. 4.18). Next, the size threshold is reapplied in another attempt to remove undesired features; however, it again has no effect on this case (Fig. 4.19).

Finally, a set of masks is applied to remove regions near large bodies of water (e.g., Great Lakes, Gulf of Mexico) in addition to placing latitudinal bounds on where a dryline can be identified. The latter mask excludes all values outside of approximately 30° N to 42° N to maintain consistency with Coffer et al. (2013). The resulting image forms the final output of the algorithm (Fig. 4.20).

4.5 Walkthrough assessment

In the 26 April 2012 case, the algorithm identified the two drylines and placed them in the locations described in section 4.1. However, the eastern dryline was allowed to extend too far north, merging with what amounts to noise around the surface low pressure region. In this case, the cyclonic (counter-clockwise) winds wrapped the moisture around the low pressure region, maintaining the high gradient magnitude boundary (this is particularly visible in the wind field; Fig. 4.6). When the true dryline was connected with the false region in the initial thresholding, the algorithm treated the two bodies as a single feature. This allowed the undesired portion to pass the vapor pressure gradient check

when it would have otherwise failed. Likewise, the continuity of the high specific humidity gradient resulted in NMS merging the two regions into a single object. Once both the threshold and NMS portions of the algorithm captured the false region as part of a true dryline, there is no method in place to remove it. In an operational situation, it would be up to forecasters to subjectively identify and correct the issue.

Figure 4.16: Binary AND of NMS and thresholded masks using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 26 April 2012.

Figure 4.17: Binary closing of combined NMS and thresholded masks using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 26 April 2012.

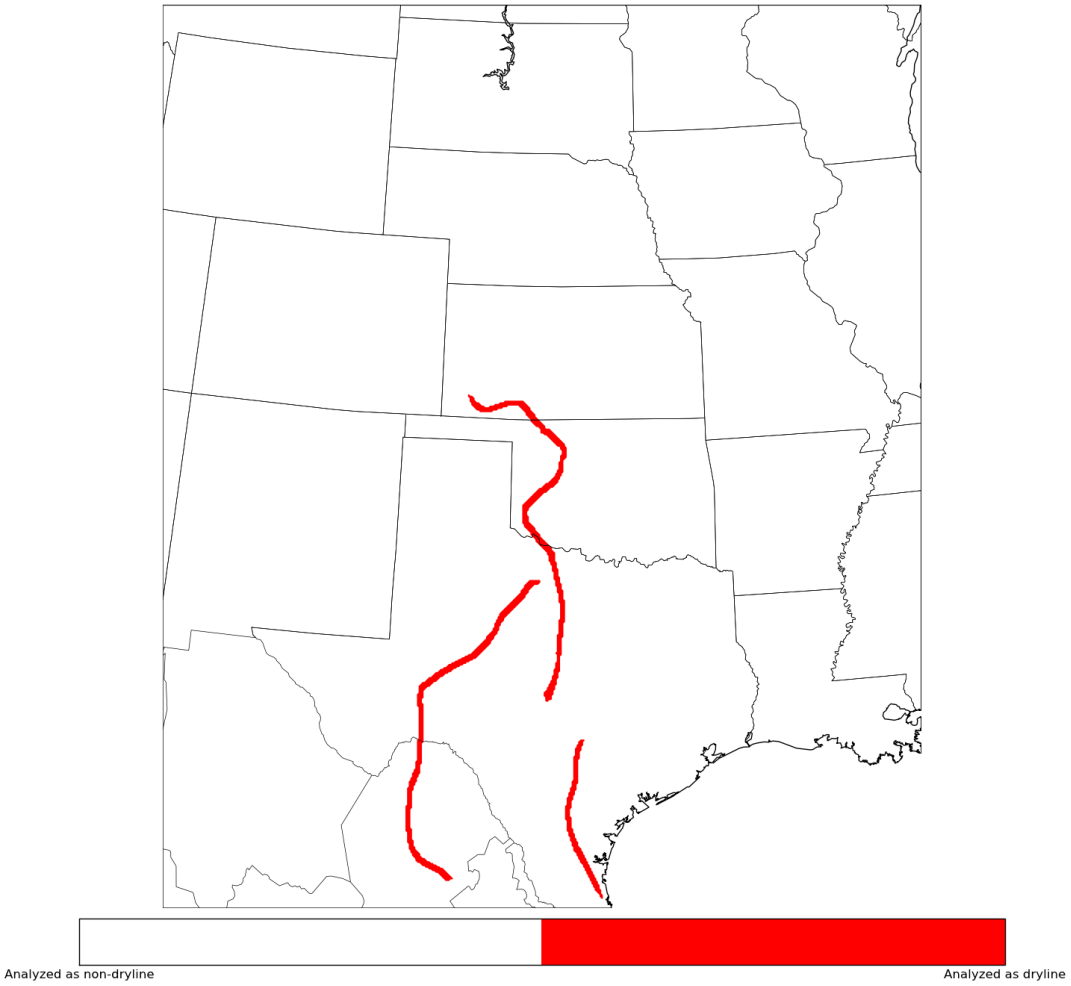


Figure 4.18: Reapplication of the vapor pressure gradient magnitude threshold [>1000 Pa/64km] applied to a binary AND of NMS and threshold masks using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 26 April 2012.

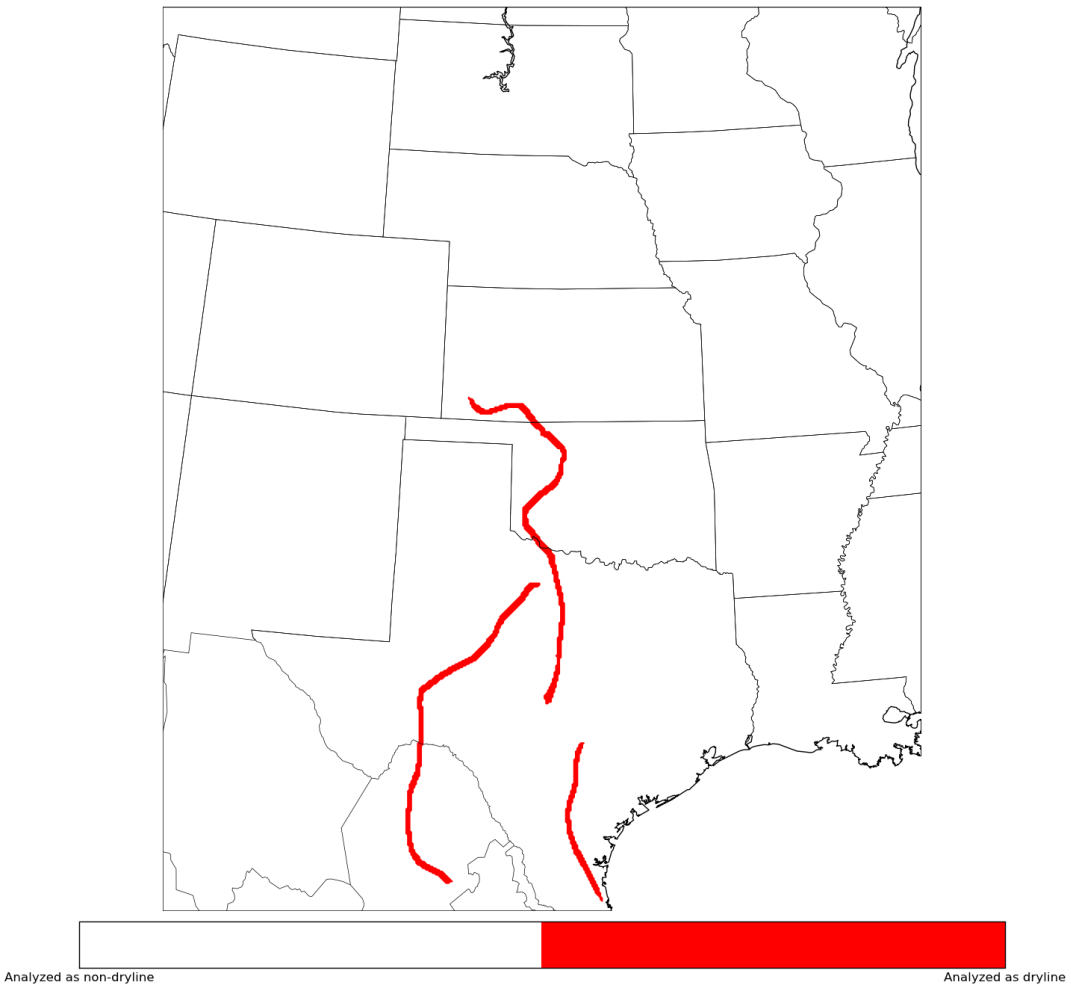


Figure 4.19: Reapplication of the size check (180km (height) x 220km (width) bounding box) applied to a binary AND of NMS and threshold masks (post-vapor pressure gradient magnitude threshold) using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 26 April 2012.

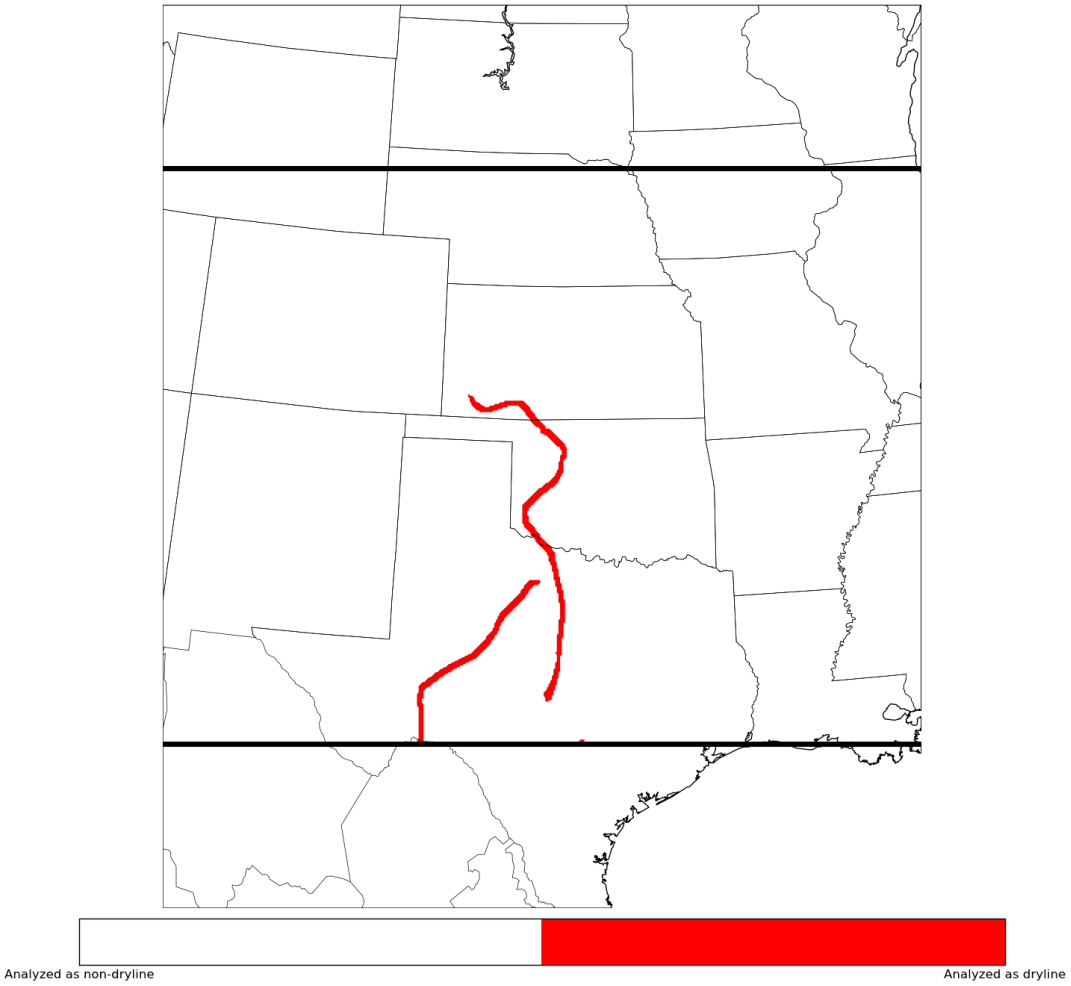


Figure 4.20: Final output of the dryline identification algorithm using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 26 April 2012. Black lines indicate ~30° N - ~42° N latitudinal bounds.

Chapter 5

Evaluation

This chapter provides an overview of the algorithm's performance. Section 5.1 begins with a quantitative assessment of the algorithm, examining both general statistical performance and specific situations in which the technique fails. It is followed by section 5.2, which describes the results of a preliminary application of machine learning to the algorithm output.

5.1 Quantitative assessment

5.1.1 Objective vs. subjective comparison

Traditional forecast (and analysis) verification methods such as False Alarm Ratio (FAR), Probability of Detection (POD), and others (discussed below) provide a technical understanding of an algorithm's performance. However, such assessments provide little understanding of why a forecast succeeded or failed. In contrast, object-based techniques provide information similar to a subjective examination by treating meteorological features as entities independent of the environmental field (Davis et al., 2006). The grouping of pixels into objects allows attributes to be assigned to specific features (e.g., the location, speed, and size of a storm) and directly compared to those of another object. For example, the path of a forecasted storm may differ significantly from its observed

counterpart's. By treating the two storms as objects, various distance statistics can be computed in addition to comparing other variables (e.g., maximum reflectivity, updraft intensity).

The ideal output of the algorithm outlined in this thesis forms the basis of an object-based identification system; however, cases in which the analyzed drylines are merged with other boundaries (or entirely non-dryline boundaries are identified) present a difficulty to its evaluation as such. The non-dryline objects will heavily influence any calculations in which they are present. For the purpose of the following evaluation, these features have been removed by only retaining identified regions falling within 30 km of the subjective analyses created by Coffer et al. (2013). The masking removes the majority of non-dryline features from subsequent examination, mitigating the errors that would otherwise be introduced. However, the process also limits the type of evaluations that can be performed. The fields remaining deal with position, both in distance between analyses and their relative locations.

Figure 5.1a shows a histogram of the distance between each objectively analyzed dryline pixel and the nearest dryline position from the subjective dataset developed by Coffer et al. (2013). Seventy five percent of the points from the objective dataset have a subjective data point within 10 km of their locations, and 90% within 15 km. These values indicate a high degree of agreement between the developed algorithm and the subjective data set, with over 90% of the identified dryline points (within the 30 km buffer) existing within four grid points of the "truth" dataset (4 km grid resolution). Figure 5.1b shows the same data, but with negative (positive) values indicating the objective analysis is further in the dry (moist) region relative to the subjective locations. Examination shows that there is very little dry/moist side bias present between the

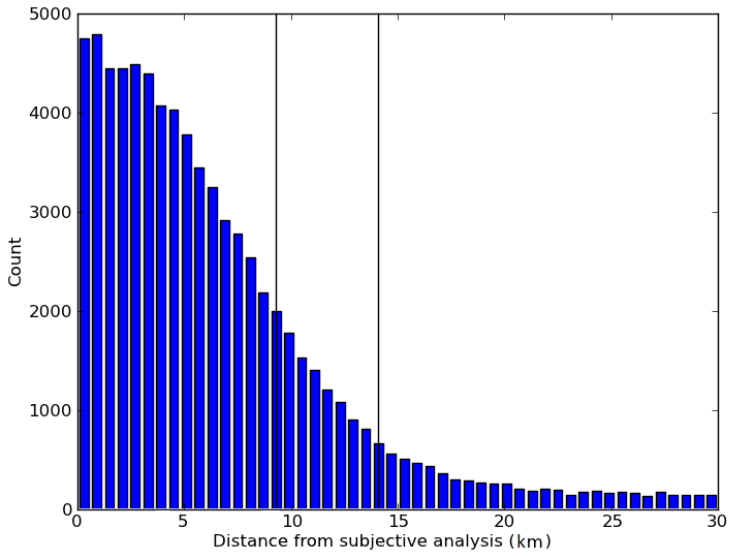
two analyses. It should be noted that the subjective dataset only contains a series of points that are linearly interpolated for comparison. It is possible that the more minute variations are partially due to this simplification.

Some of the larger distances in Fig. 5.1a are due in part to situations where the dryline algorithm follows a moisture boundary other than the subjective analyses (an extreme case can be seen in Figs. 5.2a and 5.2b). These cases are due to NMS necessarily following the strongest moisture boundary present (not to be confused with the removal of smaller variations along the same boundary discussed below) and ignoring the boundary identified by Coffer et al. (2013). While some of these cases are situations in which the objective technique was incorrect, the majority occur when the disagreement could otherwise be explained by differences in subjective opinion.

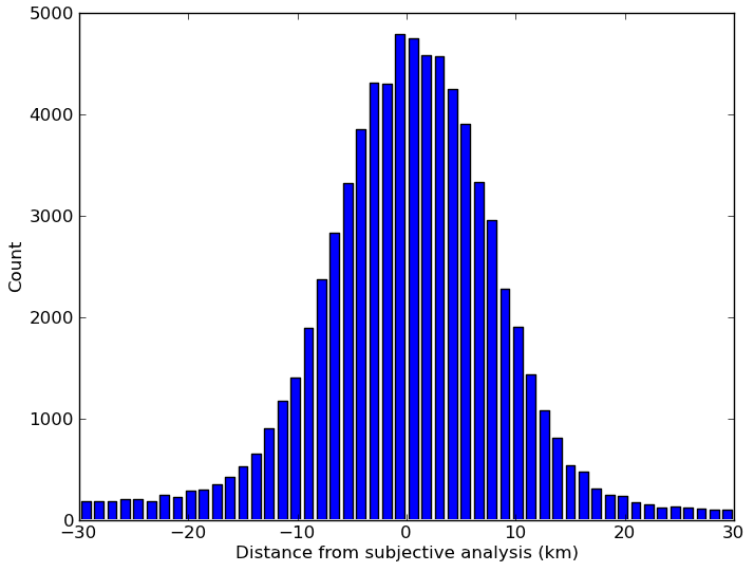
Another cause of the larger distances seen in Fig. 5.1a may be the increased smoothing factor used by NMS to prevent deviations from the true maximum specific humidity gradient. While the process removes small-scale features, it also smooths the overall curvature of the true moisture boundary. The more minute variations of the path that are observable with less smoothing are lost. For example, the combined specific humidity and dewpoint gradient threshold mask seen in Fig. 5.3a shows the presence of two slightly separated moisture boundaries from the Texas panhandle to west Texas. The two features overlap (i.e., run parallel and next to one another) in west-central Texas and are connected by a small region of high moisture gradient. While the separated regions are likely a more accurate representation of the dryline (or drylines), the smoothed field used by NMS does not maintain the separation (Fig. 5.3b). When combined with the threshold mask, the NMS region passes through the

connecting strand, resulting in the final output containing a single, continuous boundary (Fig. 5.3c).

While not shown here, the dryline algorithm also has a tendency to extend the analyzed regions further along the moisture boundary than the subjective analyses do (i.e., the objective drylines are typically longer). Unfortunately, these differences in length cannot not be calculated in their present form beyond 60 km due to the buffering process (although shorter objectively analyzed drylines can be found). The lengthening is likely due to the combination of the initial thresholds and NMS. The lines created by NMS rely on the threshold mask to provide a stopping criteria. Without it, the lines extend as far as possible while following the across-boundary gradient maximum regardless of gradient intensity. However, the initial thresholds are optimistic (i.e., intentionally low), ensuring that dryline regions are retained at the cost of keeping additional features. While the subsequent vapor pressure gradient and size thresholds aid in removing disconnected features, they cannot affect extensions attached to the dryline regions. When the threshold mask is applied to the NMS output, the lines are allowed to travel into the lower gradient regions, resulting in overly long drylines. A second thresholding has been examined for use in reducing this effect; however, we were unable to achieve consistent results. A variety of iterative endpoint removal techniques were also investigated, with similar effects.

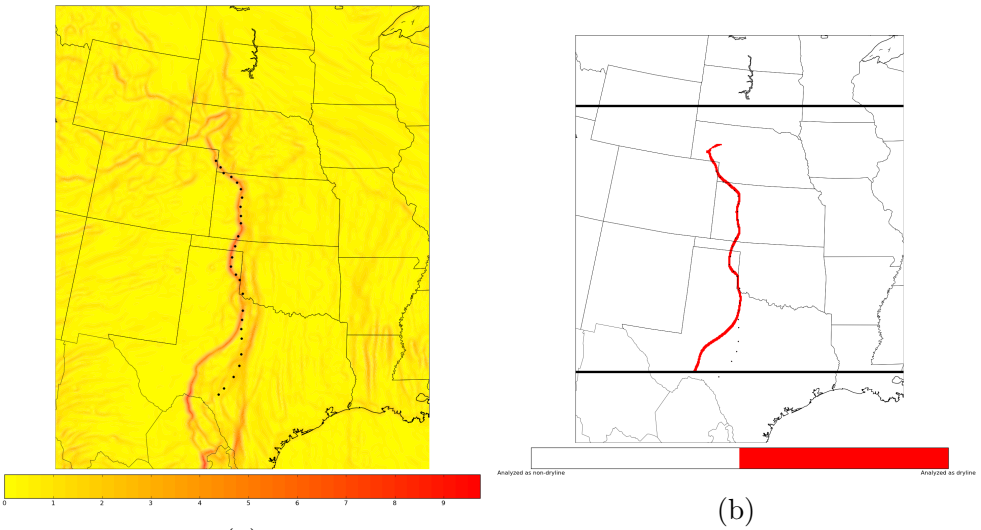


(a)



(b)

Figure 5.1: (a): Histogram of the distance between objective and subjectively analyzed points after the application of a 30 km buffer mask. The left and right black lines denote that 75% and 90% of the observed points fall within 10km and 15km of a subjective point, respectively. (b): Histogram of the distance between objective and subjectively analyzed points after the application of a 30 km buffer mask. Negative (positive) values indicate the objective analysis is placed further in the dry (moist) region than the subjective analysis.

