The significance of land use, topography, antecedent rainfall, and atmospheric conditions in relation to summertime convective storm initiation in the North Alabama region

Christopher Tracy

Follow this and additional works at: https://louis.uah.edu/uah-theses

Recommended Citation
https://louis.uah.edu/uah-theses/382

This Thesis is brought to you for free and open access by the UAH Electronic Theses and Dissertations at LOUIS. It has been accepted for inclusion in Theses by an authorized administrator of LOUIS.
THE SIGNIFICANCE OF LAND USE, TOPOGRAPHY, ANTECEDENT RAINFALL, AND ATMOSPHERIC CONDITIONS IN RELATION TO SUMMERTIME CONVECTIVE STORM INITIATION IN THE NORTH ALABAMA REGION

by

CHRISTOPHER TRACY

A THESIS

Submitted in partial fulfillment of the requirements for the degree of Master of Science in The Department of Atmospheric and Earth Science to The School of Graduate Studies of The University of Alabama in Huntsville

HUNTSVILLE, ALABAMA

2022
In presenting this thesis in partial fulfillment of the requirements for a master’s degree from The University of Alabama in Huntsville, I agree that the Library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by my advisor or in his/her absence, by the Chair of the Department or the Dean of the School of Graduate Studies. It is also understood that due recognition shall be given to me and to The University of Alabama in Huntsville in any scholarly use which may be made of any material in this thesis.

Christopher Tracy  5/23/2022
(student signature)   (date)
THESIS APPROVAL FORM

Submitted by Christopher Tracy in partial fulfillment of the requirements for the degree of Master of Science in Atmospheric Science in Atmospheric Science and accepted on behalf of the Faculty of the School of Graduate Studies by the thesis committee.

We, the undersigned members of the Graduate Faculty of The University of Alabama in Huntsville, certify that we have advised and/or supervised the candidate of the work described in this thesis. We further certify that we have reviewed the thesis manuscript and approve it in partial fulfillment of the requirements for the degree of Master of Science in Atmospheric Science.

Committee Chair

05/27/2022

(Date)

Department Chair

5/31/2022

College Dean

5/31/2022

Graduate Dean

Digitally signed by Rainer Steinwandt
Date: 2022.06.01 13:48:25 -05'00'

Digitally signed by Jon Hakkila
DH: zmJon_Hakkila, or-UF, or-Associate Provost/Graduate Dean, email: zmJon_hakkila@uh.edu, cr-US
Date: 2022.06.16 16:17:58 -05'00'

Adobe Acrobat version: 2020.033.10334
ABSTRACT

The School of Graduate Studies
The University of Alabama in Huntsville

Degree         Master of Science              College/Dept.   Atmos. & Earth Science
Name of Candidate                                         Christopher Tracy
Title The Significance of Land Use, Topography, Antecedent Rainfall, and Atmospheric
Conditions in Relation to Summertime Convective Storm Initiation in the North
Alabama Region

In the Southeast U.S., summer days with abundant “pop-up” thunderstorms are
common and strong synoptic forcing is not needed to initiate them. The goal of this study
is to explain and ultimately enhance predictability of summertime convective initiation in
these tranquil environments. Spatial non-randomness arises, with greater event density
appearing over high elevation by midday. Late in the day, event counts subside with
another mechanism emerging (urban heat island). Antecedent rainfall, instability, and
moisture are higher on average where convective initiation occurred. In terms of feature
importance, elevation is more important in the early to mid-afternoon while antecedent
rainfall and wind direction consistently are the most important overall. Future work
includes implementation of additional features, average soundings over areas of differing
event density, and further machine learning analysis.
ACKNOWLEDGEMENTS

The funding for this research was provided by the University of Alabama in Huntsville (UAH) Earth Systems Science Center (ESSC) through director John R. Christy. First, I want to thank my advisor and committee chair, Dr. John R. Mecikalski, for the offer to perform this research and the valuable guidance throughout the project. I also extend my gratitude to my other two committee members, Dr. Kevin R. Knupp and Dr. Udaysankar Nair, for the helpful feedback on the project and willingness to be on my committee.