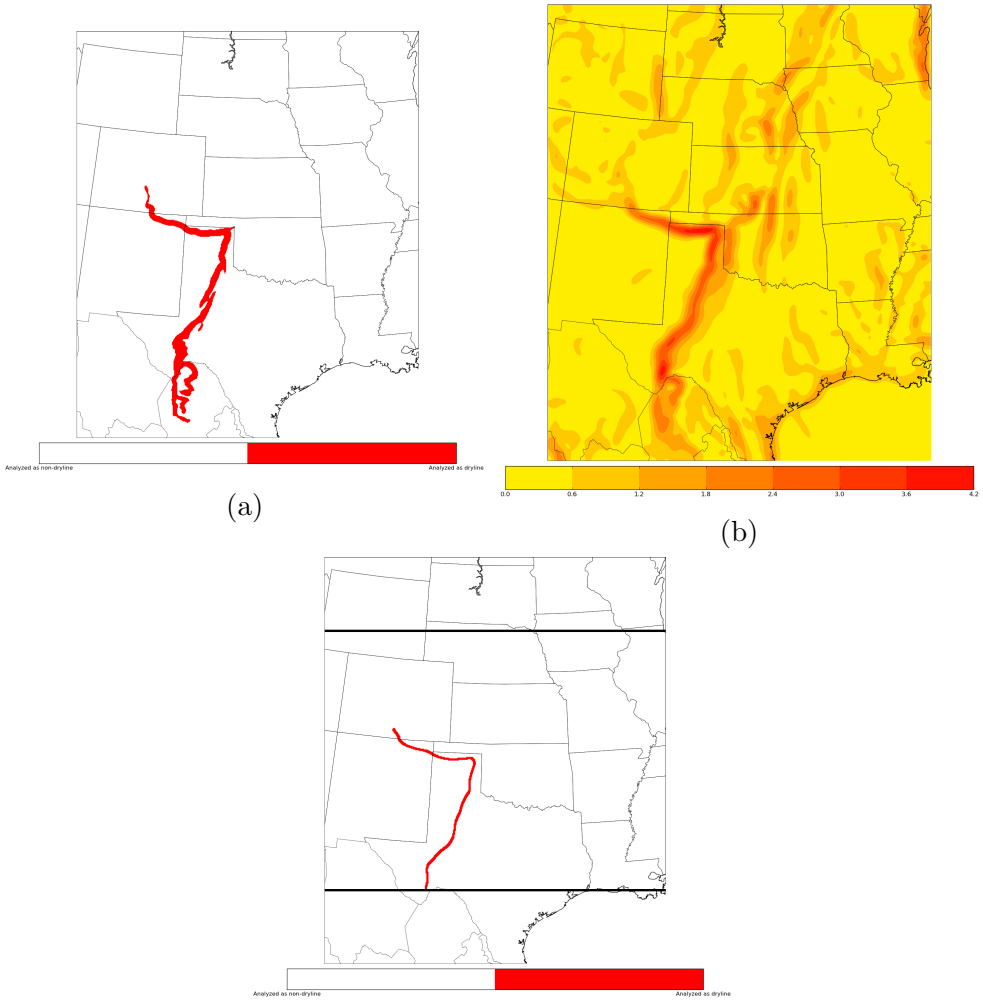


(a) (b)

Figure 5.2: Calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 01 May 2008: (a): Specific humidity gradient magnitude (g/kg/64km). (b): Final output.

5.1.2 Statistical assessment

Statistical evaluation of binary forecasts (or analyses) are commonly performed using a contingency table (Wilks, 1995). The table provides a set of values that describe a forecast's ability to predict a given variable or feature. Traditional contingency tables are created by directly comparing each location on the forecast grid to the corresponding truth data. However, these implementations result in extremely high penalties for errors in location in addition to often over-penalizing fine scale structures (e.g., drylines). An object-based adaptation of the traditional implementation can be formulated by treating groups of pixels as entities separate from the environment, allowing for improved assessment (e.g., Ebert and McBride, 2000). The modification removes the above sensitivities in addition to allowing more complex evaluations of contingency table values. For the purpose of this study, objects are defined as contiguous (or appropriately broken groups of) regions identified by the dryline algorithm.



(a)

(b)

(c)

Figure 5.3: Calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 20 May 2012: (a): Size check (final output of the initial thresholding). (b): Specific humidity gradient magnitude (g/kg/64km) (pre-processed using $\sigma = 6$ Gaussian filter). (c): Final output.

These regions are then compared to the 2011 Coffer et al. (2013) analysis as well as the independent dataset described in chapter 3.

The contingency table is comprised of four values (see Fig. 5.4). True positives (TP) occur when a dryline is present in both the objective and subjective analyses. False positives (FP) denote an incorrect identification by the objective output. A dryline that is not identified by the algorithm is categorized as a false negative (FN), while the lack of identification of any non-dryline boundary is a true negative (TN). The object-based implementation of the evaluation allows for TP, FP, TN, and FN values all to be present on the same day. In the case that multiple objectively identified drylines identify the same subjective dryline, they are treated as a single object. The associated TP, FP, and FN values are assessed as the associated fractional area of the combined objective and subjective analyses (e.g., if two or more objective drylines capture half of a subjective dryline but also capture half of an equally long cold front, the resulting values would be TP: 1/3, FP: 1/3, FN: 1/3). Any FPs not associated with a subjectively analyzed dryline are counted as separate events. It should be noted that this implementation results in double penalization of cases in which the algorithm follows a different moisture gradient than that of Coffer et al. (2013), as both a FP and FN must be assessed (see Fig. 5.2b). In addition, the entity based analysis creates difficulties in the assessment of TNs. Ideally, TNs would be determined by identifying all fronts, storm outflow boundaries, etc. present in the test data and counting each that was not objectively identified as a dryline. However, the lack of any specific measures to determine how non-dryline features should be assessed allows for widely varying scores depending on the the criteria selected. Therefore, the majority of scores using TNs have

been avoided here. The impacts of this limitation on the statistical assessment are described where appropriate.

$$FAR = \frac{FP}{TP + FP} \tag{5.1}$$

$$POD = \frac{TP}{TP + FN} \tag{5.2}$$

The false alarm ratio (FAR, Eq. 5.1) describes how frequently the algorithm misclassifies another boundary as a dryline. The score ranges from 0 to 1, with higher values signifying a greater misclassification rate. FAR is sensitive to event frequency in that rare events can provide good scores by always assuming the negative case. For example, if an event only occurs one time out of a thousand and is always assumed to not occur, the FAR is still .001. It is therefore commonly used in conjunction with the probability of detection (POD, Eq. 5.2) score. POD indicates the likelihood that a dryline is detected by the algorithm. It is defined as the fraction of correct objectively identified drylines to all subjectively identified drylines. The score has a range of 0 to 1, with higher values indicating greater detection probability. The key drawback of POD is that it can be likewise manipulated by consistently identifying regions as drylines. The lack of penalty from FPs means that a perfect score can be obtained despite an overall inaccuracy of the analysis. However, this is compensated for by FAR (and vise versa) such that an overly optimistic (pessimistic) algorithm may have good POD (FAR), but will also have a poor FAR (POD).

$$CSI = \frac{TP}{TP + FN + FP}, \tag{5.3}$$

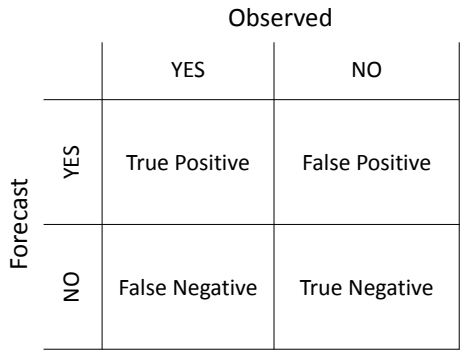


Figure 5.4: Contingency table. Forecast values correspond to the objective algorithm output while observations represent the subjective analysis performed by the author.

$$FBIAS = \frac{TP + FP}{TP + FN}, \tag{5.4}$$

$$GSS = \frac{TP - RAND}{TP + FP + FN - RAND}, \tag{5.5}$$

where

$$RAND = \frac{(TP + FN) \times (TP + FP)}{TP + FP + TN + FN}. \tag{5.6}$$

The critical success index (CSI, Eq. 5.3) provides a measure of the algorithm's accuracy while ignoring TNs and ranges from 0 to 1, where 0 indicates no skill (Schaefer, 1990). CSI only considers situations in which something was, or should have been analyzed as a dryline. The score has a slight dependence on the climatology of the event being examined. This can be accounted for by using the Gilbert Skill Score (GSS, Eq. 5.5) (also known as the Equitable Threat Score). GSS acts in much the same fashion as CSI, but attempts to remove those selections that would occur due to random chance (although this is only an approximation) (Hogan et al., 2010). The output of GSS ranges from

-.33 to 1, where 0 indicates no skill compared to random chance and 1 is the ideal output. The score does require TNs for its calculation (see Eqs. 5.5 and 5.6) but is not strongly influenced by it in this case (see below). Frequency bias (Eq. 5.4) provides the ratio of analyzed dryline events to observed dryline events. The score ranges from 0 to infinity, where scores less (greater) than 1 indicate too few (many) analyses relative to the number of observed drylines.

The scores shown in Table 5.1 are calculated from contingency table values assessed through a visual comparison of the independent 2012 dataset and the resulting dryline algorithm output by the author using the criteria outlined above. The values suggest that the technique is overly optimistic. The high POD of 0.864 indicates that very few drylines are not found, which is essential for any form of operational use. However, the other scores suggest that a large number of dryline regions are being incorrectly identified (i.e., other features are being included). CSI indicates the algorithm is only slightly skilled; however, the score may underestimate the algorithm's performance due to the relative infrequency of drylines in the dataset. The GSS value of 0.209 shows that the technique performs better than random chance, but not extremely well. While the lack of a defined value for TNs in this analysis results in some uncertainty

POD	0.864
FAR	0.717
FBIAS	3.05
CSI	0.271
GSS	0.209

Table 5.1: Skill score values for the 2012 independent dataset (see in-text description for a note on GSS calculation).

in the GSS, testing showed that the score only varied by ~.03 when using TNs ranging from 30 to 100. Regardless, the CSI and GSS values are far from perfect. The frequency bias score of 3.05 suggests that too many presumed drylines are being identified by the algorithm, which is confirmed by FAR. The high FAR of 0.717 is due to non-dryline features being independently identified or merged with drylines by the algorithm and incorporated into the output. It should be noted that the independent 2012 dataset has a slightly poorer POD, but improved FAR and other scores when compared to the evaluation of the 2011 Coffer et al. (2013) data shown in Table 5.2a (see section 5.2.2). The former difference is small enough that it can be explained by year to year variations in the datasets; however, the latter improvements are likely caused by a decrease in FPs in the 2012 dataset. This drop is due to a combination of the lack of "subjective" differences in the analyses (mentioned in section 5.1.1) and the absence of penalization due to the identification of secondary moisture boundaries.

Subsequent investigation found that there are three key situations with which the algorithm consistently has difficulty. These cases are discussed in the following sections.

5.1.2.1 Intersecting boundaries

There are very few cases in the test set in which a cold or warm front is mistaken for a dryline in its entirety. More frequently, the intersection of frontal and dryline boundaries results in their partial merger. Such a situation is shown in Fig. 5.5 on 03 April 2012. The identified region stretching from eastern New Mexico to the eastern portion of the Oklahoma panhandle is part of a cold front, while the remainder of the line is a dryline. Fig. 5.6 shows the strong

moisture gradient merging with that of the cold front in the eastern portion of the Oklahoma panhandle, overlaying the thermal gradient seen in Fig. 5.7. The cold front boundary continues northeast along the correctly unidentified weaker moisture gradient through Kansas, South Dakota, and into Iowa. This is supported by Fig. 5.8 which shows the front to be associated with a line of pressure minima, as would be expected by frontal theory (Bluestein, 1993). The strong difference in moisture gradient magnitude along the frontal boundary is caused by the differing characteristics of the two air masses that form the dryline. The wind field (Fig. 5.9) shows the air behind the front all flows from a single source region. The moisture content of this air is high enough that its intersection with the dry air to the southwest results in a large moisture gradient magnitude. However, the specific humidity gradient magnitude along the eastern half of the front is much smaller due to the relatively high moisture content of the opposing air mass to the southeast. When evaluated by NMS (Fig. 5.10), the algorithm follows the stronger moisture gradient region, merging the dryline with the western portion of the front. Combined with the original threshold masks, the NMS output results in the vapor gradient magnitude threshold check removing the disconnected eastern half of the front, while retaining that of the dryline / cold front merger. In essence, the difference in moisture content of the two air masses forming the dryline can result in differing moisture gradient magnitudes along the boundary separating a tertiary air mass. The stronger gradient magnitude portion of the non-dryline boundary may then be merged with that of the dryline instead of following along its weaker section. The high

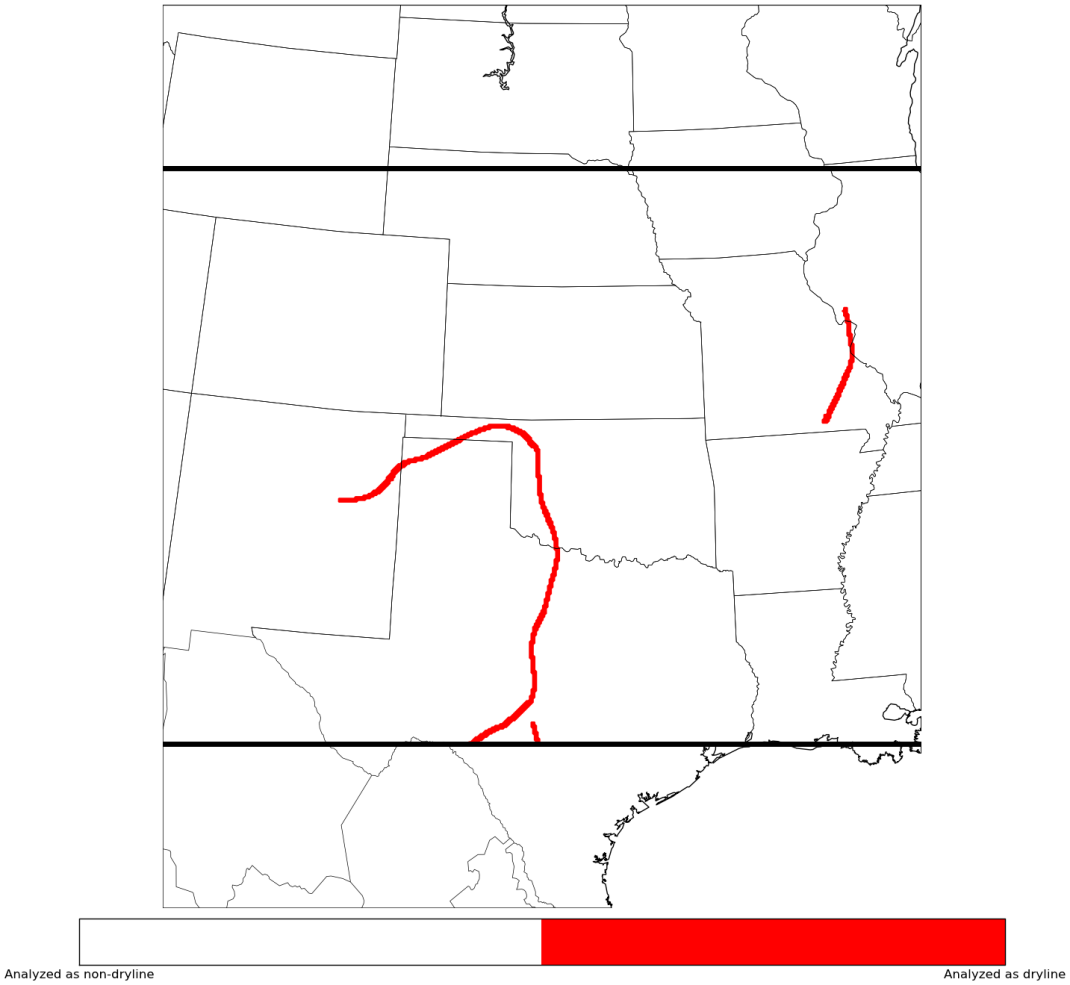


Figure 5.5: Final output calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 03 April 2012.

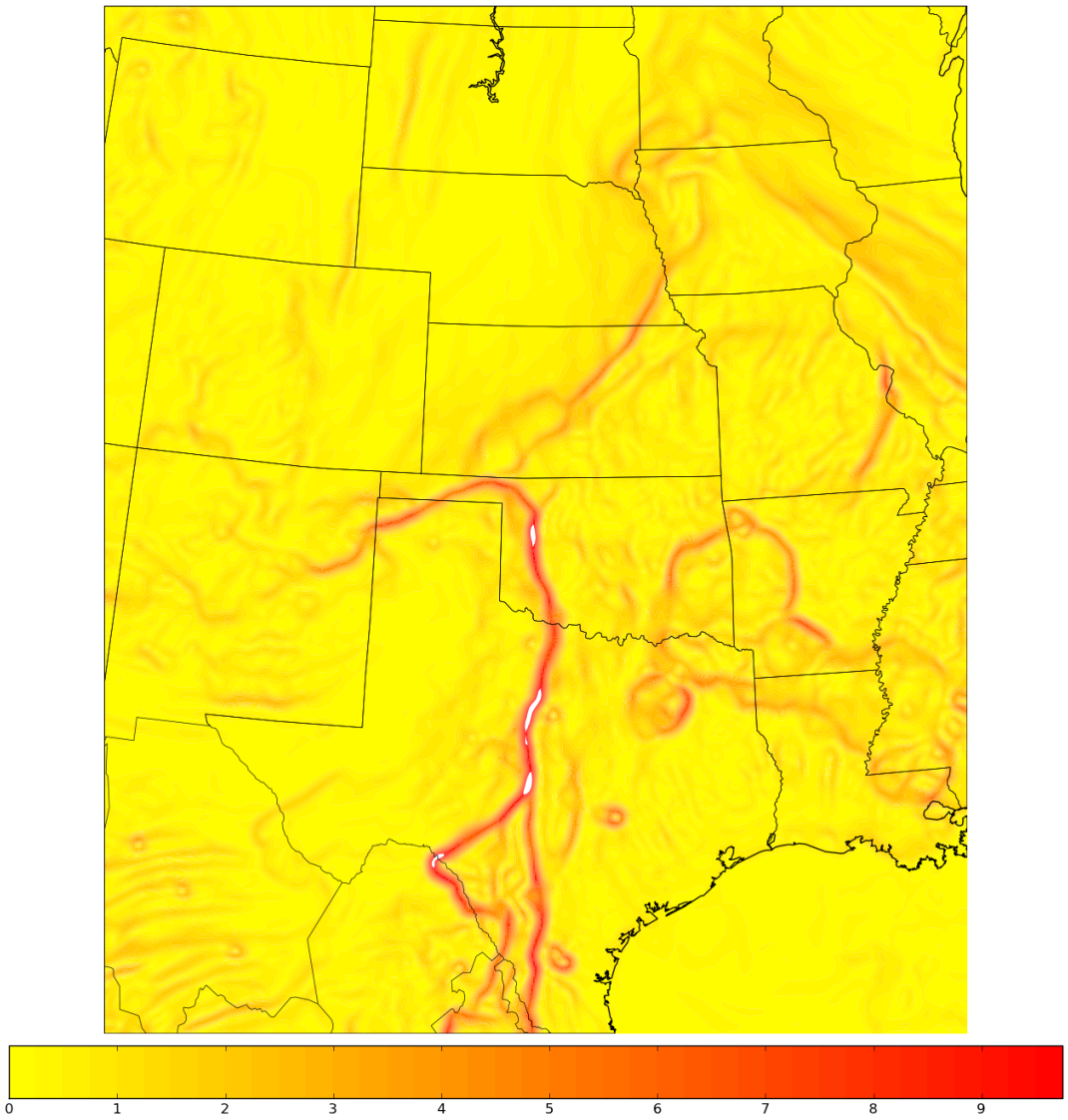


Figure 5.6: Specific humidity gradient magnitude (g/kg/64km) calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 03 April 2012.

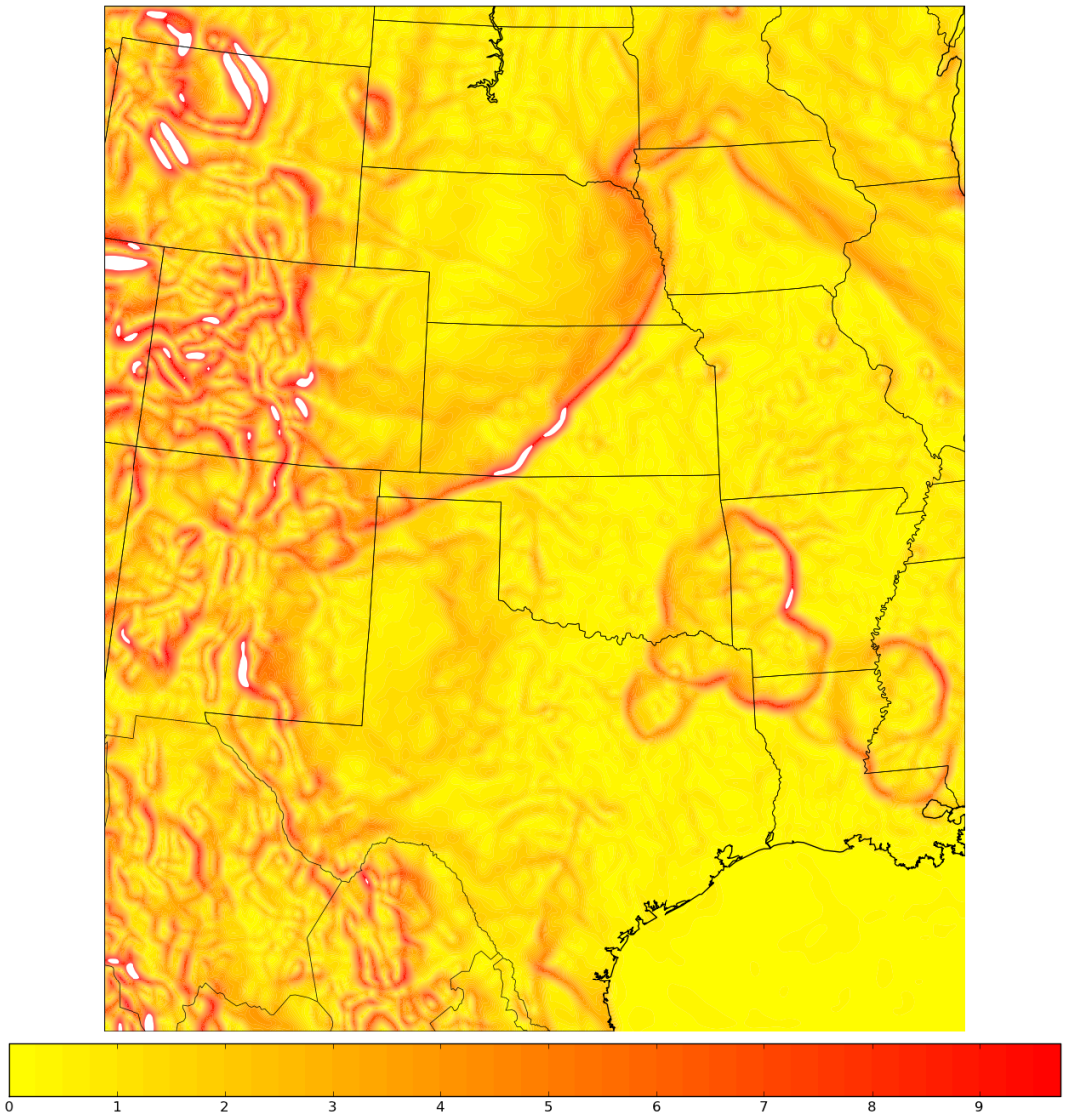


Figure 5.7: Virtual potential temperature gradient magnitude (K/64km) calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 03 April 2012.

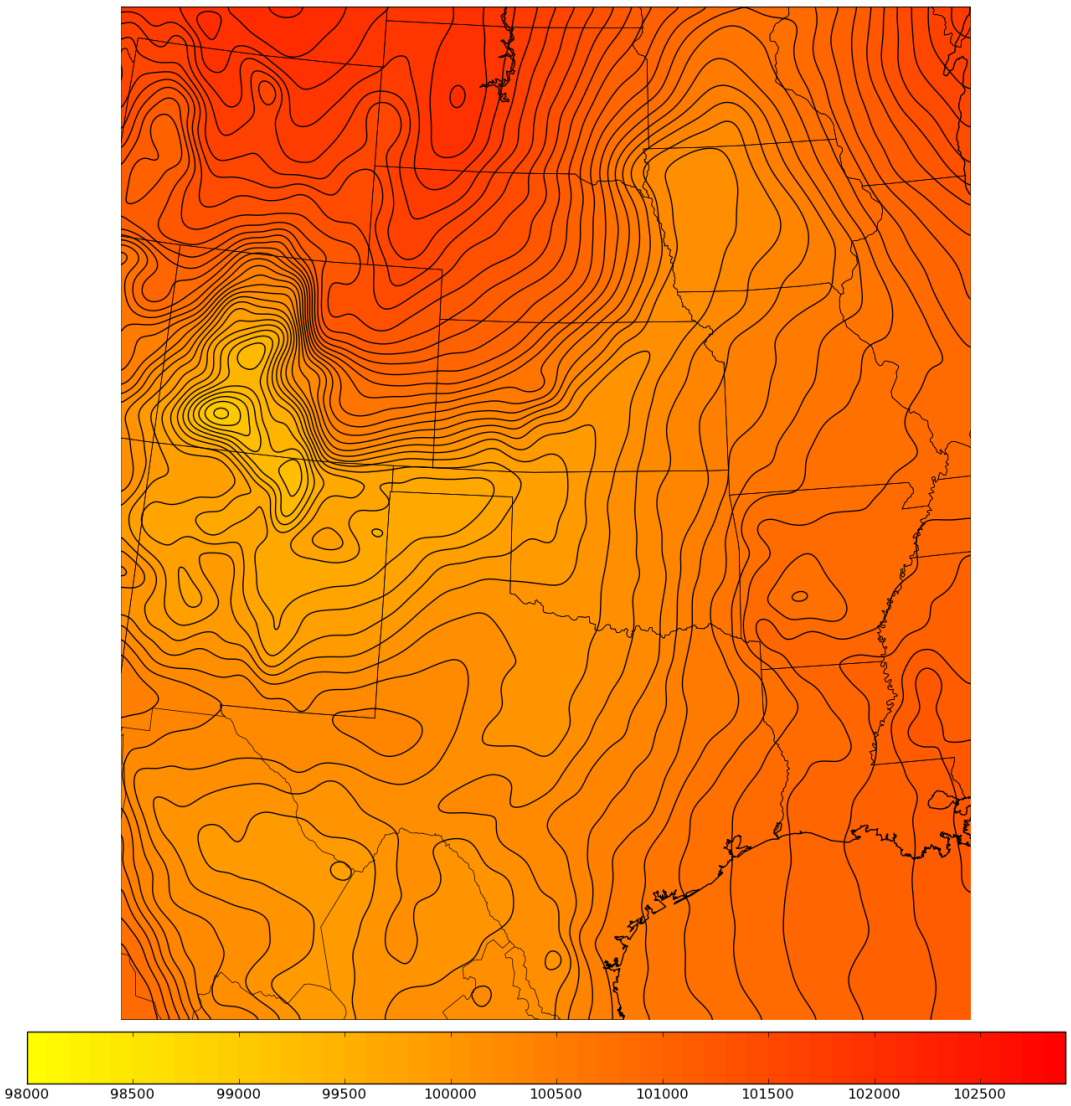


Figure 5.8: 24 hour forecast of surface pressure (Pa), contoured every 100 Pa (1 mb), from the NSSL-WRF valid 0000 UTC 03 April 2012.

moisture gradient magnitude and increased smoothing allows its inclusion by the thresholding and NMS processes, respectively.

5.1.2.2 Storm outflow

The algorithm also has difficulty dealing with convection occurring ahead of a dryline. The precipitation resulting from these storms increases the amount of moisture present at the surface. Larger or longer lived storms create significant regions of relatively high moisture that expand outward and meet with the environmental air. The intersections of the two air masses sometimes results in regions of strong moisture contrast similar to those exhibited by drylines. This is similarly true of air from a storm's rear flank downdraft (RFD), in which dry air from aloft has been transported to the surface. In these situations the algorithm will sometimes capture the resulting boundaries, especially those near a dryline as the intersecting ridges of moisture gradient magnitude may cause NMS to follow the outflow boundary instead of continuing along the true dryline. Such a situation occurred on 12 April 2012 several hours after the initiation of a large amount of convection along the dryline. Examination of the output shows a small, curled boundary protruding from the true dryline (Fig. 5.11). This extension is formed by a small pocket of dry air present beyond the well defined dryline (Fig. 5.12) interacting with the outflow of the nearby storms (Fig. 5.13). Whether the dry region is due to a jump in the dryline or dry air being mixed down by the storms is unclear. The area is larger than would be expected of a RFD and the pressure and thermal fields (Figs. 5.14 and 5.15) do not show indications of cold pool development at that location. Given the intertwining nature of the dry and moist air seen in southeastern New Mexico (Fig. 5.12), it

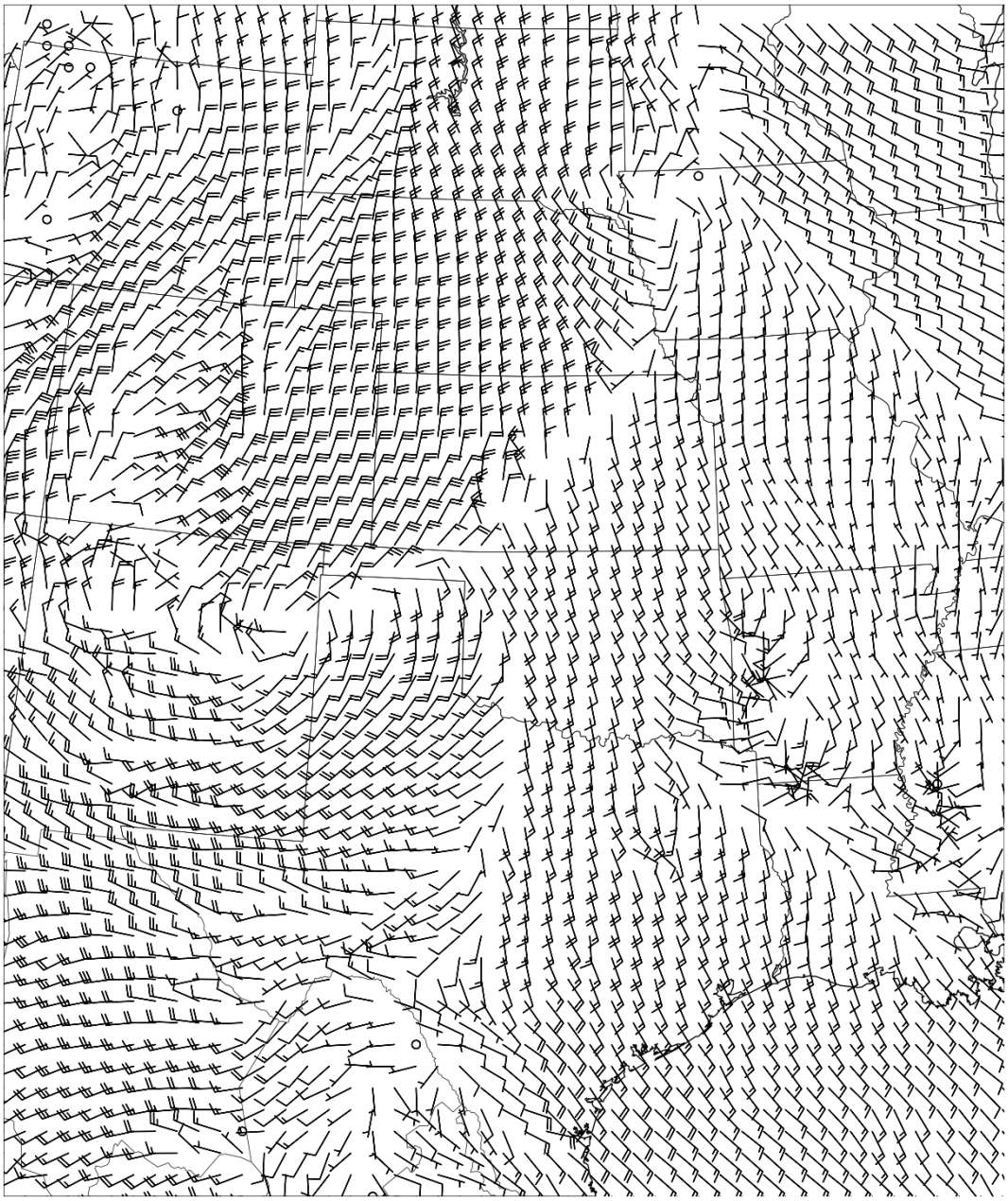


Figure 5.9: 24 hour forecast of 10 m wind speed and direction from the NSSL-WRF valid 0000 UTC 03 April 2012. Barb orientation indicates wind direction. Small flags indicate wind speeds of 5 kts, and larger flags 10 kts. Circles indicate calm winds.

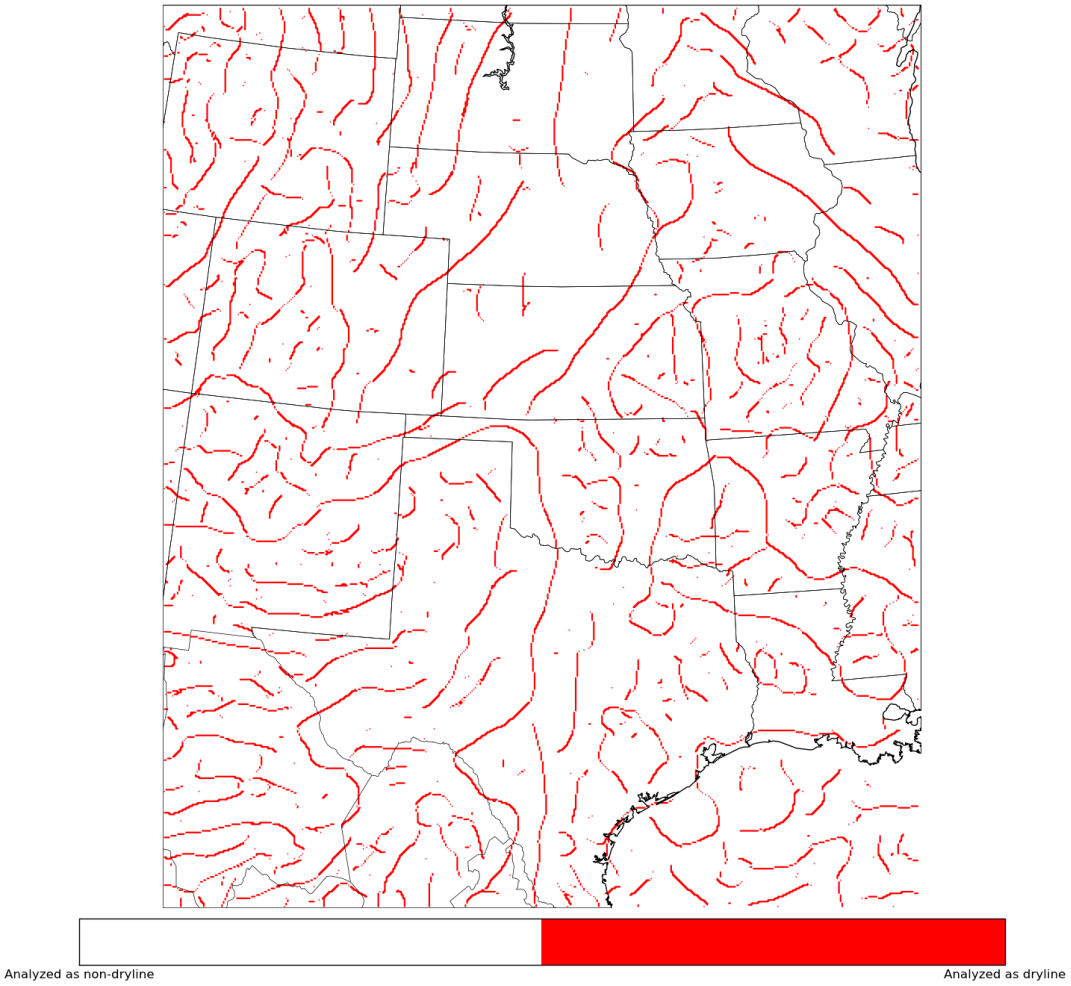


Figure 5.10: Application of NMS to specific humidity gradient magnitude ($\sigma = 6$) calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 03 April 2012.

may be a secondary moisture boundary that is in the process of being eroded by the surrounding storms.

5.1.2.3 Quasi-drylines

Drylines are typically considered to be the intersection of a continental tropical (cT) air mass and a maritime tropical (mT) air mass. That is, relatively dry (continental - over land) and warm (tropical) air meeting relatively moist (maritime - over water) and warm (tropical) air. These distinctions have come about due to the east-west terrain slope and air mass source regions present in the U.S (see chapter 2.1). However, similar features can form elsewhere in the presence of a sufficiently strong moisture gradient. Dodd (1965) shows that while the strongest April - June climatological moisture gradient is oriented east-west, a weaker gradient is also present in the north-south direction. The weaker gradient is formed by the intersection of continental polar (cP) and mT air masses, with cP being relatively dry (continental) and cool (polar). In certain cases, the resulting moisture gradient becomes strong enough for the algorithm to capture. These weaker formations rarely retain enough strength to be detected on more than a single day, although the boundary itself may persist. These features exhibit a wind shift similar to those seen by other meteorological boundaries. An example of these structures is shown in Fig. 5.16 in which the left boundary is a true dryline and the right is the described region of strong moisture gradient. The eastern boundary lies along a pressure trough (Fig. 5.19), exhibiting a weak thermal gradient (Fig. 5.17) (the strong thermal gradient along the true dryline is due to storm outflow impinging upon the boundary) and

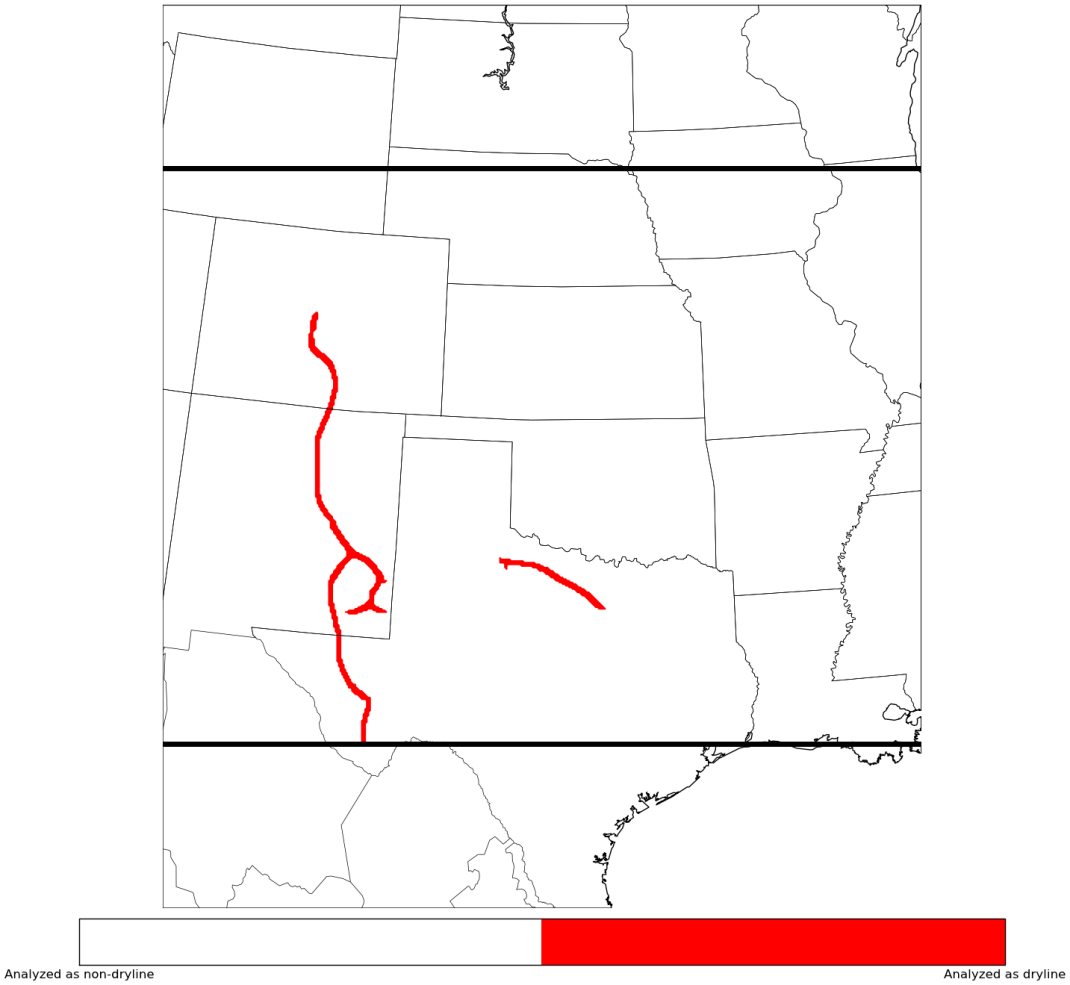


Figure 5.11: Final output calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 12 April 2012.

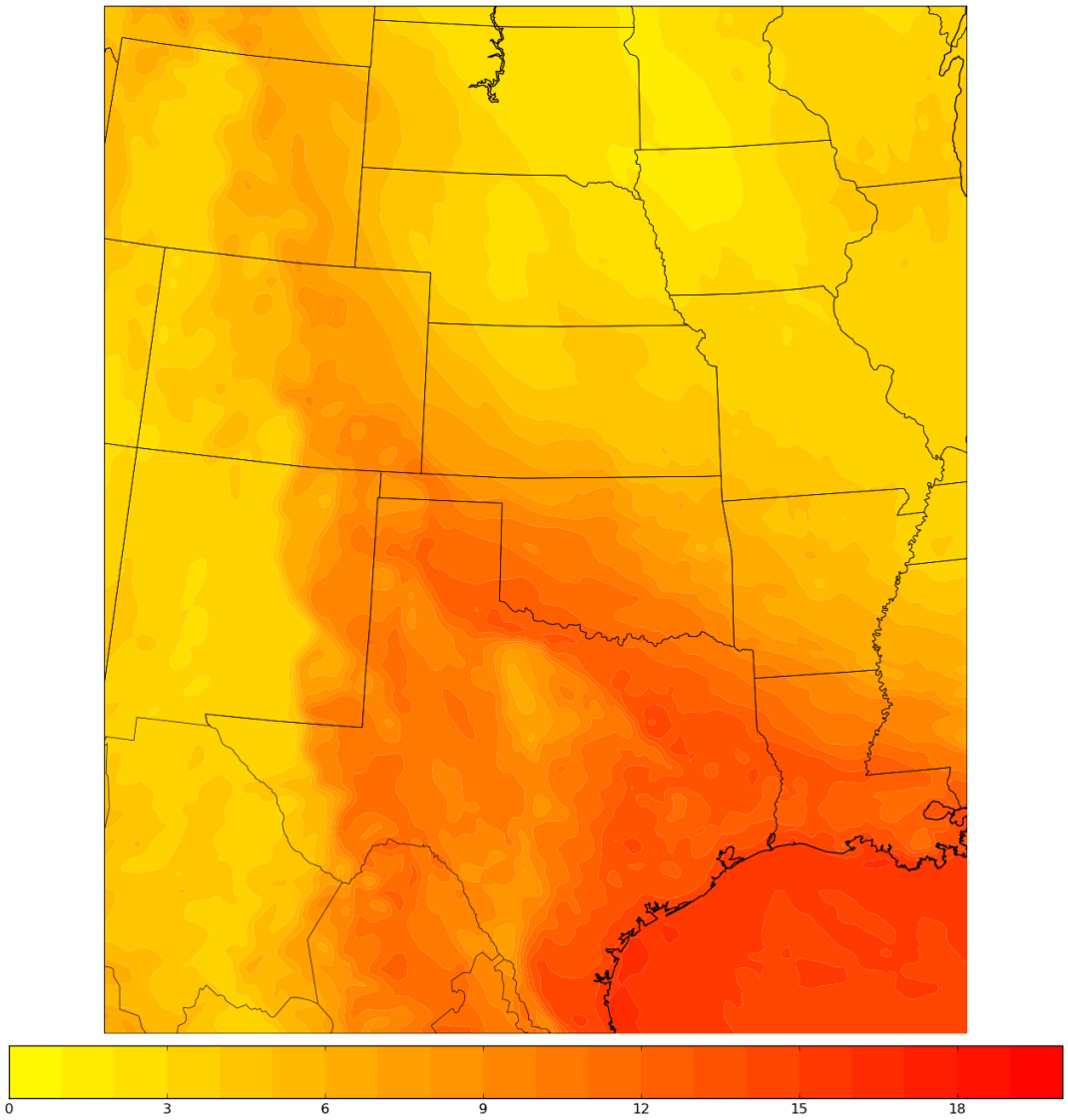


Figure 5.12: 24 hour forecast of specific humidity (g/kg) from the NSSL-WRF valid 0000 UTC 12 April 2012.

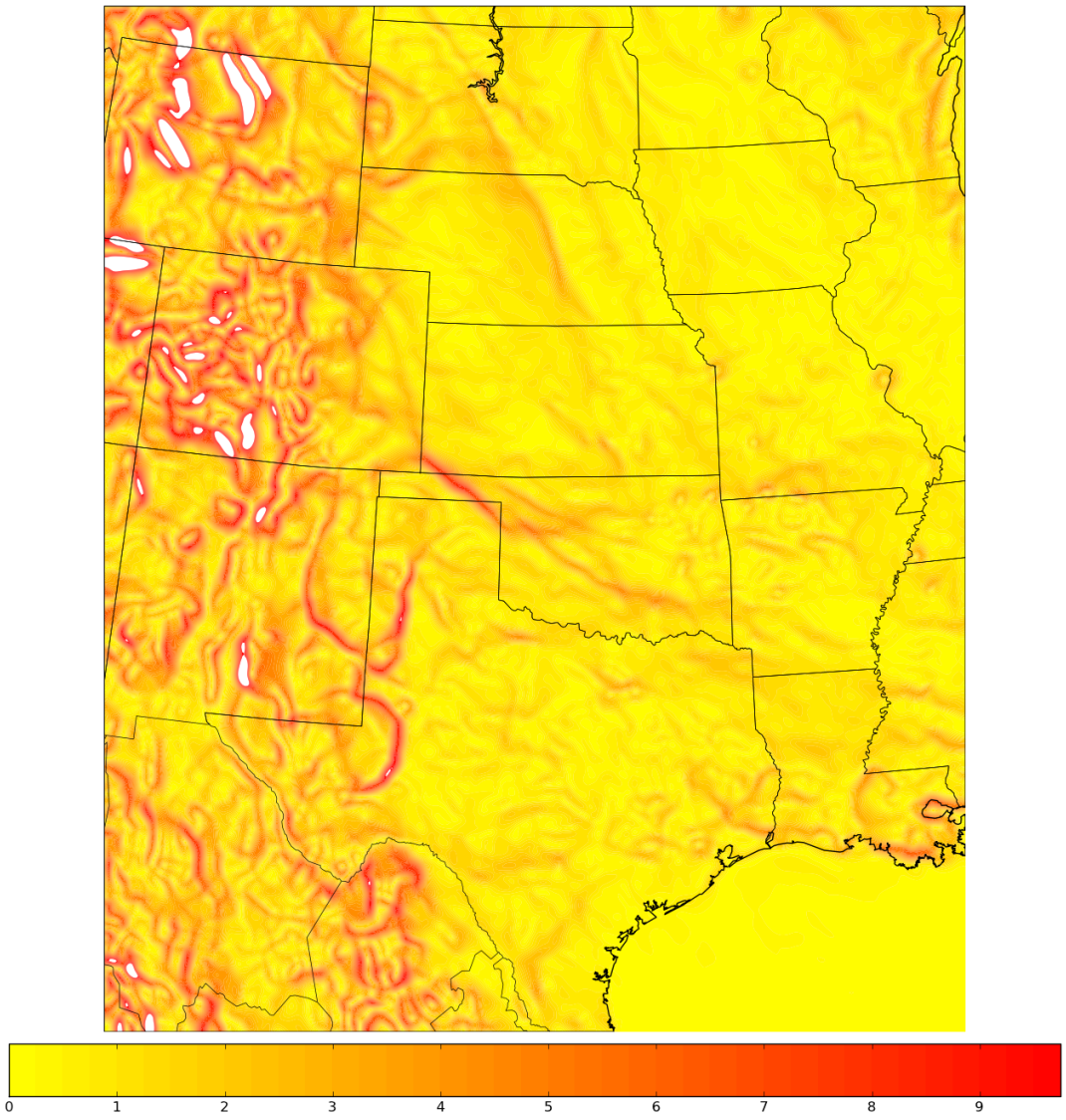


Figure 5.13: Virtual potential temperature gradient magnitude (K/64km) calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 12 April 2012.