In addition to the above people, I have a few others to recognize for their contributions. Christopher Jewett gave useful insight on efficiently downloading the radar data from the online daily archive. My officemate David Haliczer was always around to answer any questions I had, whether about my research or my project code. All of the professors who taught courses I was enrolled in have helped in enhancing my academic understanding of several main meteorological tenets (synoptic, radiation, dynamic, and thermodynamic meteorology, along with numerical weather prediction and professional development). I also want to acknowledge Denise Berendes, who helped me with using the Matrix high-performance computing (HPC) cluster for big data processing. Lastly, I would like to thank my family and colleagues (both new and old) for being there for me over the course of my graduate tenure at UAH.
# TABLE OF CONTENTS

| List of Figures                                                                 | ix |
| List of Tables                                                                  | xii|

## CHAPTER 1. INTRODUCTION AND MOTIVATION

1.1 Introduction.............................................................................. 1
1.2 Research Components & Hypothesis........................................... 3
1.3 Potential Significance of Study............................................. 4

## CHAPTER 2. LITERATURE BACKGROUND

2.1 Thunderstorm Classification................................................... 6
2.2 Convective Initiation Definition............................................. 11
2.3 Radar Echo Detection............................................................. 11
2.4 Geostationary Satellite......................................................... 13
2.5 Convergent Boundaries........................................................... 14
2.6 Machine Learning & Convective Echoes...................................... 16
2.7 Geographical CI Mechanisms................................................... 17
2.8 Boundary Layer Processes........................................................ 22
2.9 Evapotranspiration................................................................. 24

## CHAPTER 3. DATA AND METHODS

3.1 Getting the Data & Study Period............................................. 26
3.2 Radar Data for CI Detection................................................... 29
3.3 RAP Analysis............................................................................. 31
3.4 Elevation Data.......................................................................... 33
3.5 Land Use Data

3.6 Antecedent Rainfall Data

3.7 CI Algorithm Integration & Aspects

3.8 Spatial Statistical Methods

3.9 Random Forest Model for Feature Importance Analysis

CHAPTER 4. RESULTS

4.1 Atmospheric Analysis of Two Case Days

4.2 Spatial CI Event Distributions

4.2.1 Early Day Group

4.2.2 Middle Day Group

4.2.3 Late Day Group

4.3 Cumulative CI Event Heat Maps

4.4 CI Event Frequency Relationships of Static Features & Wind Direction

4.4.1 Elevation and Elevation Gradient

4.4.2 Land Use

4.4.3 Wind Direction

4.5 CI versus Non-CI Antecedent Rainfall & RAP Analysis Field Distributions

4.5.1 Antecedent Rainfall

4.5.2 Surface-Based CAPE

4.5.3 2-Meter Dewpoint Temperature

4.5.4 Surface-Based Lifted Index

4.5.5 Surface Temperature

4.5.6 950 hPa Vertical Velocity
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wind Direction Standard Deviation</td>
<td>101</td>
</tr>
<tr>
<td>Wind Speed</td>
<td>102</td>
</tr>
<tr>
<td>Higher Event Density Subregion</td>
<td>104</td>
</tr>
<tr>
<td>Subregion Elevation</td>
<td>105</td>
</tr>
<tr>
<td>Subregion Elevation Gradient</td>
<td>107</td>
</tr>
<tr>
<td>Subregion Select RAP Model Features</td>
<td>111</td>
</tr>
<tr>
<td>Random Forest Importance</td>
<td>114</td>
</tr>
<tr>
<td>Feature Importance over Whole Domain</td>
<td>115</td>
</tr>
<tr>
<td>Feature Importance Over Subdomain</td>
<td>121</td>
</tr>
</tbody>
</table>