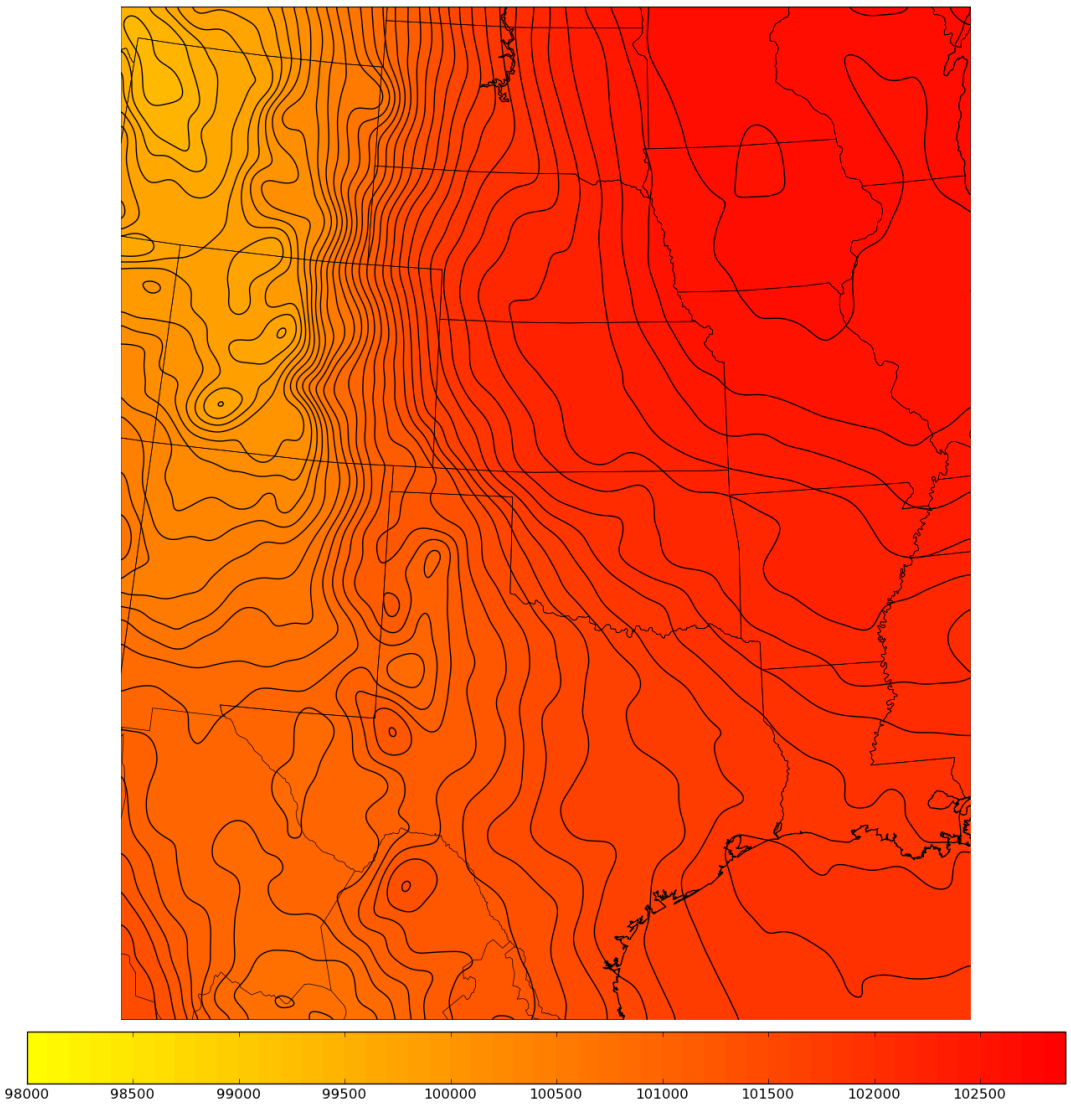


Figure 5.14: 24 hour forecast of surface pressure (Pa), contoured every 100 Pa (1 mb), from the NSSL-WRF valid 0000 UTC 12 April 2012.

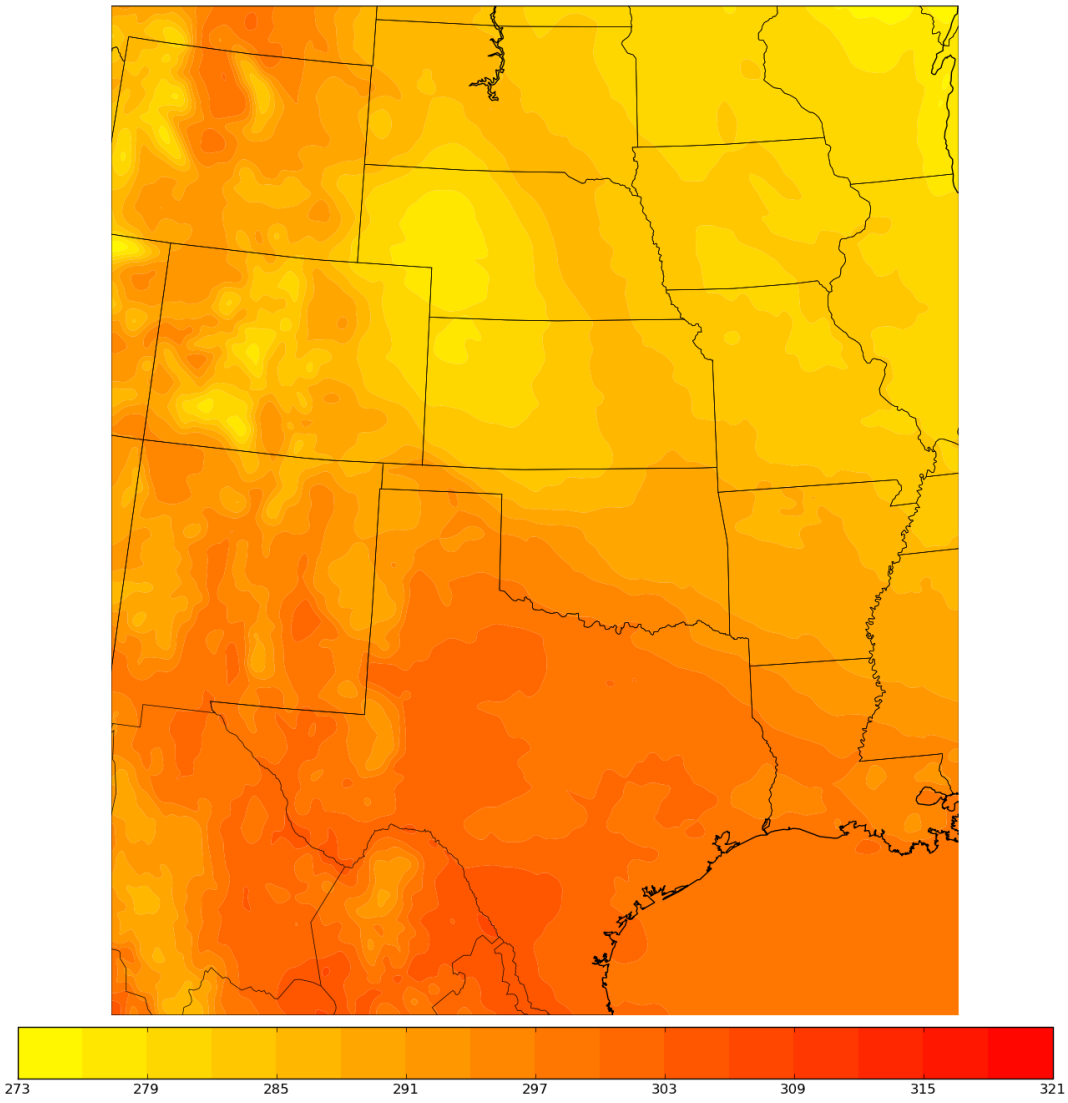


Figure 5.15: 24 hour forecast of virtual potential temperature (K) from the NSSL-WRF valid 0000 UTC 12 April 2012.

cyclonic (counterclockwise in the Northern Hemisphere) wind shift (Fig. 5.18) across the boundary. The structure of the feature is akin to a baroclinic trough as described by Sanders (1999); however, its strong moisture gradient suggests it is more closely related to an active dryline than a typical front. It should be noted that Sanders did not address moisture in his discussion.

5.2 Preliminary machine learning results

The key issue plaguing the current algorithm output is the excessive capture of non-dryline features. Subsequent examination of the algorithm output found that many of the FPs could be removed through simple thresholds (e.g., on the thermal gradient). Unfortunately, these methods were often only applicable in certain situations and would typically result in poorer outcomes when not valid (see appendices A.2 and A.3). However, these types of conditions are often ideal for use by machine learning techniques. In addition, the object-based nature of the algorithm's output allows for easy creation of datasets for use by said techniques. We therefore applied the random forest (RF; Breiman, 2001) algorithm (100 trees, maximum depth 5) to the objectively analyzed drylines using the Waikato Environment for Knowledge Analysis (WEKA, Hall et al., 2009) toolkit.

RFs are a machine learning technique that uses an ensemble of decision trees (another machine learning structure) to output a predictand for a given sample. Decision trees are generally formed by recursively splitting their training dataset along an attribute threshold chosen to maximize the reduction of information remaining in the dataset. The process is repeated until a predictand is selected

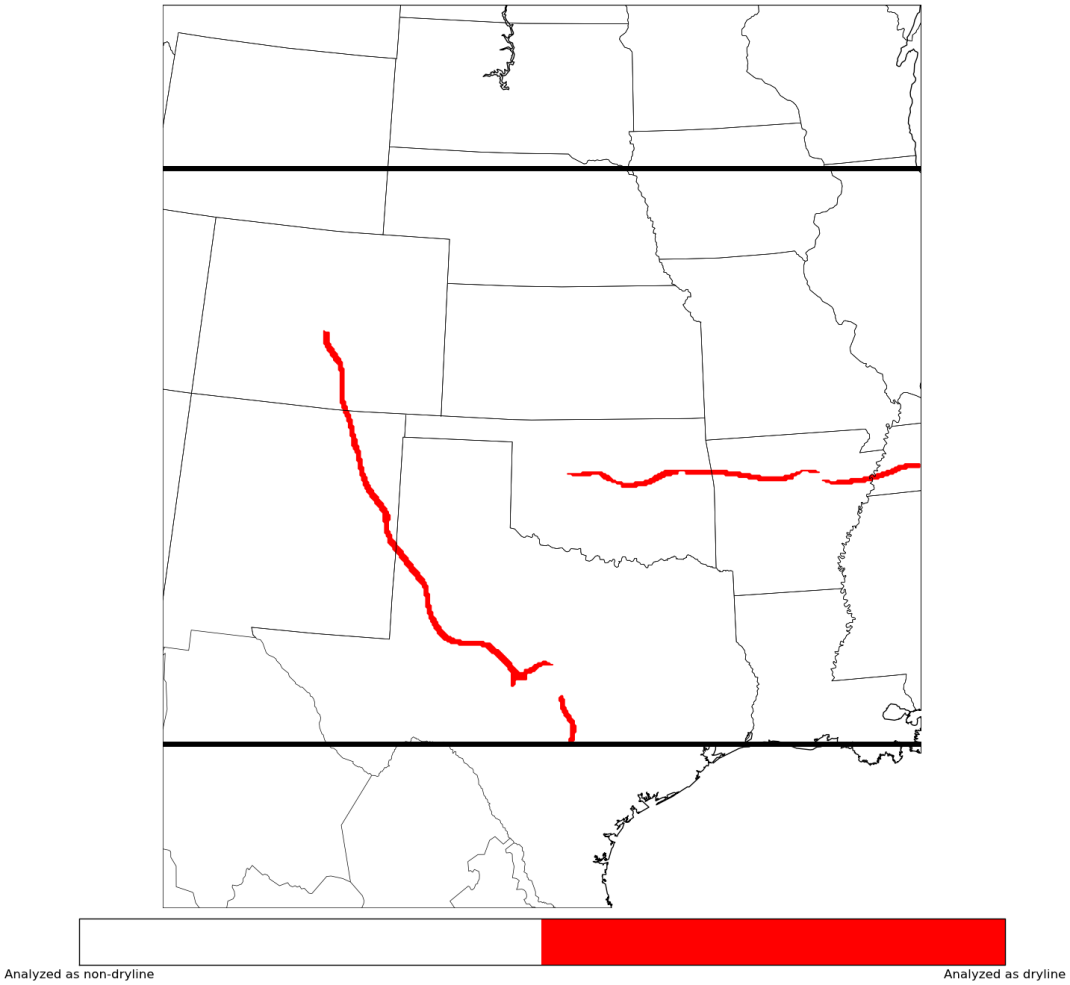


Figure 5.16: Final output calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 27 April 2012.

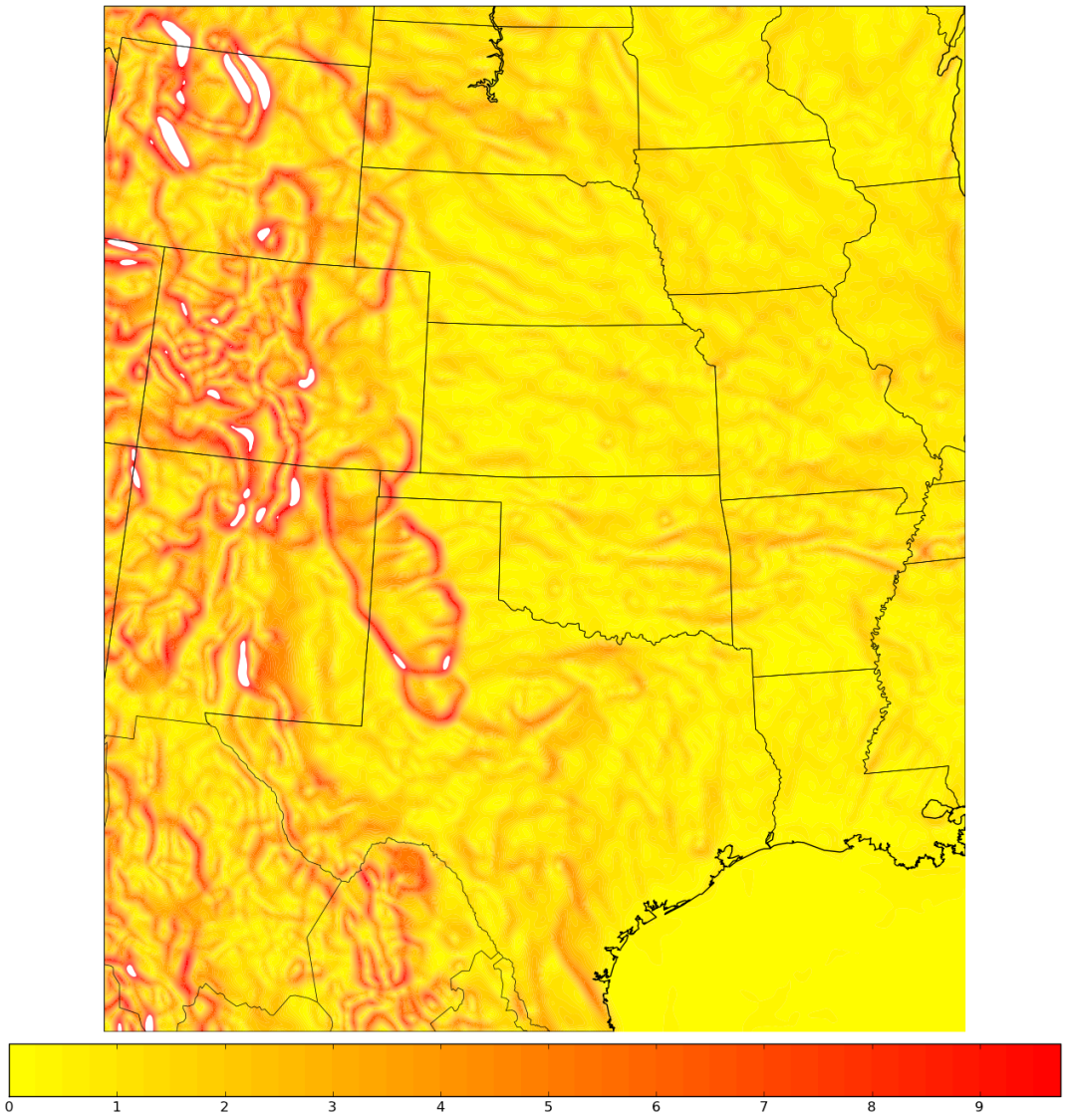


Figure 5.17: Virtual potential temperature gradient magnitude (K/64km) calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 27 April 2012.

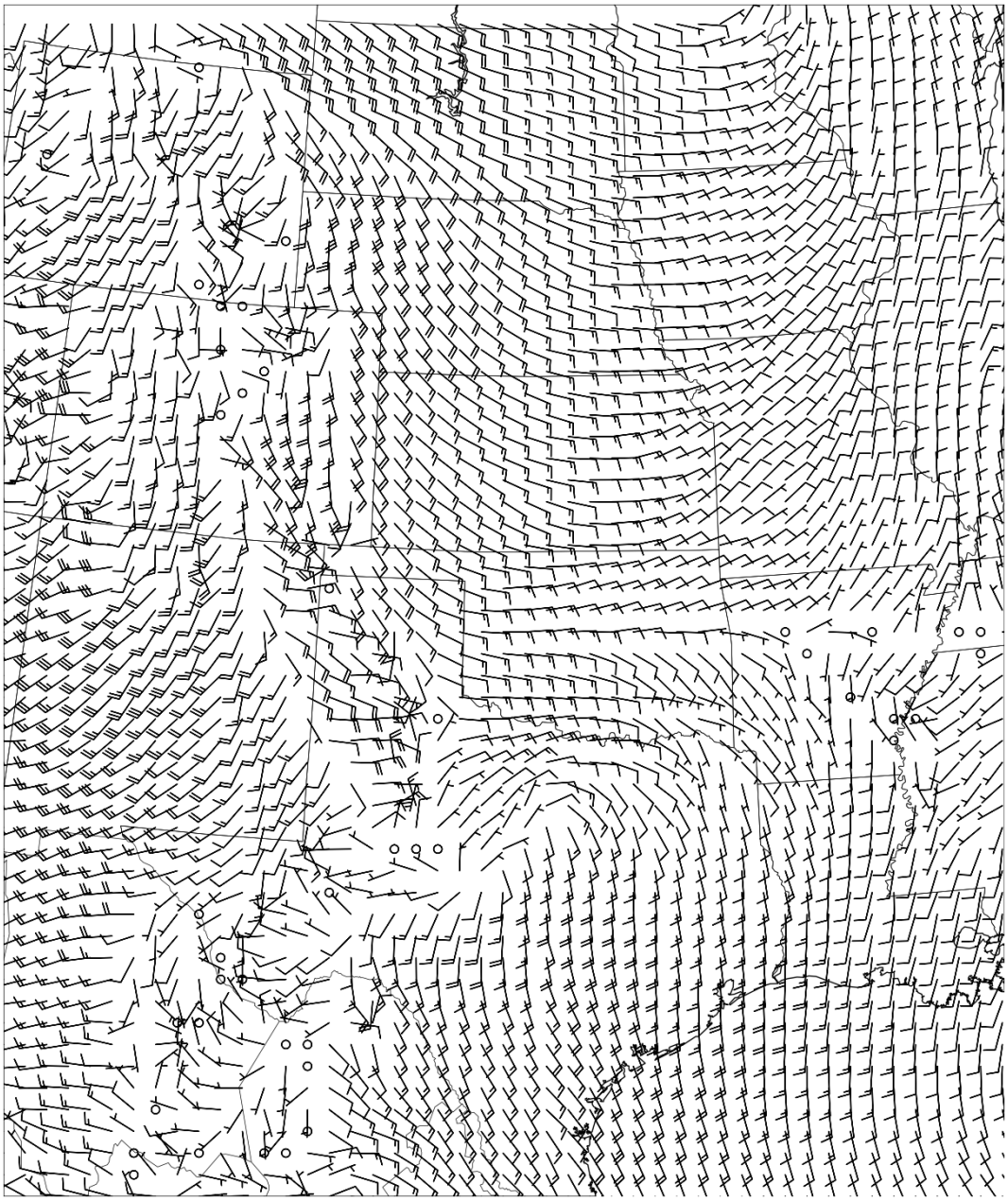


Figure 5.18: 24 hour forecast of 10 m wind speed and direction from the NSSL-WRF valid 0000 UTC 27 April 2012. Barb orientation indicates wind direction. Small flags indicate wind speeds of 5 kts, and larger flags 10 kts. Circles indicate calm winds.