**CHAPTER 5. DISCUSSION**

**CHAPTER 6. CONCLUSIONS**

**REFERENCES**
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1. A conceptual model of the life cycle of a typical air mass thunderstorm in a low-shear environment</td>
<td>8</td>
</tr>
<tr>
<td>2.2. Seasonal average precipitation amounts due to isolated precipitation features over the Southeast U.S. from 2009-12</td>
<td>10</td>
</tr>
<tr>
<td>2.3. In (a), average convective cloud (CC) scores of each MODIS land class over the Southeast U.S. for the July 8-14 2006 period between 1500-1900 UTC (left)</td>
<td>21</td>
</tr>
<tr>
<td>2.4. Spatial distributions of satellite-derived cloud frequency (a) and convective cloud frequency (b, defined as cloud tops &lt; -41°C) at 1200 CST for the June-August 1986 time period over part of the Deep South</td>
<td>23</td>
</tr>
<tr>
<td>3.1. A hierarchal clustering dendrogram of similar features selected for the present study using the Ward’s linkage algorithm, with linkage distance (smaller distance at node = stronger clustering) on the y-axis (a) and corresponding Spearman correlation matrix (showing feature intercorrelation) of those same features (b)</td>
<td>28</td>
</tr>
<tr>
<td>3.2. The spatial domain over which radar (for determining a CI event) and feature data was collected for each case day, indicated by the red box</td>
<td>38</td>
</tr>
<tr>
<td>3.3. Map of Alabama counties for future reference in the spatial analysis (Geology.com)</td>
<td>39</td>
</tr>
<tr>
<td>4.1. Spatial plots (a-f) at 1800 UTC from two case days of: MRMS reflectivity at 10°C in units of dBZ (a, d), NCEP Reanalysis (NARR) 2-meter dewpoint in units of Celsius with 10-meter wind barbs in knots (b, e), and NARR 500 hPa (mb) geopotential height in units of meters with 500 hPa wind barbs in knots (c, f)</td>
<td>46</td>
</tr>
<tr>
<td>4.1. (continued) 0000 UTC soundings from the Birmingham (BMX) site on each of the case days are shown in (g, h) with derived parameters</td>
<td>47</td>
</tr>
<tr>
<td>4.2. Spatial plots of CRM elevation (a) and calculated elevation gradient (b)</td>
<td>51</td>
</tr>
<tr>
<td>4.3. MODIS land use data over Alabama</td>
<td>53</td>
</tr>
<tr>
<td>4.4. Spatial CI event counts for the early (1600-1900 UTC) day group overlaid on elevation (a-c)</td>
<td>57</td>
</tr>
</tbody>
</table>
4.4. (continued) Spatial CI event counts for the early (1600-1900 UTC) day group overlaid on MODIS land use (d-f).................................................................58

4.5. As in Figure 4.4a-c, but for the middle day time group (1900-2200 UTC).................61

4.5. (continued) As in Figure 4.4d-f, but for the middle day time group (1900-2200 UTC).................................................................62

4.6. As in Figure 4.4, but for the late time group (2200-0000 UTC).............................65

4.7. Spatial heat maps of CI events tallied in ~4 x 4 km bins for the early (1600-1900 UTC) (a), middle (1900-2200 UTC) (b), and late (2200-0000 UTC) (c) time groups over all case days.................................................................70

4.8. Binned elevation (x-axis) versus relative CI event frequency (CI%) of each bin (y-axis) with second-degree polynomial regression curves overlaid..........................74

4.8. (continued) CI versus non-CI elevation box plots for all three time groups............75

4.9. As in Figure 4.8.a-c, but for elevation gradient (binned every 20 meters/pixel)........78

4.9. (continued) As in Figure 4.8.d, but for elevation gradient.....................................79

4.10. MODIS land class versus relative CI event frequency (CI%) of each class over all case days.............................................................................82

4.11. As in Figure 4.10, but for 10-meter wind direction..............................................85

4.12. CI versus non-CI one-day antecedent rainfall box plots for all three time groups.........................................................................................88

4.13. As in Figure 4.12, but for two-day antecedent rainfall.......................................89

4.14. As in Figure 4.12, but for five-day antecedent rainfall.......................................91

4.15. As in Figure 4.12, but for surface-based CAPE...............................................93

4.16. As in Figure 4.12, but for 2-meter dewpoint.....................................................94

4.17. As in Figure 4.12, but for surface-based LI......................................................96

4.18. As in Figure 4.12, but for surface temperature...............................................98

4.19. As in Figure 4.12, but for 950 hPa (mb) vertical velocity ($\omega_{950}$)....................100

4.20. As in Figure 4.12, but for 10-meter wind direction standard deviation ($\Theta_{\text{std}}$)....102
4.21. As in Figure 4.12, but for 10-meter wind speed…………………………………104

4.22. Inset plot of high event density subregion in the early/middle time groups (red box)………………………………………………………………105

4.23. As in Figure 4.8, but only with elevation data at grid points over the subregion……107

4.24. As in Figure 4.23, but for elevation gradient (binned every 10 meters)…………110

4.25. As in Figure 4.12, but for surface-based CAPE in the early and middle time groups over the subregion…………………………………………………………………112

4.26. As in Figure 4.25, but for 2-meter dewpoint……………………………………113

4.27. As in Figure 4.25, but for ω950…………………………………………114

5.1. A conceptual model of summertime pulse CI over USGS topography (The University of Texas at Austin) in north Alabama based on the results of this study………………………………………………………………………………130

5.2. A 3D version of the conceptual model in Figure 5.1 over a smaller spatial scale for the early day group, with additional features identified…………………………131

5.3. A schematic of a mesoscale differential heating circulation over land…………..132
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1. All 36 case dates selected for the CI analysis</td>
<td>28</td>
</tr>
<tr>
<td>3.2. MODIS land use classifications and their respective numerical assignments</td>
<td>38</td>
</tr>
<tr>
<td>4.1. Early day group Gini feature importance values over the entire domain for four different random states, an exclusion run w/ three most important and two least important features excluded, and feature average permutation importance at a random state of 42 over five repetitions</td>
<td>116</td>
</tr>
<tr>
<td>4.2. As in Table 4.1, but for the middle day group over the entire domain</td>
<td>118</td>
</tr>
<tr>
<td>4.3. As in Table 4.1, but for the late day group over the entire domain</td>
<td>121</td>
</tr>
<tr>
<td>4.4. As in Table 4.1, but for the early day group over the subregion</td>
<td>123</td>
</tr>
<tr>
<td>4.5. As in Table 4.1, but for the middle day group over the subregion</td>
<td>124</td>
</tr>
</tbody>
</table>
CHAPTER 1. INTRODUCTION AND MOTIVATION

1.1 Introduction

Thunderstorms are one of the most common hazards in global weather. They are typically initiated by a convergence mechanism in moderate-to-high amounts of low-level moisture, enough of which to develop an initial cumulus cloud into a towering cumulonimbus cloud capable of producing heavy rain, lightning, strong winds, and even hail. Once a vigorously growing cumulus cloud reaches the cumulonimbus stage, a convective thunderstorm is imminent (Roberts and Rutledge 2003). With the presence of thunderstorms over an area, it is essential that weather forecasters know their motion and intensity trends so that downstream warnings can be issued. Even a storm that is short-lived (< 30 min.) can cause abrupt disruptions to daily life. The unpredictable development, behavior, and evolution of these types of storms is seen in the summer months during the afternoon hours in the Southeast U.S. Despite recent meteorological advances in technology and research, the predictability of these “pulse” storms in weak synoptic environments is still low relative to storms that are initiated in stronger synoptic environments.