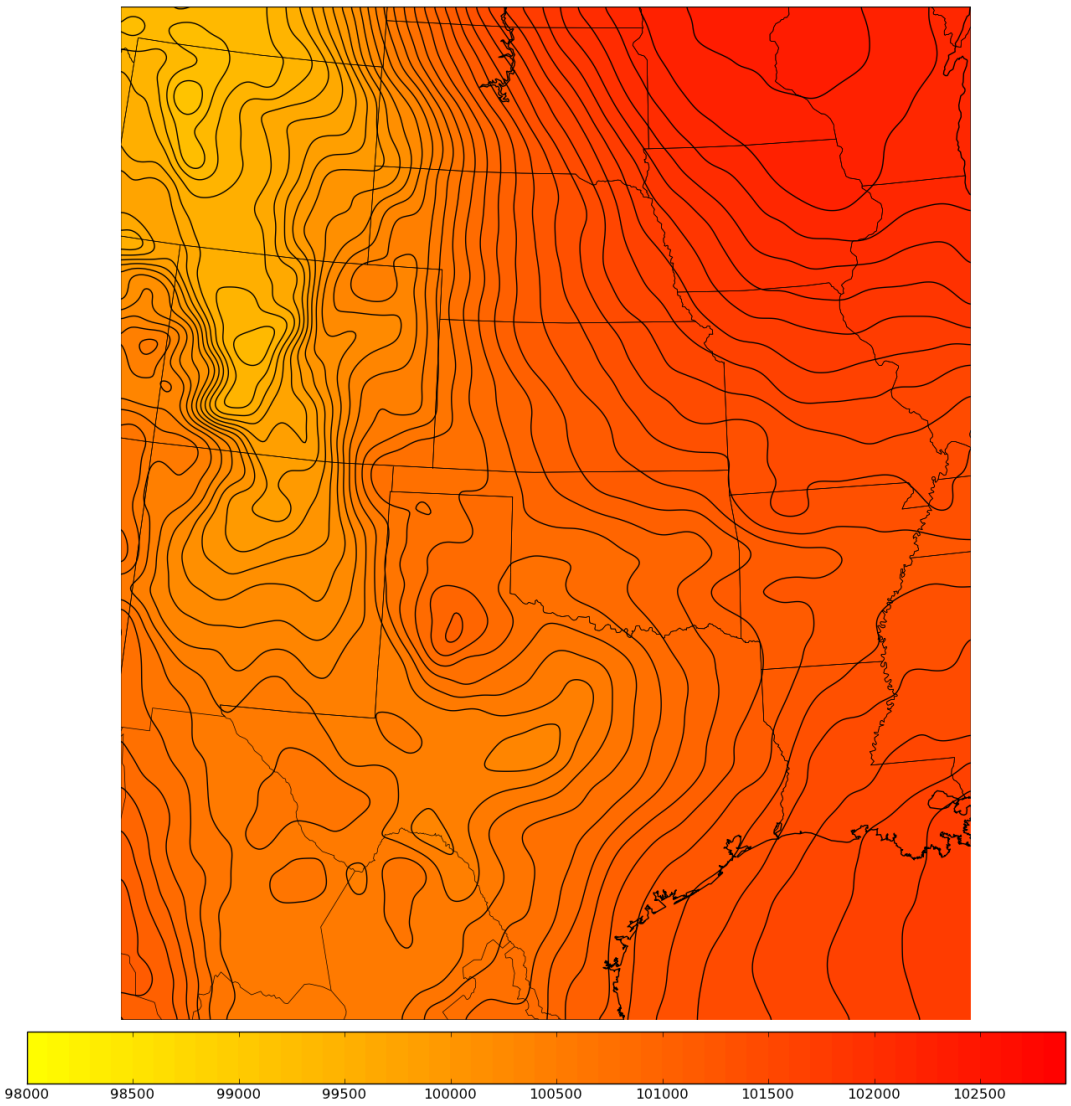


Figure 5.19: 24 hour forecast of surface pressure (Pa), contoured every 100 Pa (1 mb), from the NSSL-WRF valid 0000 UTC 27 April 2012.

or too few samples remain (in this case the probabilities of the possible outcomes are typically returned). Each decision tree in the RF is "perturbed" by denying it access to certain attributes during its creation.

The RF technique was chosen for three reasons. First, the structure of RFs is well suited for classification (although they can be applied to regression problems). Second, the importance of the attributes being provided to the RF are unknown. There has been no assessment of how useful each may be to classification; however, the "perturbations" of the forest's component decision trees, combined with the algorithm's ensemble approach to prediction, enables the technique to handle unimportant attributes without significant impact. Finally, RFs are human readable (or at least their decision trees are). This allows for easier understanding of the selection process formed by the learning algorithm and subsequent examination the underlying causes of its selections (i.e., why certain attribute/threshold combinations were/were not important).

5.2.1 Training and testing datasets

The performance of RFs is commonly assessed using cross-validation. The full dataset is broken into n subsets under the assumption that the contained samples are independent. The RF is then trained (i.e., created) on n-1 of the subsets and tested (i.e., its performance is assessed) on the remainder. However, this method cannot be used here as any two sampled points may be drawn from neighboring locations on the same dryline, which breaks the independence assumption. Therefore, a single training dataset and a single testing dataset are used, with data sampled from the 2007-2010 and 2011 outputs, respectively. Each dataset was formed by selecting a random point from a randomly selected

objectively analyzed dryline from the appropriate time frame. A total of 219 attributes were then assessed, including pressure, moisture, and thermal variables, variable variances along the boundary using a number of distance thresholds, and dry and moist side values (using the specific humidity gradient orientation as a proxy for dryline orientation). The process was repeated until 10000 TP and 10000 FP data points (per the evaluation performed above) were accumulated to ensure that each dataset was balanced (i.e., it has the same number of samples of each output label).

5.2.2 Results

An initial estimate of the RF's performance is assessed by performing an object-based statistical evaluation using the same settings described in section 5.1.2. Tables 5.2a and 5.2b show the 2011 object-based skill scores of the dryline algorithm before and after the application of the RF, respectively. The learning technique decreases the FAR by nearly 20% while not greatly impacting the POD. In addition, the frequency bias, CSI, and GSS have all improved. These show some promise for future examination of the RF application; however, these statistics are unable to demonstrate the full impact of the algorithm. In many situations, the RF removes the large majority, but not all, of a false feature (e.g., Fig. 5.20). When reevaluated using object-based techniques, the small regions are treated as individual objects and contribute a full FP to the contingency table (this is due to this section's preliminary nature and related lack of additional processing). Examination of the test dataset contingency table (Table 5.3b; sample by sample comparison - not object-based) shows that the RF is able to correctly remove ~68% of the non-dryline pixels while only incorrectly removing ~10% of the dryline pixels. This is particularly useful

as it may allow for easier removal of non-dryline features during subsequent processing despite their partial retention. The training set contingency table (Table 5.3a) has superior values; however, this is likely a sign of overfitting by the RF as opposed to a desirable trait.

Future considerations involve the incorporation of improved attributes, removal of noisy attributes, and reduction of overfitting (e.g., through increasing the minimum number of samples required to continue branching). Subsequent processing of the output (e.g., considering the percent of neighboring pixels that are retained, reapplication of the vapor pressure gradient magnitude threshold, application of a secondary machine learning technique) may also provide significant benefits.

POD	0.905
FAR	0.868
FBIAS	6.87
CSI	0.130
GSS	0.116

(a) Skill score values for pre-RF 2011 Coffer et al. (2013) dataset (see in-text description for a note on GSS calculation).

POD	0.890
FAR	0.713
FBIAS	3.06
CSI	0.276
GSS	0.228

(b) Skill score values for post-RF 2011 Coffer et al. (2013) dataset (see in-text description for a note on GSS calculation).

Table 5.2: 2011 Coffer et al. (2013) dataset skill scores before and after RF application.

	Dryline (RF)	Not Dryline (RF)
Dryline (truth)	9410	590
Not Dryline (truth)	1703	8297

(a) Training set (2007-2010) contingency table.

	Dryline (RF)	Not Dryline (RF)
Dryline (truth)	8941	1059
Not Dryline (truth)	3150	6850

(b) Test set (2011) contingency table

Table 5.3: RF training and test set contingency tables

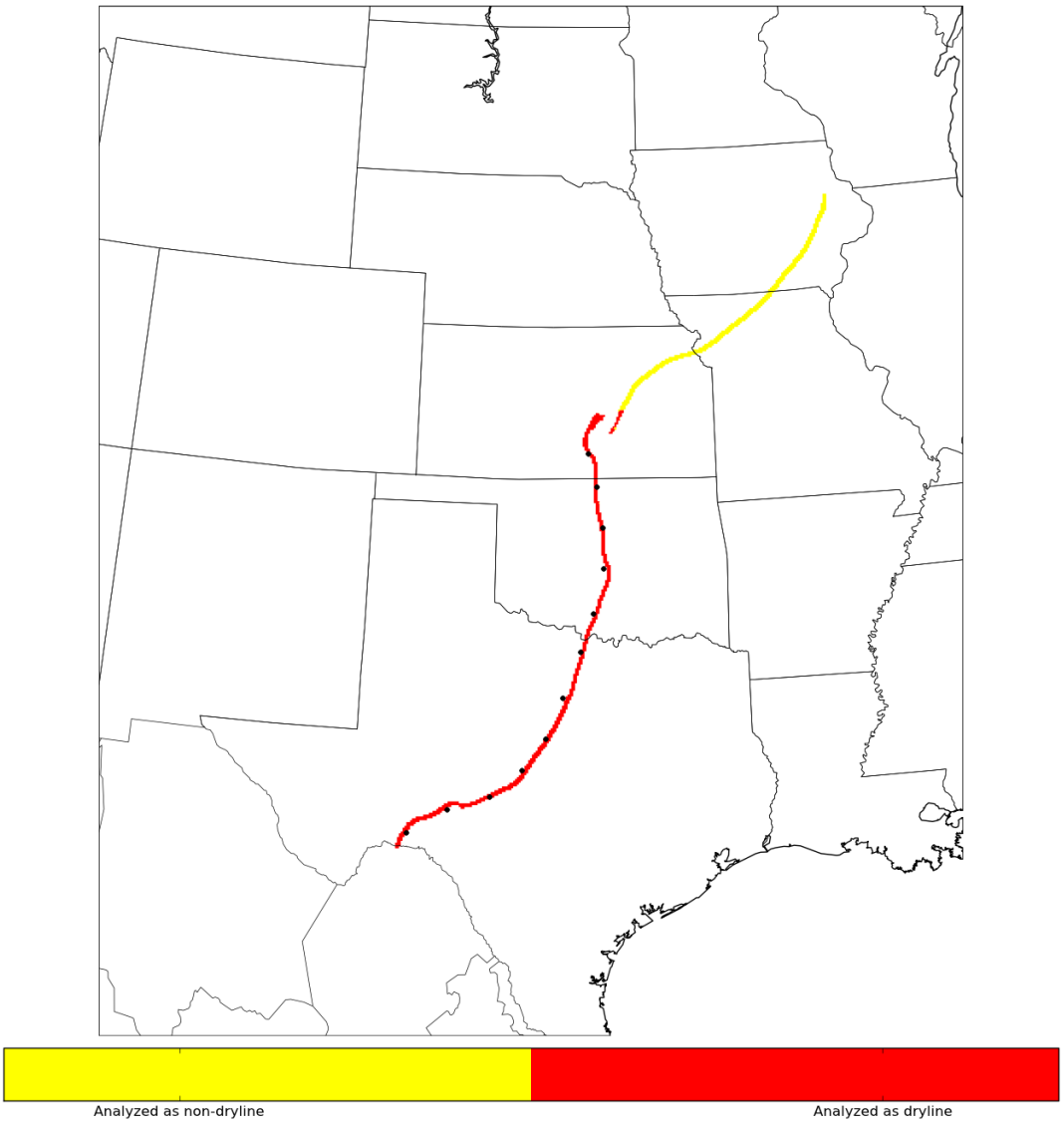


Figure 5.20: Output of random forest applied to the dryline algorithm assessment of 24 hour forecast data from the NSSL-WRF valid 0000 UTC 04 April 2011. Yellow areas are regions that the random forest has removed from consideration, while red indicates that the area has been retained.

Chapter 6

Conclusions and Future Expansion

This thesis describes an automated, multi-parameter dryline identification algorithm designed to aid in the analysis of the large quantities of gridded numerical model data available to the meteorological community. The technique utilizes a number of image processing and morphological transformations to identify drylines present in a single timestep of NWP output. Initial evaluation of the algorithm was performed via comparison with subjectively analyzed drylines from Coffer et al. (2013) and showed consistent agreement between objective and subjective analyses (in the case of true drylines) with little bias. The generally small differences of the two analyses are likely due to a combination of small scale deviations being lost through increased smoothing and the algorithm's strict adherence to following only the strongest moisture gradients. In addition, it was found that the algorithm tends to extend drylines beyond the regions identified by the subjective analysis. This is likely caused by the optimistic threshold not properly bounding the lines found by NMS.

Statistical evaluation of the method showed the algorithm to be overly optimistic. The high POD indicated that the majority of drylines were found by the method; however, its other scores (particularly FAR) suggest that a large number of non-dryline features are being retained. Several techniques have been attempted to reduce these occurrences; however, the inconsistency of values along the dryline makes blanket removal extremely difficult. Fortunately,

the object-oriented structure of the algorithm is ideally suited for application of learning techniques to specify and remove undesirable regions. Preliminary use of a random forest was shown to provide a significant decrease in the number of FPs while minimally impacting TPs. In addition, the capability to identify the cause of individual false alarms (e.g., frontal boundary vs. gust front) allows for the development of a dataset that can be used in other identification algorithms (e.g., cold front identification).

The algorithm devised in this thesis lays the groundwork for a final product with the potential to provide significant contributions to a variety of meteorological applications ranging from model evaluation to operational forecasting. The main issue facing the method is the inclusion of non-dryline features. In addition to expansions of the machine learning application, several other avenues may aid the future development of the algorithm. First, the current implementation only examines the surface characteristics of the dryline. However, these boundaries typically have a well defined vertical structure that can be exploited. For example, the moisture boundary at the surface extends well above the ground and then further into the horizontal. Provided sufficient resolution, the gradient region should be identifiable and provide another structure for examination. Similarly, the algorithm could be modified to search for the large air masses that form the dryline boundary. This would allow for examination of the data at a much larger scale and may result in easier subsequent identification. For example, one air mass forming a suspected dryline could be required to interact with the Gulf of Mexico while maintaining a vertical moisture gradient that connects with a corresponding surface boundary. Another avenue is the development and use of techniques designed to identify other meteorological features. The output of these methods could be combined with current data

to form the basis for a more robust algorithm that can make selections based on input from multiple techniques (similar to potential implementations of the learning methods described above, but incorporating all non-dryline features, not just those incorrectly identified as drylines). In particular, a method to identify surface low pressure systems would help in preventing the overextension of dryline boundaries as described in section 5.1.1.

Reference List

Atkins, N. T., R. M. Wakimoto, and C. L. Ziegler, 1998: Observations of the finescale structure of a dryline during VORTEX 95. *Monthly Weather Review*, **126**, 525–550.

Bluestein, H. B., 1993: *Synoptic-Dynamic Meteorology in Midlatitudes: Volume II: Observations and Theory of Weather Systems*. Oxford University Press.

Bluestein, H. B. and S. S. Parker, 1993: Modes of isolated, severe convective storm formation along the dryline. *Monthly Weather Review*, **121**, 1354–1372.

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

Buban, M. S., L. Z. Conrad, 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. *Monthly Weather Review*, **135**, 2473–2505.

Buban, M. S., C. L. Ziegler, E. R. Mansell, and Y. P. Richardson, 2012: Simulation of dryline misovortex dynamics and cumulus formation. *Monthly Weather Review*, **140**, 3525–3551.

Chartrand, R., 2011: Numerical differentiation of noisy, nonsmooth data. *ISRN Applied Mathematics*, **2011**.

Chen, F. and J. Dudhia, 2001: Coupling an advanced land surface–hydrology model with the Penn State–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Monthly Weather Review*, **129**, 569–585.

Coffer, B., L. Maudlin, P. Veals, and A. Clark, 2013: Dryline position errors in experimental convection-allowing NSSL-WRF model forecasts and the operational NAM. *Weather and Forecasting*, **28**, 746–761.

Crawford, T. and H. Bluestein, 1997: Characteristics of dryline passage during COPS-91. *Monthly Weather Review*, **125**, 463–477.

Davis, C., B. Brown, and R. Bullock, 2006: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Monthly Weather Review*, **134**, 1772–1784.

Dixon, M. and G. Wiener, 1993: TITAN: Thunderstorm Identification, Tracking, Analysis, and Nowcasting - A radar-based methodology. *Journal of Atmospheric and Oceanic Technology*, **10**, 785–797.

Dodd, A. V., 1965: Dew point distribution in the contiguous United States. *Monthly Weather Review*, **93**, 113–122.

Dudhia, J., 1989: Numerical study of convection observed during the Winter Monsoon Experiment using a mesoscale two-dimensional model. *Journal of Atmospheric Science*, **46**, 3077–3107.

Ebert, E. E. and J. L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *Journal of Hydrology*, **239**, 179–202.

Fujita, T., 1958: Structure and movement of a dry front. *Bulletin of the American Meteorological Society*, **39**, 574–582.

Geusebroek, J., A. Smeulders, and J. van de Weijer, 2002: Fast anisotropic Gauss filtering. *Proc. ECCV*, 99–112.

Goudenhoofdt, E. and L. Delobbe, 2012: Statistical characteristics of convective storms in Belgium derived from volumetric weather radar observations. *Journal of Applied Meteorology and Climatology*, **52**, 918.

Gunn, S. R., 1998: Edge detection error in the discrete Laplacian of Gaussian. *Proc. ICIP*, **2**, 515–519.

Hall, M., E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, 2009: The WEKA data mining software: An update. *SIGKDD Explorations*, **11**.

Hane, C. E., R. M. Rabin, T. M. Crawford, H. B. Bluestein, and M. E. Baldwin, 2002: A case study of severe storm development along a dryline within a synoptically active environment. Part II: Multiple boundaries and convective initiation. *Monthly Weather Review*, **130**, 900–920.

Hane, C. E., C. L. Ziegler, and H. B. Bluestein, 1993: Investigation of the dryline and convective storms initiated along the dryline: Field experiments during COPS-91. *Bulletin of the American Meteorological Society*, **74**, 2133–2145.

Hoch, J. and P. Markowski, 2005: A climatology of springtime dryline position in the U.S. Great Plains region. *J. Climate*, **18**, 2132–2137.

Hogan, R. J., C. A. T. Ferro, I. T. Jolliffe, and D. B. Stephenson, 2010: Equitability revisited: Why the "Equitable Threat Score" is not equitable. *Weather and Forecasting*, **25**, 710–726.

Hong, S.-Y. and J.-O. J. Lim, 2006: The WRF Single-Moment 6-Class Microphysics scheme (WSM6). *Journal of the Korean Meteorological Society*, **42**, 129–151.

Johnson, J., P. L. MacKeen, A. Witt, E. D. Mitchell, G. J. Stumpf, M. D. Eilts, and K. W. Thomas, 1998: The Storm Cell Identification and Tracking algorithm: An enhanced WSR-88D algorithm. *Weather and Forecasting*, **13**, 263–276.

Jones, P. A. and P. R. Bannon, 2002: A mixed-layer model of the diurnal dryline. *Journal of Atmospheric Science*, **59**, 2582–2593.

Koch, S. E., 1979: Mesoscale gravity waves as a possible trigger of severe convection along a dryline. PhD dissertation, University of Oklahoma.

Lakshmanan, V., 2004: A separable filter for directional smoothing. *Geoscience and Remote Sensing Letters, IEEE*, **1**, 192–195.

Lakshmanan, V., 2012: *Automating the Analysis of Spatial Grids*. Springer.

Lakshmanan, V., K. Hondl, and R. Rabin, 2009: An efficient, general-purpose technique for identifying storm cells in geospatial images. *Journal of Atmospheric and Oceanic Technology*, **26**, 523–537.

Lakshmanan, V., R. Rabin, and V. DeBrunner, 2003: Multiscale storm identification and forecast. *Journal of Atmospheric Research*, **67**, 367–380.

Lakshmanan, V. and T. Smith, 2009: Data mining storm attributes from spatial grids. *Journal of Atmospheric and Oceanic Technology*, **26**, 2353–2365.

Lakshmanan, V. and T. Smith, 2010: An objective method of evaluating and devising storm tracking algorithms. *Weather and Forecasting*, **29**, 721–729.

Lanicci, J. M., T. N. Carlson, and T. T. Warner, 1987: Sensitivity of the Great Plains severe-storm environment to soil-moisture distribution. *Monthly Weather Review*, **115**, 2660–2673.

Markowski, P. and Y. Richardson, 2010: *Mesoscale Meteorology in Midlatitudes*. Wiley-Blackwell.

Marr, D. and E. Hildreth, 1980: Theory of edge detection. *Proc. Royal Society of London. Series B, Biological Sciences*, **207**, 187–217.

Martin, W. J. and A. Shapiro, 2007: Discrimination of bird and insect radar echoes in clear air using high-resolution radars. *Journal of Atmospheric and Oceanic Technology*, **24**, 1215–1230.

Matteson, G. T., 1969: The west Texas dry front of June 1967. M.S. thesis, University of Oklahoma.

McCarthy, J. and S. E. Koch, 1982: The evolution of an Oklahoma dryline. Part I: A meso- and subsynoptic-scale analysis. *Journal of Atmospheric Science*, **39**, 225–236.

McGuire, E. L., 1962: The vertical structure of three dry lines as revealed by aircraft traverses. Tech. Rep. R7, National Severe Storms Project.

Mellor, G. L. and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Reviews of Geophysics*, **20**, 851–875.

Miao, Q. and B. Geerts, 2007: Finescale vertical structure and dynamics of some dryline boundaries observed in IHOP. *Monthly Weather Review*, **135**, 4161–4184.

Michalakes, J. and M. Vachharajani, 2008: GPU acceleration of numerical weather prediction. *Parallel Processing Letters*, **18**, 531–548.

Mitchell, E. D., S. V. Vasiloff, G. J. Stumpf, A. Witt, M. D. Eilts, J. T. Johnson, and K. W. Thomas, 1998: The National Severe Storms Laboratory Tornado Detection Algorithm. *Weather and Forecasting*, **13**, 352–366.

Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. *Journal of Geophysical Research: Atmospheres*, **102**, 2156–2202.

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. *Monthly Weather Review*, **134**, 406–430.

Nixon, M. S. and A. S. Aguado, 2012: *Feature Extraction & Image Processing for Computer Vision*. 3d ed., Academic Press.

NOAA, 1998: Automated Surface Observing System (ASOS) User's Guide. Manual.

NWS, 1999: Oklahoma/southern Kansas tornado outbreak of May 3, 1999. NOAA Service Assessment.

Ogura, Y., H. Juang, K. Zhang, and S. Soong, 1982: Possible triggering mechanisms for severe storms in SESAME-AVE IV (9–10 May 1979). *Bulletin of the American Meteorological Society*, **63**, 503–515.

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

Peterson, R. E., 1983: The west Texas dryline: Occurrence and behavior. *Preprints, 13th Conf. on Severe Local Storms, Tulsa, OK, American Meteorological Society, J9 - J11.*

Rhea, J. O., 1966: A study of thunderstorm formation along dry lines. *Journal of Applied Meteorology*, **5**, 58–63.

Roebber, P. J., D. M. Schultz, and R. Romero, 2002: Synoptic regulation of the 3 May 1999 tornado outbreak. *Boundary-Layer Meteorology*, **17**, 399–429.

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. *Boundary-Layer Meteorology*, **82**, 235–262.

Sanders, F., 1999: A proposed method of surface map analysis. *Monthly Weather Review*, **127**, 945–955.

Sanders, F. and C. A. Doswell, 1995: A case for detailed surface analysis. *Bulletin of the American Meteorological Society*, **76**, 505–521.

Schaefer, J. T., 1973: The motion and morphology of the dryline. Tech. Rep. COM-74-10043, NOAA Technical Memo.

Schaefer, J. T., 1974: The life cycle of the dryline. *Journal of Applied Meteorology*, **13**, 444–449.

Schaefer, J. T., 1986: The dryline. *Mesoscale Meteorology and Forecasting*, P. S. Ray, Ed., American Meteorological Society.

Schaefer, J. T., 1990: The Critical Success Index as an indicator of warning skill. *Weather and Forecasting*, **5**, 570–575.

Schultz, D. M., C. C. Weiss, and P. M. Hoffman, 2007: The synoptic regulation of dryline intensity. *Monthly Weather Review*, **135**, 1699–1709.

Shaw, B. L., R. A. Pielke, and C. L. Ziegler, 1997: A three-dimensional numerical simulation of a Great Plains dryline. *Monthly Weather Review*, **125**, 1489–1506.