In Brown and Arnold (1998), a weak synoptic environment is defined by the following criteria: no frontal mechanisms within 500 km, 500 hPa winds < 7.5 m s\(^{-1}\) (~15 knots), surface winds < 5 m s\(^{-1}\) (~10 knots), and surface dewpoint temperatures > 17ºC. These criteria were applied in this study with the 500 hPa wind threshold extended slightly: **500 hPa winds < 20 knots, surface winds < 10 knots over the study domain.** Even with recent improvements in ensemble-based numerical weather prediction (NWP)
of severe thunderstorm initiation through data assimilation from a wide array of meteorological observations (radar, satellite-based, aircraft, surface observations, etc.), there is still a lack of confidence in the statistical relationships between NWP guidance and resultant forecast skill in relation to precisely predicting convective conducive ness (Coniglio et al. 2019).

Large-scale atmospheric patterns are not the sole driver of convective storms. It is not intuitive to think of geographical features (e.g., elevation and land use) as significant factors that can directly influence the timing and location of pulse storm formation. Past studies, to be discussed in-depth in Chapter 2, have looked at these mechanisms related to pulse convective initiation (CI). Two key discoveries are that upslope winds over heterogeneous terrain can induce mechanical lifting of air (e.g., Kuo and Orville 1973; Weckwerth et al. 2014) and land surface heterogeneity can produce differential heating boundaries that support localized convergence (Wilson and Schreiber 1986; Yan and Anthes 1988; Segal and Arritt 1992; Hong et al. 1995; Trier et al. 2004). A combination of these two can act in unison to form areas of enhanced convective formation (e.g., Wilson and Schreiber 1986). The Southeast U.S. has considerable mesoscale terrain features that should be studied further, such as the foothills of the Appalachian Mountains (which span from New York down to northern Georgia/Alabama). With these factors considered, the potential exists for non-random diurnal summertime pulse convective patterns over the Southeast U.S., with mesoscale variations in feature variables sought after. However, specific interactions on the microscale (e.g., turbulent eddies in the boundary layer) with any larger mesoscale boundaries are not able to be resolved in this study. Still, a comprehensive summertime convective storm climatology that incorporates
both geographical (e.g., Gambill and Mecikalski 2011) and non-geographical (meteorological) variables over the Southeast U.S. in an intercomparison spatial and importance analysis has yet to be performed. The latter includes antecedent rainfall, which can have an effect on evapotranspiration levels over a land area, and RAP model analysis. Similar convective climatology studies revolving around convective echo identification utilizing radar and/or satellite have been conducted over multiple regions including the Black Hills, Colorado High Plains, and the Brazilian Amazon (Kuo and Orville 1973; Klitch et al. 1985; Lima and Wilson 2008). With this central idea, several primary questions arise which serve as the foundation of this study.

1.2 Research Components & Hypothesis

This study is centered around several components. First is the cumulative spatial distribution patterns of summertime CI events across the study region. Then, it is examined how distinct elevation features (< 1000 m) and land use over the study region influence pulse CI events over the Southeast U.S. Next, it is determined whether summertime CI is significantly correlated with locations that recently experienced rainfall over prior days through the presence of higher amounts of antecedent rainfall amid weak background winds. The last component is how important geographical features are relative to meteorological features in pulse CI over the study period. The primary goal of this 2020-2021 summertime analysis is to explain and ultimately enhance predictability of summertime CI in tranquil synoptic environments over the Southeastern U.S. This analysis was accomplished through a qualitative assessment of the degree of spatial randomness in tallied CI occurrences over the study region, along with quantitative
statistical analyses of different diurnal timeframes between CI events and features across a set of case days. Then, any features that are more significantly correlated with pulse CI events are examined further and compared with relevant previous studies.

The proposed hypothesis for this study is that summertime pulse CI events in the Southeast U.S. occur in a non-random manner and static features (terrain and land use) are most important in dictating whether pulse CI occurs, especially early in a given day. The remaining meteorological features would not show any significant relation with CI in part due to their non-static nature. If none of the examined features are found to be significantly correlated with pulse CI over the study period, then it is to be inferred that these CI events are expected to be randomly distributed across the region.

1.3 Potential Significance of Study

The outcomes have the potential to advance understanding of summertime pulse CI event patterns in the Southeast U.S. Conceptual models that describe the most significant mechanisms in pulse storm formation can be developed. Regarding short-term numerical weather prediction, eventual regional and seasonal adjustments could be made to convective parameterizations within NWP models such as the Global Forecasting System (GFS) and North American Mesoscale Model (NAM). Using the meteorological and geographical features assessed in this study, convective indices can also be developed based on real-time weather conditions and specific geography at a location (with potential integration into existing convective nowcasting systems). Thus, the results of this study could lead to broad applications in the operational realm of meteorology when it comes to synoptically tranquil and humid environments. In addition, future
research will also likely need to be conducted in order to expand upon unanswered aspects of the work.
CHAPTER 2. LITERATURE BACKGROUND

2.1 Thunderstorm Classification

The summer season (late May – early September) over the eastern U.S. consists of lush vegetation, humid conditions (dewpoint temperatures > 17°C, after Gambill and Mecikalski 2011), and convective potential. The poleward retraction of the mid-latitude jet stream strongly supports a long-lasting diurnal convective cycle in these months with a plentiful supply of warm, moist air in the lower levels of the atmosphere. There are three primary convective storm types that occur in the U.S.: supercellular, multicellular, and ordinary cells (also termed “pulse” or “air mass”).

The first storm type, supercells, can form a mesoscale environment of their own and produce significant tornadoes and destructive hail. These are typically fueled with moderate-to-high instability, strong synoptic forcing, and high vertical wind shear oriented near-perpendicular to a convergent boundary. The multicellular classification can also involve a considerable amount of synoptic forcing and includes sub-classifications such as mesoscale convective systems (MCS) and derechos (can bring destructive wind damage over a long path). Most multicellular storms are forced on the synoptic scale with a moderate amount of vertical wind shear. These multicellular systems can either take the form of broken meso-β (spatial/temporal scale of 20-200 km and 1-6 hours, respectively) or nonbroken meso-α (spatial/temporal scale of 200-2000 km and > 6 hours, respectively) linear structures, with longer system duration maintained by new cells constantly initiating along a gust front on the south end of the system (Markowski and Richardson 2010). Embedded cells within a multicellular system are
harder to track than individual discrete cells, which can take advantage of large
directional wind shear and lack of interference from neighboring storms in order to
remain discrete for a long period of time.

The third type, which is the storm type of focus in this study, is typical in an
environment with moderate-to-high convectively available potential energy > 1000 J/kg
(CAPE) and low vertical wind shear < 20 knots (Markowski and Richardson 2010). This
type of convection is known for producing outflow (or “gust fronts”), a localized area of
cold, dense air (high pressure) that propagates outward from the storm. This results in
high Bulk Richardson number values, which implies a dominance of outflow over inflow
in a thunderstorm and thus a tendency for pulse storms to be rather short-lived (Weisman
and Klemp 1982; Rotunno et al. 1988; Markowski and Richardson 2010) in comparison
to longer-lived/larger-scale storms such as supercells (e.g., Wilson 1966). The number of
these storms tends to peak in the early to mid-afternoon hours, in concert with the highest
amount of diurnal heating from incoming solar radiation (e.g., Lima and Wilson 2008,
Rickenbach et al. 2015). Figure 2.1 shows the evolution of this type of storm in a limited-
shear environment. When a positively buoyant air parcel reaches its level of free
convection (LFC), it will likely continue to rise and form a towering cumulus cloud.
Eventually, it matures into an ordinary storm cell with a defined updraft. Since an
ordinary cell tends to be strongly aligned in the vertical direction (lack of vertical updraft
tilt associated with wind shear), eventually the increase in evaporative cooling caused by
falling precipitation helps accelerate the stage.
The predominance of different storm types also varies by region and season. Precipitation climatology studies over the Southeast U.S. have been performed, with clear diurnal patterns being observed. For one, summertime convection typically prevails onshore over land during the day (especially after 1800 UTC) and shifts offshore at night, which is very apparent over the Florida Gulf Coast (Rickenbach et al. 2015). As diurnal heating becomes a predominant factor in afternoon convective development, scattered ordinary cells can form at a rapid pace. As evident in Figure 2.2, this isolated daytime convection has a distinct seasonal trend in the Southeastern U.S. Maximum rainfall amounts attributed to this convection type are seen in the summer months (Figure 2.2c), whereas high rainfall amounts from mesoscale precipitation systems are more common in the winter and spring months (Figure 2.2a, b). These mesoscale precipitation systems account for 70-90% of the total annual precipitation despite being much less abundant in
the summer months (Rickenbach et al. 2015). Not only are spatial relationships in storm type observed in the Southeast precipitation climatology, but trends in intensity and duration as well. On average, rainfall events with heavier rain rates were found to peak in areas along the Gulf Coast from 1960-2017, decreasing further inland from the coast (Brown et al. 2019). In addition to this, a significant portion (82%) of a set of precipitation-measuring stations in Brown et al. (2019) showed a decrease in the average duration of precipitation events across the Southeast region over this same time period. This observed rise in the number of short-lived, intense rain events is worth noting, especially in relation to the short-lived nature of pulse CI events in the summertime months.
Figure 2.2 Seasonal average precipitation amounts due to isolated precipitation features over the Southeast U.S. from 2009-12. Units are in mm day$^{-1}$. Scale range of average isolated precipitation is depicted in the color bars. The four subplots show month ranges of December-February (top left, a), March-May (top right, b), June-August (bottom left, c), and September-November (bottom right, d). Figure adapted from Rickenbach et al. (2015).
2.2 Convective Initiation Definition