Sipprell, B. D. and B. Geerts, 2007: Finescale vertical structure and evolution of a preconvective dryline on 19 June 2002. *Monthly Weather Review*, **135**, 2111–2134.

Skamarock, W. C., et al., 2008: A description of the advanced research WRF version 3. Tech. Rep. TN-475+STR, National Center for Atmospheric Research.

Sun, C. and P. Vallotton, 2009: Fast linear feature detection using multiple directional non-maximum suppression. *Journal of Microscopy*, **234**, 147–157.

Thompson, R. L. and R. Edwards, 2000: An overview of environmental conditions and forecast implications of the 3 May 1999 tornado outbreak. *Weather and Forecasting*, **15**, 682–699.

Uccellini, L. W., S. F. Corfidi, N. W. Junker, P. J. Kocin, and D. A. Olson, 1992: Report on the surface analysis workshop held at the National Meteorological Center. *Bulletin of the American Meteorological Society*, **73**, 459–472.

Wakimoto, R. M. and H. V. Murphey, 2010: Frontal and radar refractivity analyses of the dryline on 11 June 2002 during IHOP. *Monthly Weather Review*, **138**, 228–241.

Wakimoto, R. M., H. V. Murphey, E. V. Browell, and S. Ismail, 2006: The "Triple Point" on 24 May 2002 during IHOP. Part I: Airborne doppler and LASE analyses of the frontal boundaries and convection initiation. *Monthly Weather Review*, **134**, 231–250.

Weeks, A., 1996: *Fundamentals of Electronic Image Processing*. Wiley-IEEE Press.

Wilks, D. S., 1995: *Statistical methods in the Atmospheric Sciences*. Academic Press.

Williams, R. J., 1976: Surface parameters associated with tornadoes. *Monthly Weather Review*, **104**, 540–545.

Witt, A., M. D. Eilts, G. J. Stumpf, J. T. Johnson, E. D. Mitchell, and K. W. Thomas, 1998: An enhanced hail detection algorithm for the WSR-88D. *Weather and Forecasting*, **13**, 286–303.

Xue, M. and W. J. Martin, 2006: A high-resolution modeling study of the 24 May 2002 dryline case during IHOP. Part II: Horizontal convective rolls and convective initiation. *Monthly Weather Review*, **134**, 172–191.

Zhang, T. Y. and C. Y. Suen, 1984: A fast parallel algorithm for thinning digital patterns. *Communications of the ACM*, **27**, 236–239.

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

Ziegler, C. L., T. J. Lee, and R. A. Pielke, 1997: Convective initiation at the dryline: A modeling study. *Monthly Weather Review*, **125**, 1001–1026.

Ziegler, C. L. and E. N. Rasmussen, 1998: The initiation of moist convection at the dryline: Forecasting issues from a case study perspective. *Weather and Forecasting*, **13**, 1106–1131.

Appendices

Appendix A

Development Process and Notable Formulations

This appendix discusses the initial examination of the data as well as a number of intermediate algorithms leading to the development of the final technique. Unless otherwise specified, an isotropic Gaussian filter with a standard deviation of 2.5 grid points (approximately 10 km at 4 km resolution) has been applied to each field prior to use. The process helps remove smaller features while highlighting dryline sized regions.

A.1 Initial Examination

Initial examination of the data was performed using a combination of smoothing parameters and edge/gradient detection techniques. The effects of each are shown using the 22 April 2008 case due to the presence of a variety of meteorological boundaries, including a strong dryline from western Texas into southwest Oklahoma. The dryline impinges upon a stationary front exhibiting a strong moisture gradient along part of its length, with a cold front also present further northeast (Fig. A.1). Figs. A.2 - A.13 show the effect of various standard deviations on the Sobel and LoG (see appendix B.2) specific humidity calculations. As expected, fewer spurious features are visible in the outputs using higher σ

values; however, finer variations in the true boundary location are lost (this is of greater importance in less well defined drylines). One benefit of the Sobel operator is the retention of a first order gradient magnitude value that can be easily related to the strength of the dryline, which the LoG does not have. LoG can be implemented instead by examining the difference of the output provided by the application of maximum and minimum filters to the second order derivative field. This provides a sense of magnitude along the zero crossing region (see appendix B.2); however, it also introduces lobe-like structures that outline the desired regions despite also acting as a filter for smaller features (Fig. A.12).

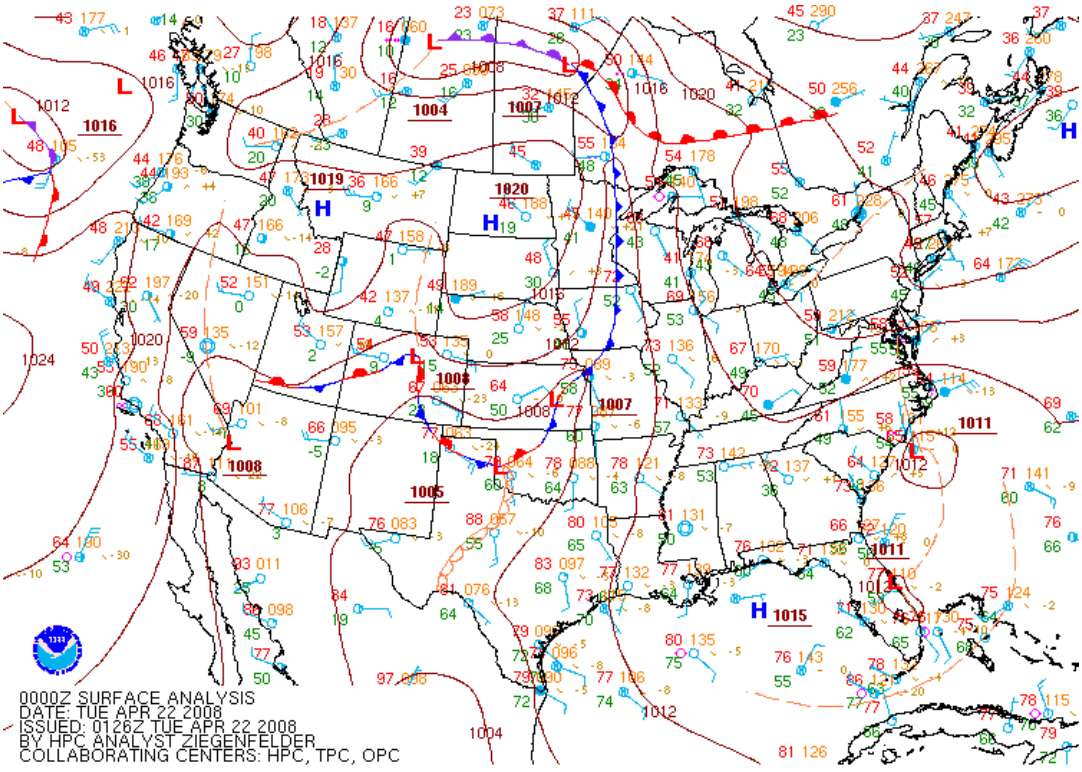


Figure A.1: Valid 22 April 2008, 0000 UTC: Hydrometeorological Prediction Center (HPC) surface analysis corresponding to the analysis date. The light brown, bumped line denotes the dryline position. Cold fronts are represented by the blue line with triangles. Warm fronts are denoted with a red, bumped line. The line that alternates blue triangles and red humps indicates the presence of a stationary front.

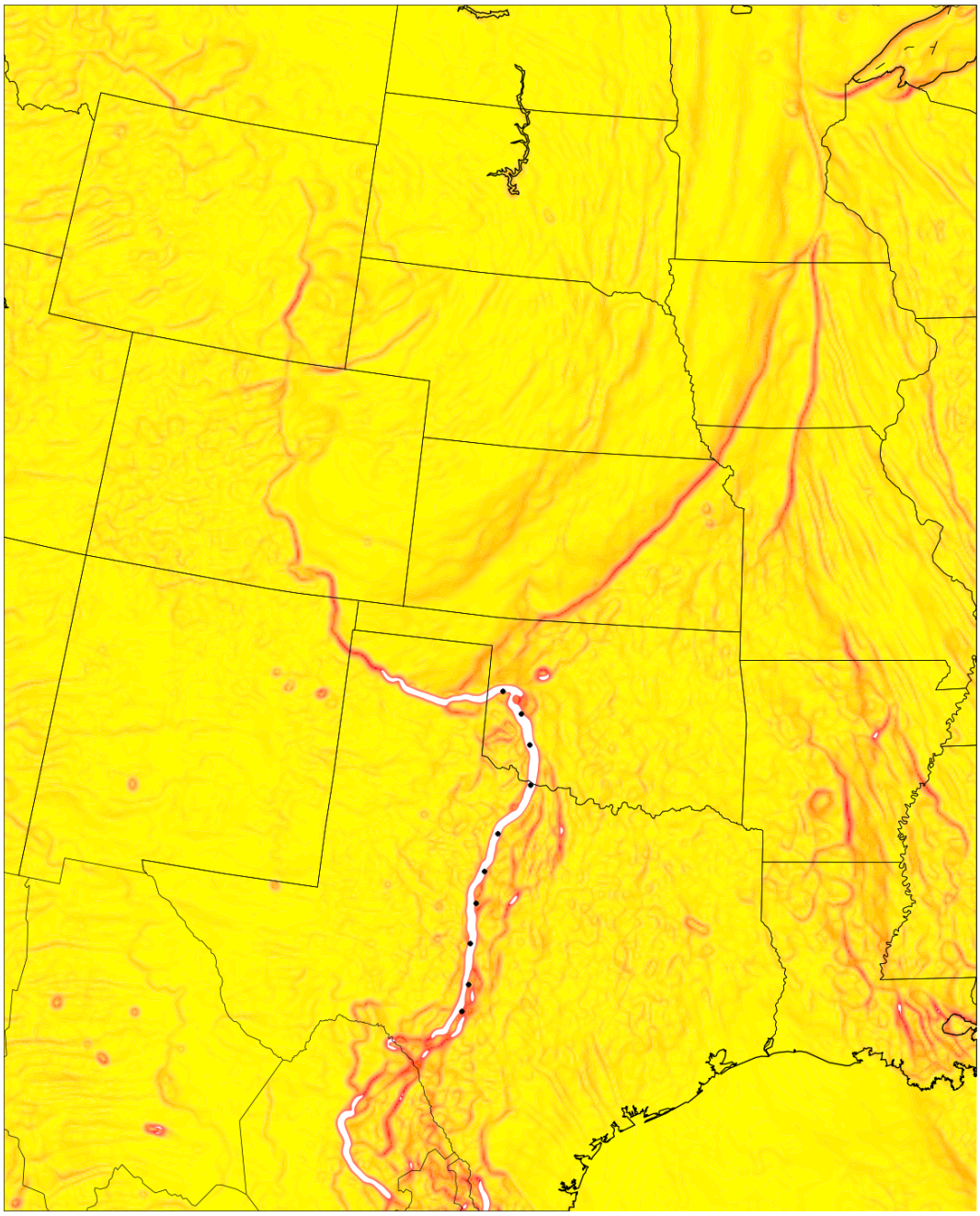


Figure A.2: Sobel of specific humidity, sigma = 1.0, calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 22 April 2008.

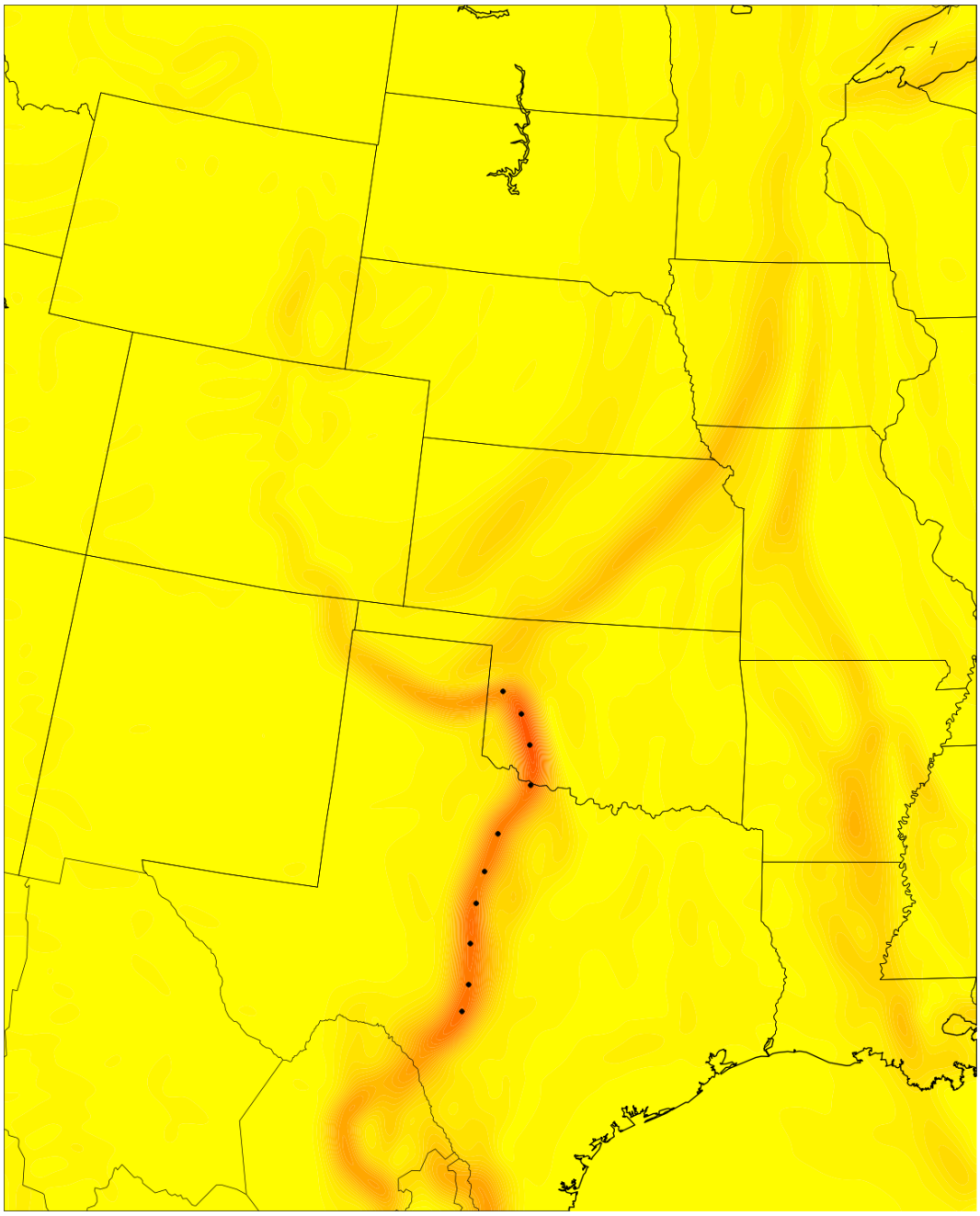


Figure A.3: Sobel of specific humidity, sigma = 2.5, calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 22 April 2008.

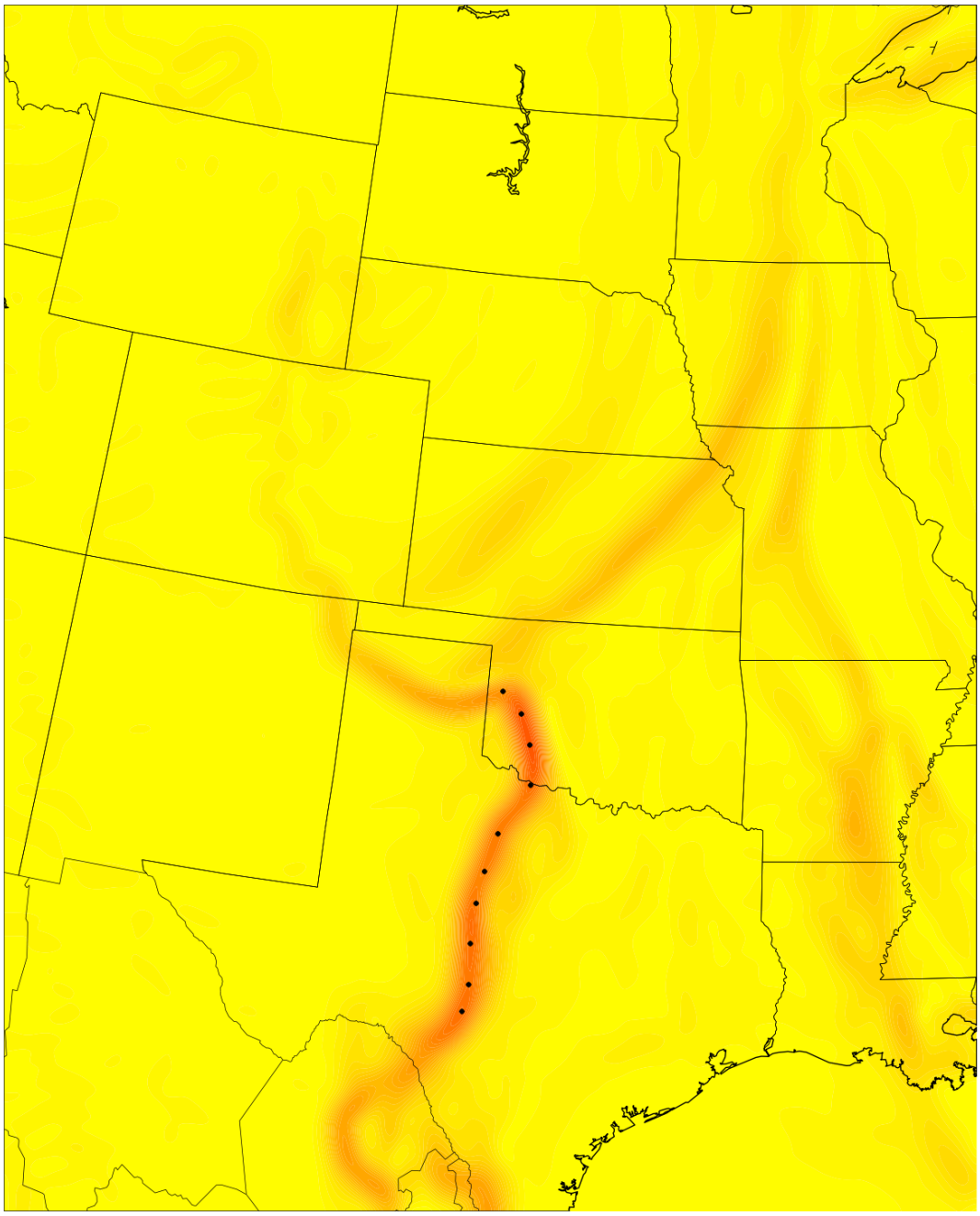


Figure A.4: Sobel of specific humidity, sigma = 6.0, calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 22 April 2008.

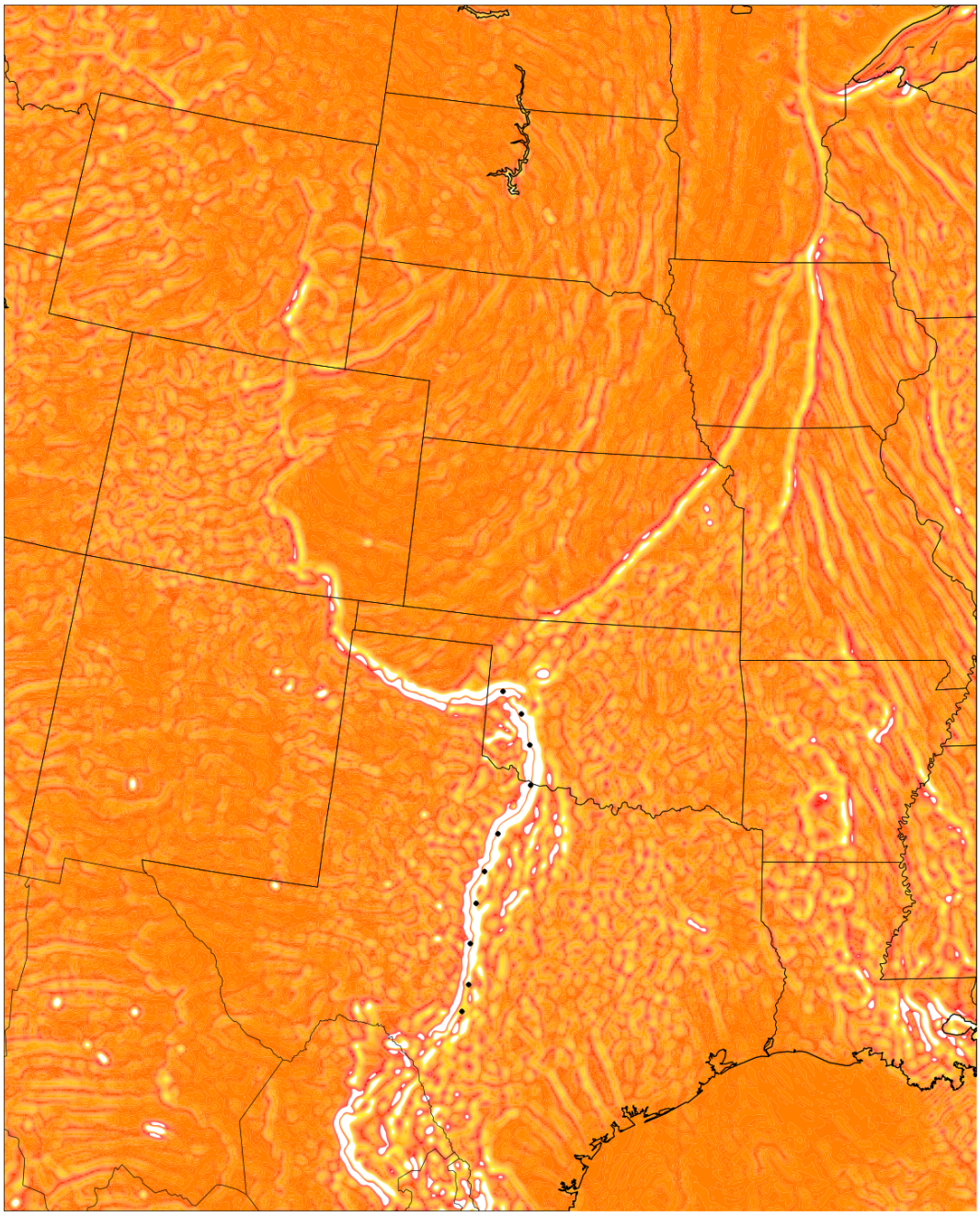


Figure A.5: Raw LoG of specific humidity, sigma = 1.0, calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 22 April 2008.

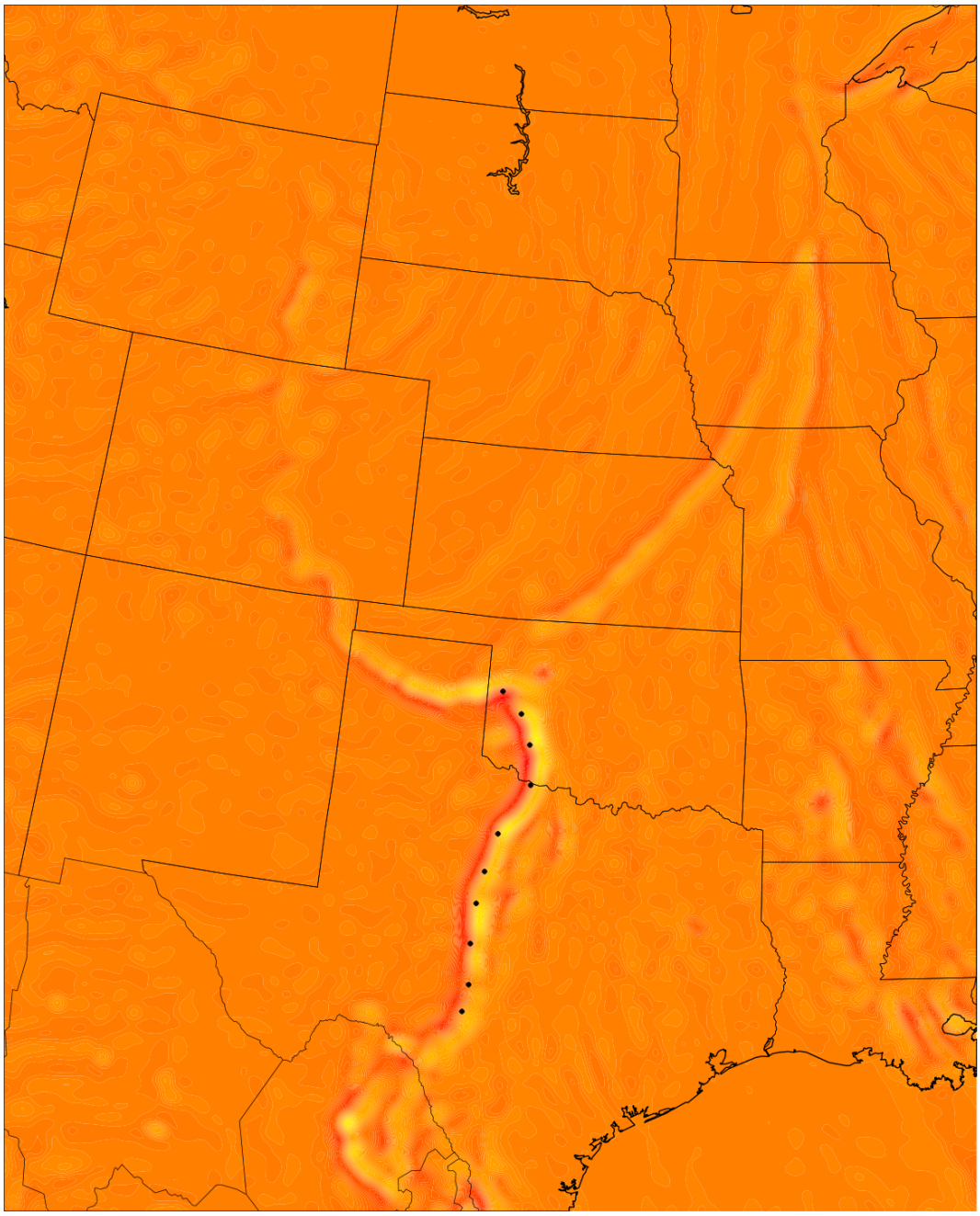


Figure A.6: Raw LoG of specific humidity, sigma = 2.5, calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 22 April 2008.

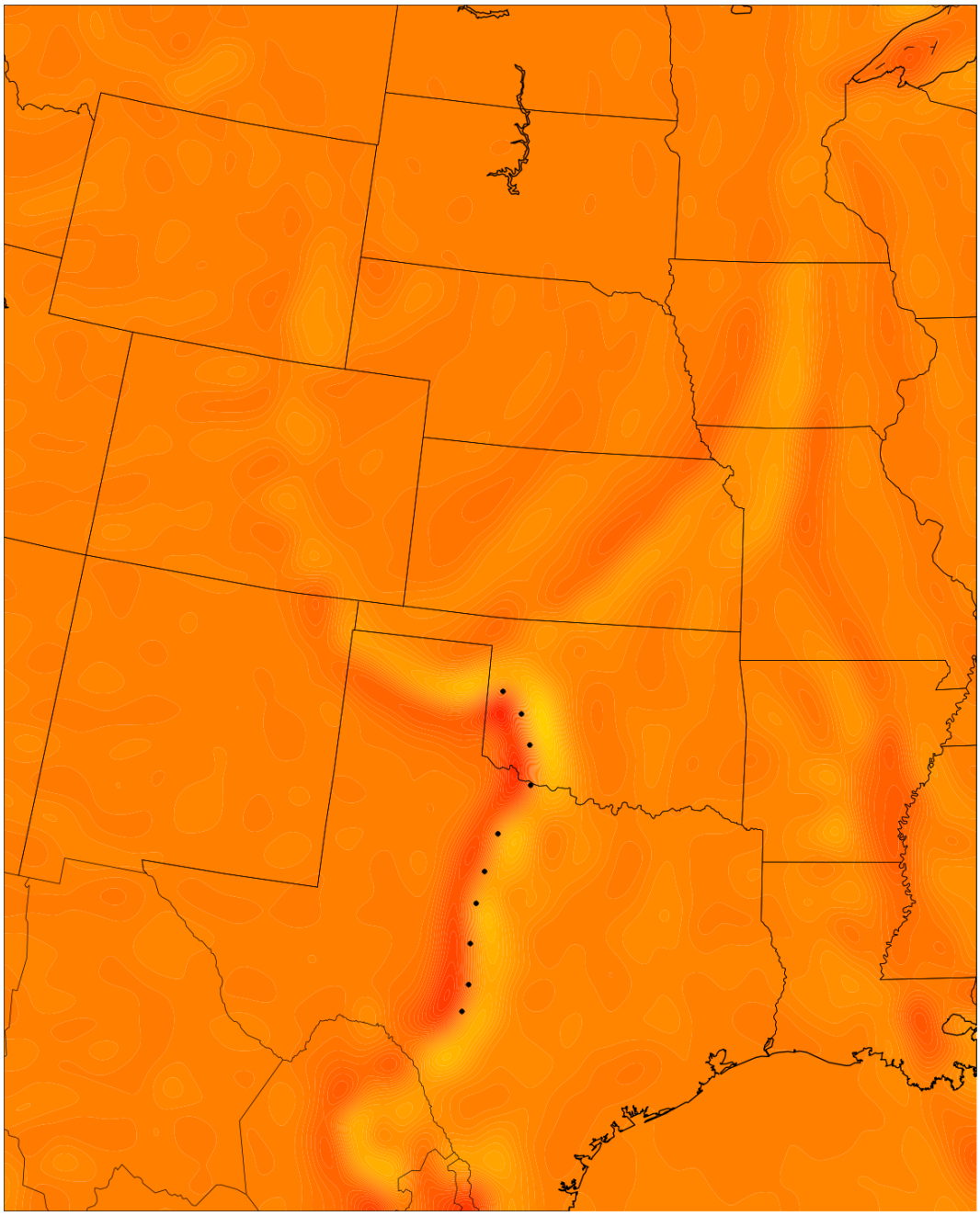


Figure A.7: Raw LoG of specific humidity, sigma = 6.0, calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 22 April 2008.

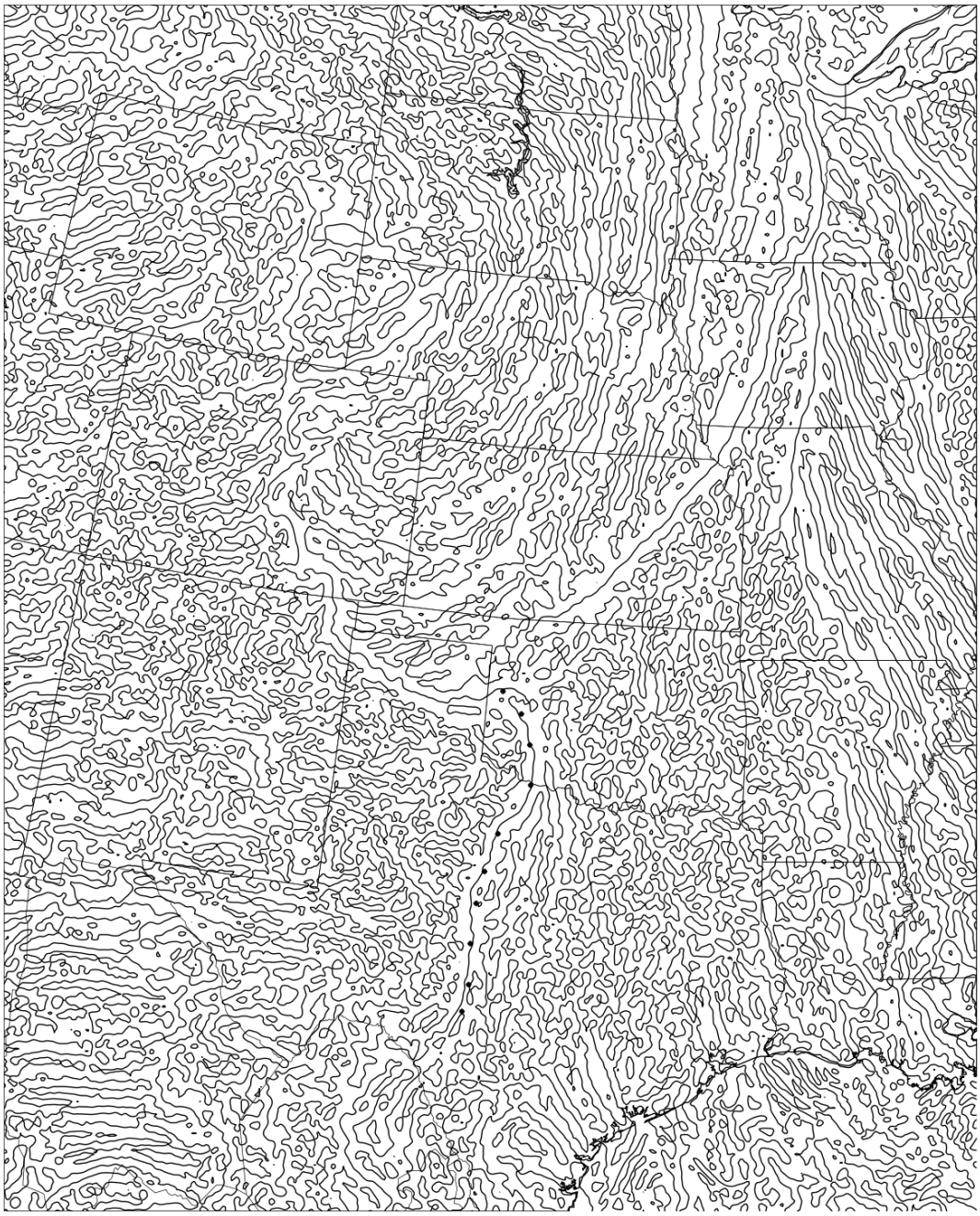


Figure A.8: LoG of specific humidity zero crossings, sigma = 1.0, calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 22 April 2008.

Figure A.9: LoG of specific humidity zero crossings, sigma = 2.5, calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 22 April 2008.

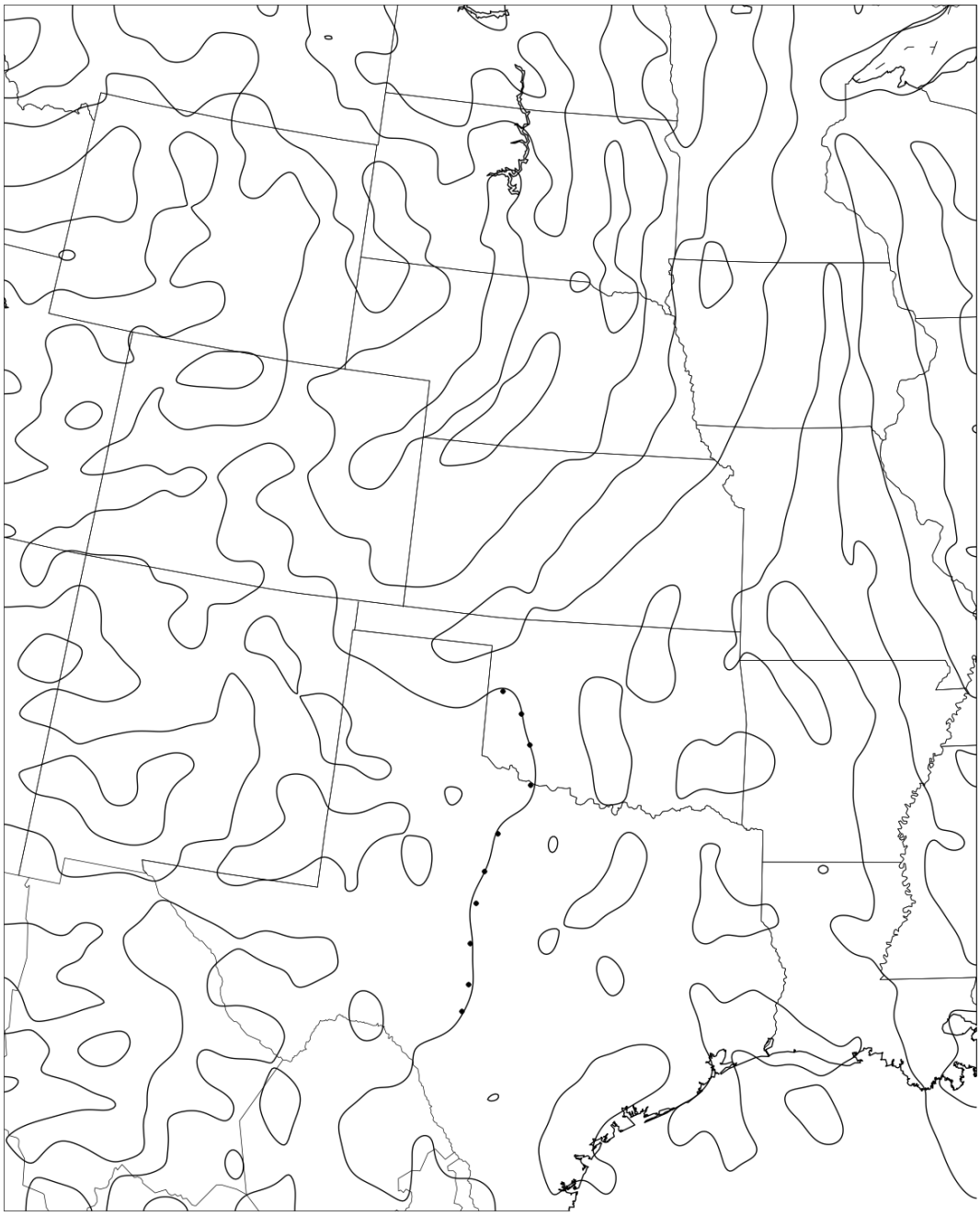


Figure A.10: LoG of specific humidity zero crossings, sigma = 6.0, calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 22 April 2008.

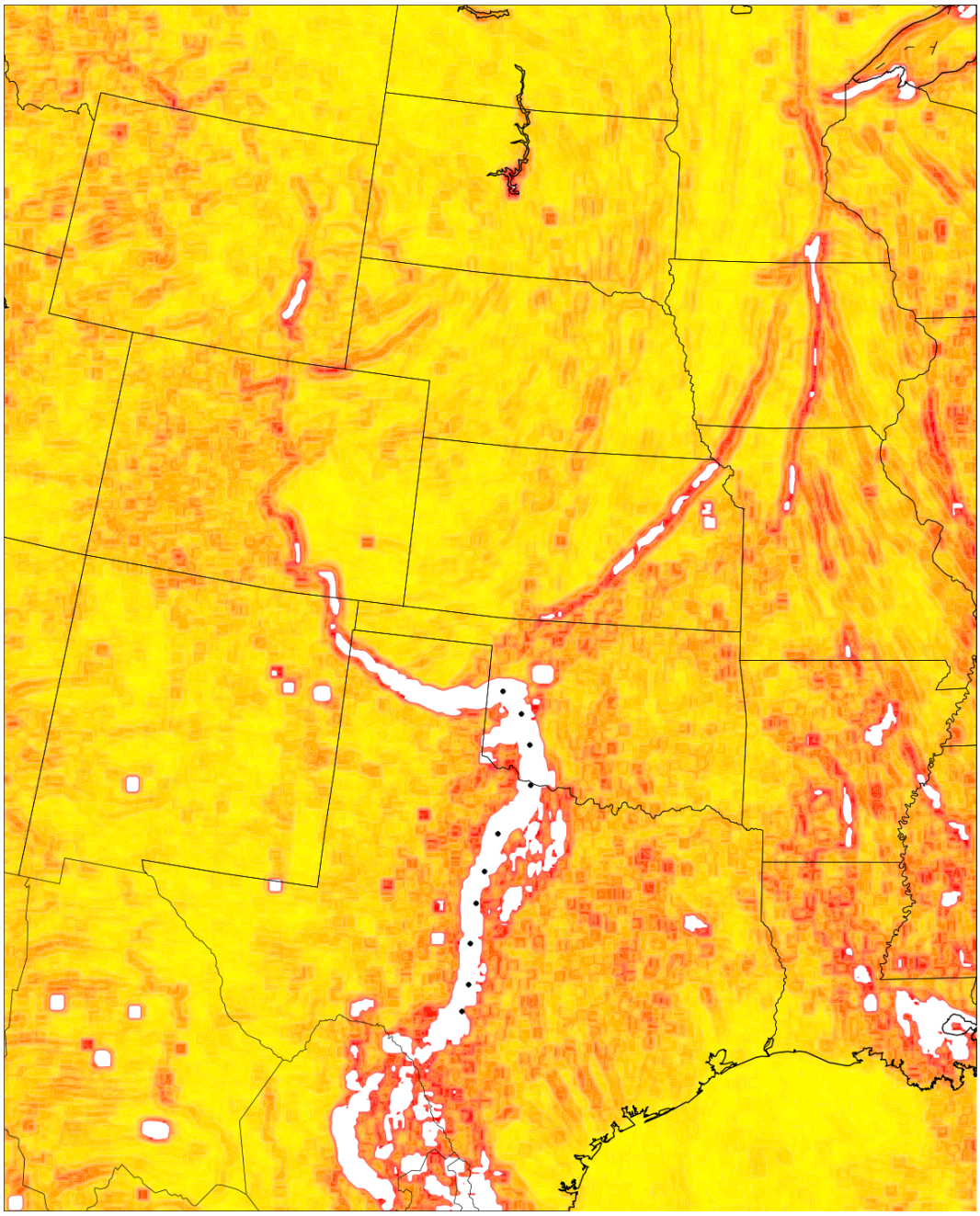


Figure A.11: LoG of specific humidity max - min, sigma = 1.0, calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 22 April 2008.

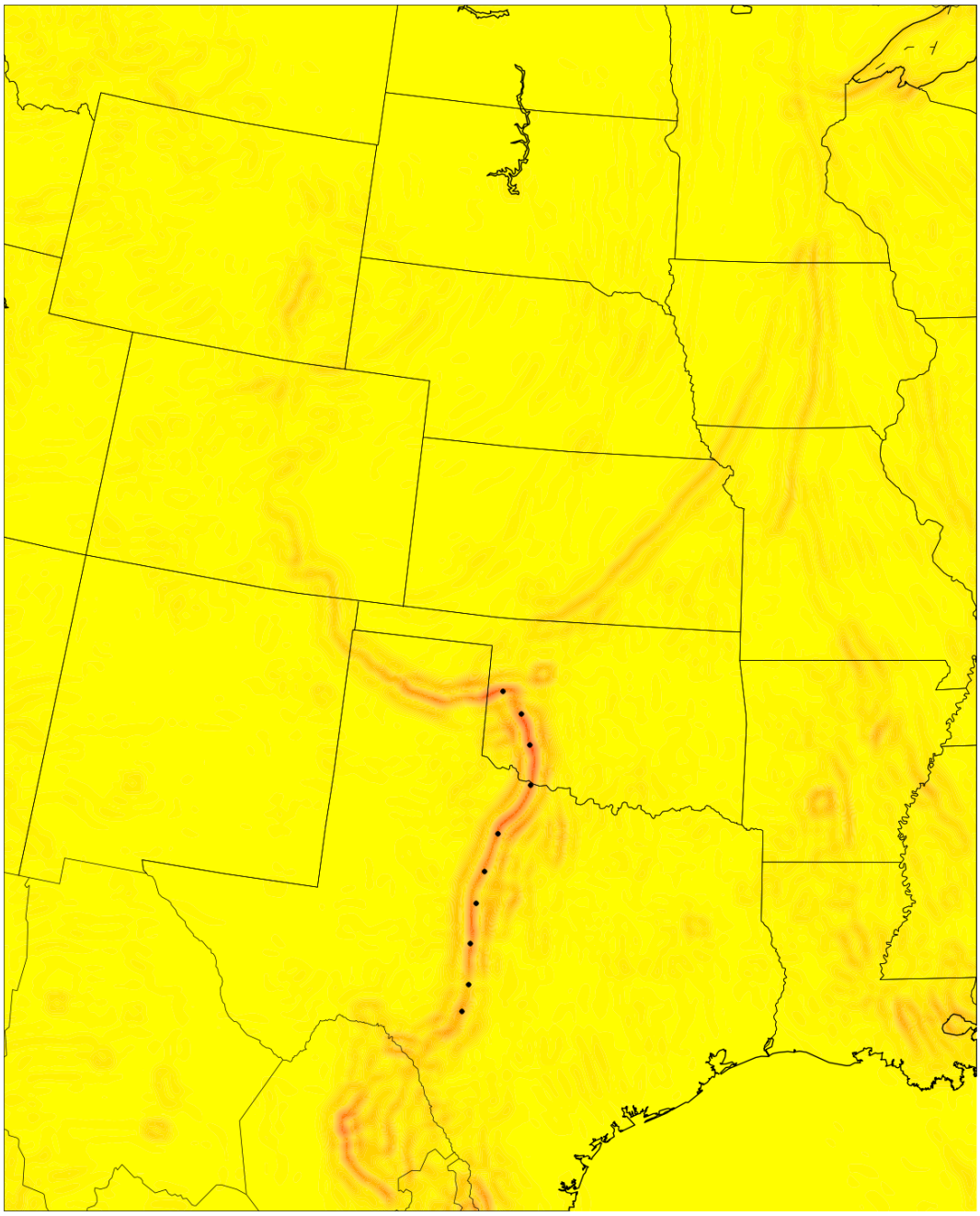


Figure A.12: LoG of specific humidity max - min, sigma = 2.5, calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 22 April 2008.