In order to effectively conduct the analysis, a set definition for CI needs to be applied. Some previous studies have defined CI using a radar reflectivity threshold of 30 dBZ at the 1-km height level (Schreiber 1986; Wilson and Schreiber 1986). Another valid definition is when an isolated radar echo reaches 35 dBZ at the -10°C isotherm level. In a convectively conducive environment, the 35-dBZ threshold has multiple atmospheric features and applications. Convective cloud-top glaciation and ice nucleation processes can occur at this level, which can both act as a strong indication of ongoing and/or deepening convection (e.g., Schreiber 1986; Roberts and Rutledge 2003; Mecikalski and Bedka 2006). This CI threshold has also been used to track convective echoes over time. Patou et al. (2018) adapted a satellite tracking procedure for clouds through the use of bounding boxes around cloud “objects”, assigning each one a unique numerical index value based on the conditional presence of other clouds in and around the box edges. The 35-dBZ threshold has been applied not only on a point-by-point basis, but also over a defined minimum storm area for tracking convective echoes (e.g., Lima and Wilson 2008). With this the case, using solely this threshold in the prediction of CI is still unproven and should only be used for present CI identification in an isolated growing echo (Wilson and Schreiber 1986).

2.3 Radar Echo Detection

In meteorological applications, Doppler radar is capable of measuring various properties and characteristics of atmospheric targets within a sample volume through backscattered microwave radiation. For instance, the presence of high amounts of large
Cloud hydrometeors will return a much stronger signal back to the radar than a cloud with smaller hydrometeor sizes, hence why reflectivities > 35 dBZ play a key role in identifying convective echoes. Over an area scannable by an S-band radar, which has a pulse wavelength of ~10 cm, the first appearance of non-convective cloud echoes can actually occur with reflectivities < 0 dBZ (Knight and Miller 1993). These estimations do not come without drawbacks such as boundary layer clear-air scattering from insects and Bragg scattering above the boundary layer that can result in detectable signals (Wilson et al. 1994). Bragg scattering occurs with target separation at around half the radar wavelength and primarily stems from variations of the refractive index in a well-mixed boundary layer, maximizing chances for constructive wave interference. Higher returns from Bragg scattering are seen at larger wavelengths (e.g., S-band) where higher sensitivity thresholds are needed (Knight and Miller 1993). Since a single radar cannot effectively classify all mantel echoes as a result of the Bragg scattering, growing echoes in their early stage have been classified as cumulus congestus (Knight and Miller 1993; Mueller et al. 2003). Low-level moisture corrections performed using radar refractivity data from individual radars were found to improve prediction of future CI in a favorable synoptic setup with a reduction in forecast error from the cycling of moisture refractivity fields (Gasperoni et al. 2013). Thunderstorm tracking has also been done using radar. In Nisi et al. (2018), a thunderstorm tracker implementing the 35-dBZ threshold was used for tracking of hail cores within strong thunderstorms over the European Alps, with longer tracks found in westerly and southwesterly synoptic flow regimes. A main reason for this outcome is the correspondence of these two flow regimes in the Northern Hemisphere with well-defined synoptic support that can sustain longer-lasting storms.
Shorter storm tracks were also found in the study during the afternoon hours when air mass convection was the dominant storm mode (Nisi et al. 2018).