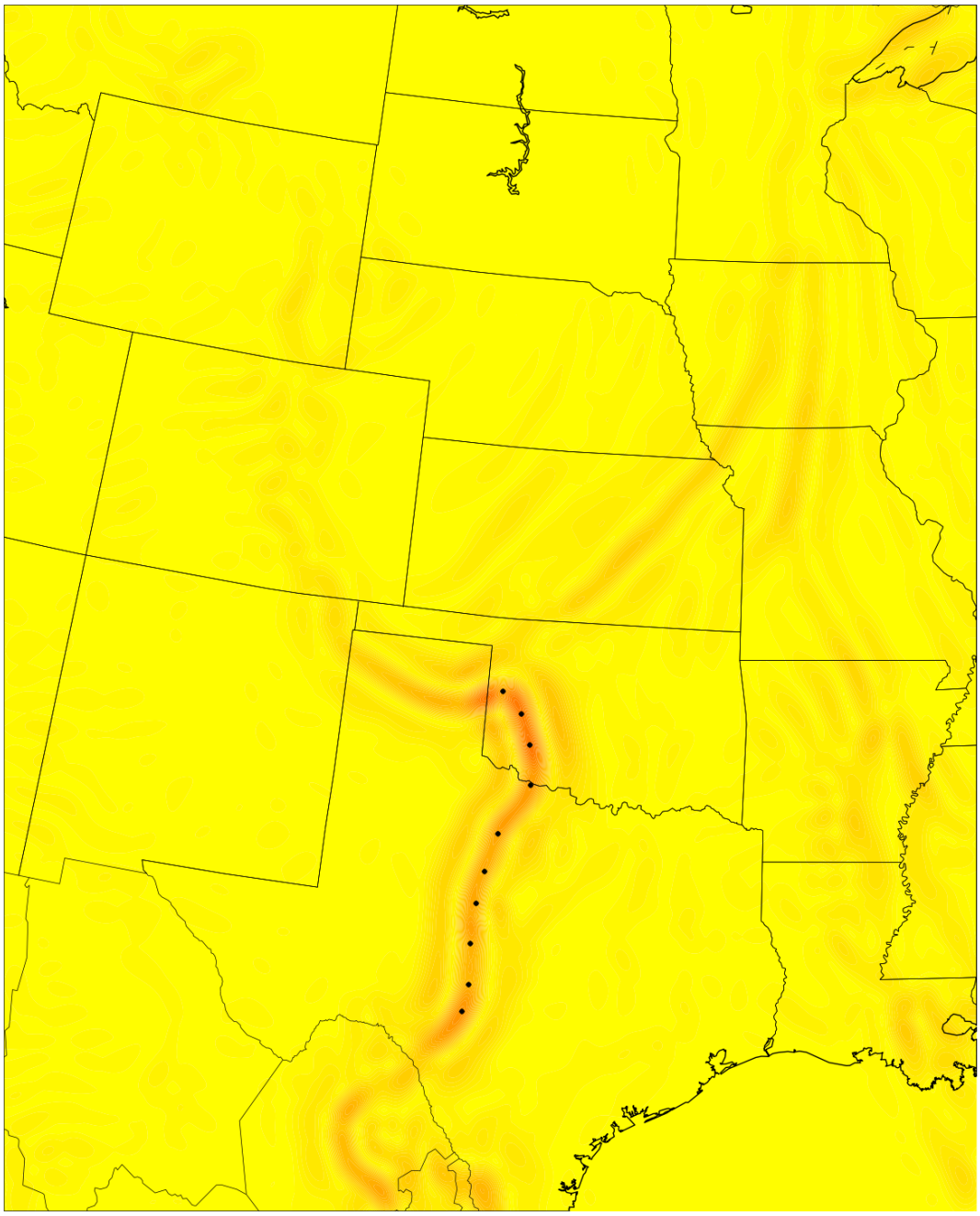


Figure A.13: LoG of specific humidity max - min, sigma = 6.0, calculated using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 22 April 2008.

Sobel and LoG perform similarly when applied to wind direction (not shown). There are two key differences that should be noted, however. First, both Sobel and LoG have a significant increase in the amount noise present in their

output. This is expected, especially in the presence of low wind speeds. The slower winds enable local, smaller scale forcings to more easily override the larger scale flow, resulting in greater variation of local directional values. This is particularly clear in observational data, although it is merely an artifact of instrumentation in some cases. Second, the stationary front seen in the data has a much stronger wind shift profile than that of the dryline, while the cold front has values similar to those of the dryline. These types of values are seen consistently throughout the dataset, suggesting that a simple threshold on the wind shift is insufficient for ruling out boundaries similar to drylines.

Virtual potential temperature generally remains relatively constant across the dryline regions, while showing larger variations along regions of the cold front. To investigate the consistency of the virtual temperature gradient across the boundary, slices were sampled every 50 km along the length of each feature identified in the truth dataset. The process also allowed for examination of along dryline variation of other variables. The slices were extracted by repeatedly traveling 50 km along a great circle arc between dryline segments as denoted by the truth dataset. A 140 km line perpendicular to and centered on the dryline was placed at each stop. Data was interpolated to 100 evenly spaced points along each line, with basic statistics such as mean and standard deviation calculated and combined with individual slices to create composite plots (e.g., Fig. A.14). Examination of the slices generally showed the structure expected of drylines; however, it confirmed that assuming drylines to always have an approximately constant virtual temperature structure is not realistic.

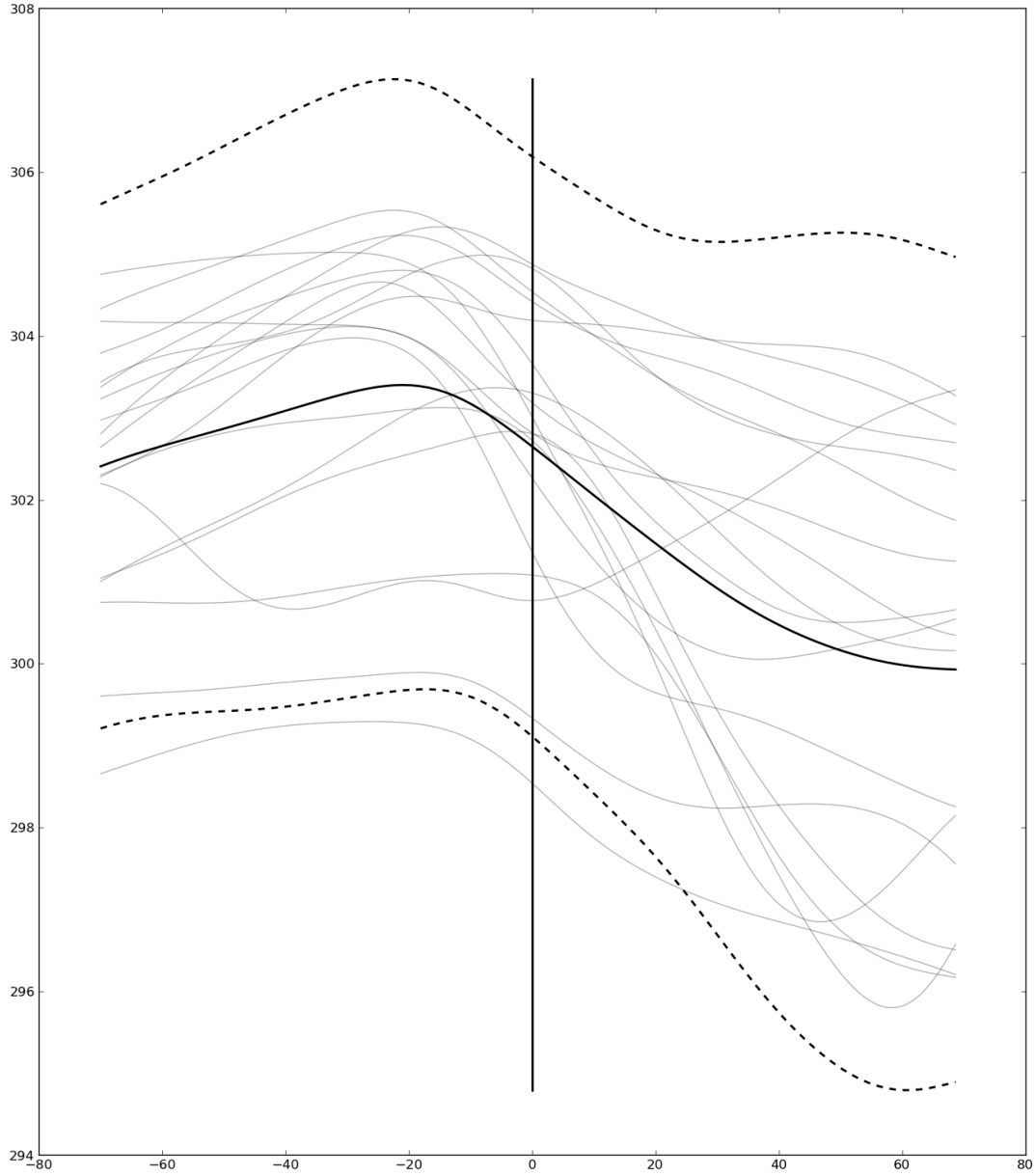


Figure A.14: Slices of across dryline virtual potential temperature (K) (y coordinates), created using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 26 April 2009. Negative (positive) x coordinates indicate depth into the dry (moist) region (km). Light solid lines indicate sampled slices across the dryline. The dark, wavy line represents the mean of the slices, and the dashed lines show plus and minus two standard deviations. The dark, vertical line denotes the position of the dryline.

A.2 Dry to moist wind shift

As described in chapter 2.1, drylines are typically associated with confluent wind fields. The source regions of the air masses that form drylines in the US result in the mean flow along the boundary being oriented approximately 90° counterclockwise of the gradient vector of the moisture field. The dry side wind typically has a stronger east-west component than the moist, although this may change based on the dryline's overall orientation. The across dryline wind direction gradient magnitudes (i.e., the wind shift) were examined by using the specific humidity gradient orientation as a proxy for the boundary. Wind directions were then evaluated using a variety of methods (e.g., comparing a mean wind direction to dryline orientation, dx vs dy gradient components); however, the results lacked in consistency, regardless of implementation. While most spurious features were removed, the true drylines were occasionally broken apart such that they no longer formed a continuous feature. These situations are likely caused by outflow from nearby storms eroding the original wind structure of the dryline; however, synoptic scale features (e.g., a passing short wave) could manipulate the surface wind field such that it was no longer recognizable as being associated with a dryline.

A.3 Dry Warm, Moist Cool

Temporal information was generally ignored in this study due to a combination of data limitations and a desire to keep the final algorithm as generalizable as possible. However, several diurnal assumptions were examined in the hopes of using them to create a set of criteria ideal for the near peak solar heating environment that could then be relaxed for application to other times. One of

the more promising assumptions examined is that the dry region of the dryline will be warmer than the moist from mid-afternoon to late evening. During the course of the day, solar radiation heats the area surrounding the dryline. However, the high specific heat of water vapor results in the moist region warming less than the dry. In addition, the moist side heating is inhibited by the greater amount of vegetation typically present. Finally, the low level moisture increases the likelihood of stratiform cloud formation, which would reduce the amount of solar radiation reaching the surface. This general structure has been consistently seen in multiple observational studies (e.g., Ziegler and Hane, 1993; Crawford and Bluestein, 1997). It may also play a role in the apparent variation of virtual potential temperature gradients seen in various studies (e.g., McGuire (1962) vs. Parsons et al. (1991)) (Crawford and Bluestein, 1997). Unfortunately, the diurnal assumption was occasionally overridden by strong synoptic scale effects (especially when the dryline had a strong east-west orientation) as well as local events. As with the dry to moist wind shift discussed above, the output lacked consistency. The algorithm was able to identify dryline regions, but often only after breaking the continuity of the features.

A.4 Skeletonization and filter banks

In appendix A.2, we examined the use of the across boundary wind shift in dryline identification. Similar to the wind shift, confluence is also a decent indicator of dryline presence, although it doesn't have any particular use in removing other boundaries from consideration. Unfortunately, when combined with the specific humidity gradient threshold, a confluence threshold results in

breaks in the defined boundary (although not to the extent the two implementations above). To counteract this issue, a new technique was devised. The process involved slowly extending the ends of the identified regions until they reconnected. A filter bank of oriented Gaussian filters (appendix B.1) was used to accomplish this, resulting in preferential expansion along the direction of the "lines". However, the process needed to be repeated multiple times in order to be effective, with the regions widening on each iteration. The smoothing process was therefore paired with skeletonization (appendix B.3) after each iteration, reducing the lines to a single pixel in width. While skeletonization risked removing some of the extension caused by the smoothing, a sufficiently elongated Gaussian was able to overcome the loss (although this, in turn, risked sampling less from the line during a region with high curvature, again risking skeletonization overriding the lengthening process). While the output typically correctly identified the drylines, it was overly optimistic, capturing stray strands extending from the main region (e.g., Fig. A.15) as well as other boundaries.

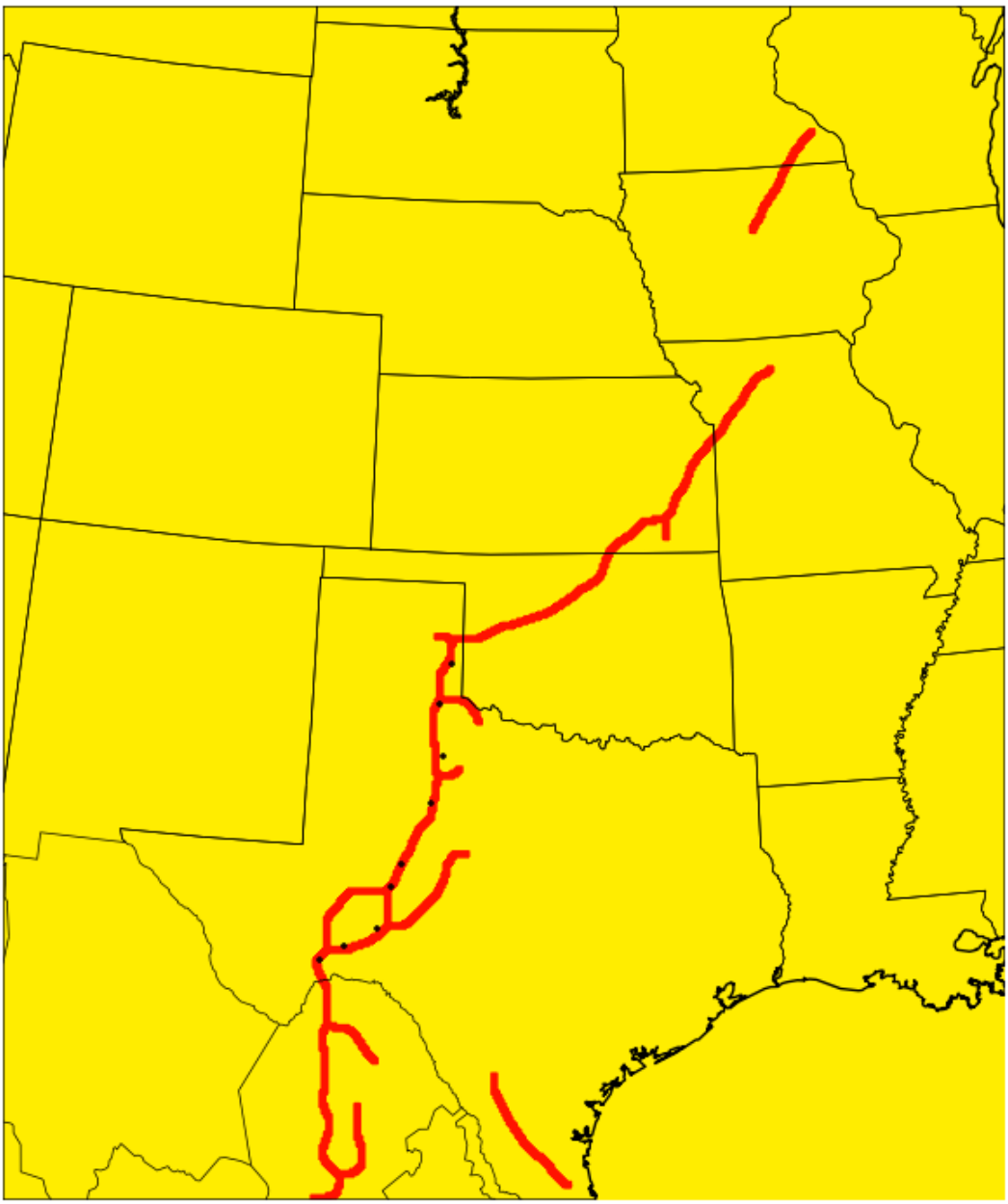


Figure A.15: Skeletonization and filter banks output, performed using 24 hour forecast data from the NSSL-WRF valid 0000 UTC 04 April 2010.

Appendix B

Data Processing Overview

The following provides an overview of various techniques that were not used in the final algorithm, but were key to its development.

B.1 Filter Banks

When smoothing data, an anisotropic Gaussian filter may be desired if the general shape of the feature to be identified is known (e.g., Lakshmanan, 2004). Consider the identification of tightly packed vertical stripes. A Gaussian filter with a large vertical extent would likely be of more benefit than an isotropic Gaussian as it would reduce the influence of neighboring stripes during smoothing. However, if the orientation of the lines were not known, a series of filters can be applied in succession, each designed for a specific orientation. The version that best reduced the local deviation of a pixel's value would be retained (the output of different filters may be kept for different parts of the image). The individual filters combine to form a filter bank, or a set of filters designed to capture different features at different scales, orientations, etc.

B.2 Laplacian of Gaussian

The Laplacian of Gaussian (LoG) is an edge detection technique designed to apply across multiple scales (Marr and Hildreth, 1980). The method utilizes zero crossings of the second order derivative associated with changes in spatial intensity to identify edges (i.e., where the value changes from positive to negative or vise versa). The LoG is written as:

$$LoG = \vec{\nabla}^2(G * I) \tag{B.1}$$

where $\vec{\nabla}^2$ is the Laplacian operator, G is a Gaussian matrix, and I is the input data matrix. However, the equation is more commonly rewritten as

$$LoG = \vec{\nabla}^2(G) * I \tag{B.2}$$

to increase computational efficiency.

The accuracy of LoG is dependent upon the condition of linear variation which states that "the intensity variation near and parallel to the line of zero-crossings should be locally linear" at the examined scale (Marr and Hildreth, 1980). When met, the condition ensures that linearly retrieved zero crossings are true zero crossings. Otherwise, the calculated zero crossings are subject to an offset defined by the true nature of the variation (e.g., Fig. B.1). In addition, the discrete implementation of LoG is subject to both bias and zero crossing errors due to truncation, as demonstrated by Gunn (1998). However, the errors can be reduced by increasing the size of the convolution mask at the cost of computational efficiency.

In addition to identifying zero crossings, a maximum - minimum filter may be applied to the smoothed second order derivative to identify edges. While the precision of the operation is decreased, the size of the footprint will act to provide supplementary filtering of small features.

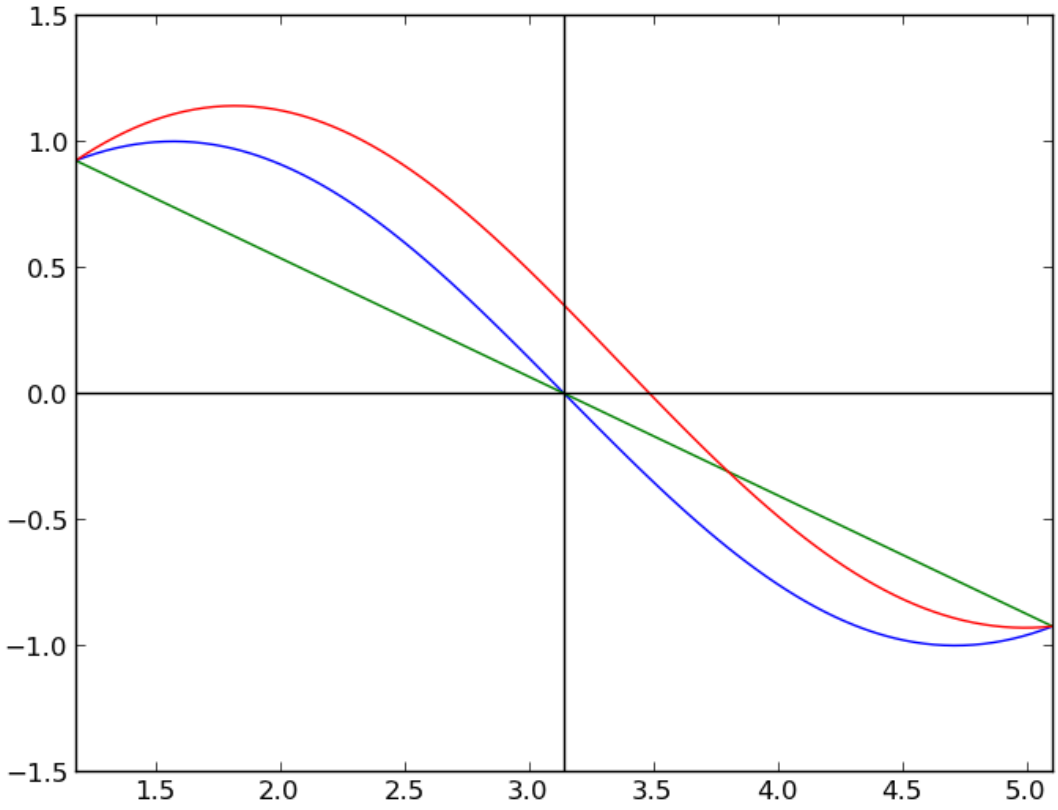


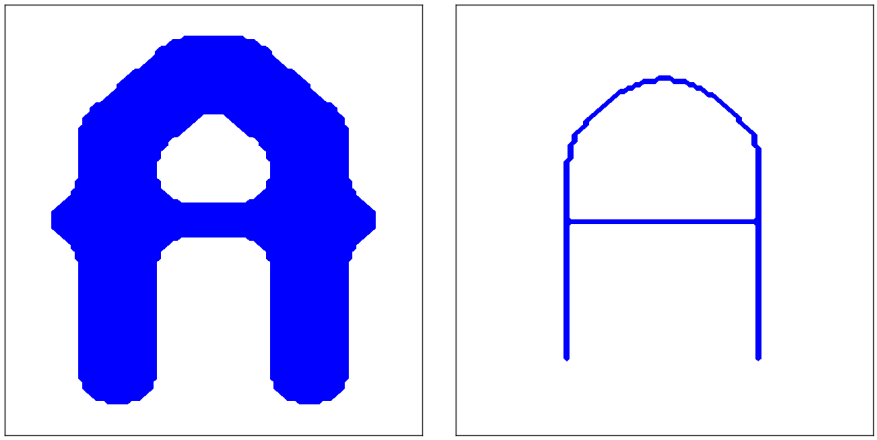
Figure B.1: LoG Zero Crossing estimation. The vertical black line indicates the estimated position of the zero crossing. The green line (linear) results in a correct assessment of the zero crossing. Despite its nonlinearity, the blue line is likewise correct. However, the red line will result in the incorrect identification of the zero crossing location.

B.3 Skeletonization

Skeletonization is an image processing technique designed to reduce a binary image to a series of lines representative of the original shape. Depending on

the particular implementation, skeletonization may also be called a medial axis transform. That is, the image is reduced to its medial axes (i.e., a connected region of those points existing at the median of the nearest edges). The output can be achieved through either a single pass through the image or an iterative process (e.g., Zhang and Suen, 1984). An example of the process can be seen in Fig. B.2. It should be noted that while skeletons can be fairly sensitive to noise, they are not significantly impacted by it in the situations occurring in this study. The objects being skeletonized in this project are thin and line-like. As the structures are already only a few pixels wide, there is little opportunity for noise to impact the results. However, it should be noted that this may not hold true for lower resolution models as the deviation of a single pixel will have different spatial impacts.

Another benefit of skeletonization occurs when the process is applied to line-like structures (e.g., the trace of a tree branch). In this case, a secondary set of algorithms can be used in conjunction with skeletonization to identify endpoints and intersections within the resulting image. This allows for the removal of tiny, nub-like branches, in addition to several other applications (not discussed here).



(a) Unprocessed blob. (b) Skeletonized blob.

Figure B.2: Unprocessed and skeletonized blob. Note that the cross bar of the A protrudes from the sides of the unprocessed image, but is lost during skeletonization.