2.4 Geostationary Satellite

The use of geostationary satellite data fields has proven to be effective in applications such as cloud frequency studies (e.g., Gibson and Vonder Haar 1990; Gambill and Mecikalski 2011) and the tracking of convective echoes through the examination of various cloud-top properties. For instance, a CI lead time of 30-45 minutes can be attained with the use of several Geostationary Operational Environmental Satellite (GOES) data fields in combination with the 35-dBZ threshold (Mecikalski and Bedka 2006). Using a group of eight CI predictors in the study, a GOES pixel that satisfied at least seven of the eight CI predictor thresholds was defined as having a high likelihood of CI. When supplemented with a cumulus cloud mask, channel differencing can be used to identify cumulus echoes of interest. This is how several of the IR predictors in Mecikalski and Bedka (2006) were derived. Better performance of this CI nowcasting algorithm was also found in the study under synoptically tranquil conditions from CI pixels having higher correlation with local reflectivity maximums within a cumulus cloud of interest.

Similarly, Mecikalski et al. (2008) also used the 35-dBZ threshold to calculate skill scores for several of the IR fields, with maximum optimization when only three or four of the interest fields were considered. Additionally, a low false alarm rate was found with cloud-top glaciation ~30 min. before CI detection (Mecikalski et al. 2008). These rapid cloud-top temperature drops are a solid indication of CI and are thus important for
distinguishing stronger convection from weaker convection. Typically, ~30 min. can pass between a cloud top reaching the freezing level (0°C) and CI (Roberts et al. 2003). Along with increases in cloud ice content, the cloud-top temperature drops can coincide with high rain rates (Patou et al. 2018). The incorporation of NWP meteorological information into a GOES-R CI framework, which included the use of statistical models and radar CI detection, was found to reduce the number of false alarms (Mecikalski et al. 2015).

Unlike radar, the use of GOES radiance data comes with parallax error, which stems from the angle of a satellite scan relative to the nadir point (constant point location of the GOES satellite aligned with the equator). This source of error, causing a displacement cloud echoes in the opposite direction of the satellite, is more extreme at higher latitudes (above 50°N in Northern Hemisphere) and geometric corrections need to be made to account for it.

2.5 Convergent Boundaries

Convergent boundaries are a CI mechanism that has been previously studied. One particular type of boundary is outflow from existing convection, which can aid in the initiation of new storms where it collides with the background flow outside of the thunderstorm. This type of mechanism is not supported by a weakly-sheared environment, however, due to the lack of pronounced lift associated with the enhanced shear and vertical dynamics (Markowski and Richardson 2010). Despite this, multiple convergent boundaries that collide in this environment can still provide a mechanism that can trigger additional convection. Various convergent boundaries have been observed with radar and surface analysis data in eastern Colorado, including the Palmer Divide
among other boundaries brought about by the intersection of descending mountain flow and moist, warmer air from the east (Wilson and Schreiber 1986). These can be mechanisms of CI even under suppressed synoptic conditions. In addition to radar data, the preceding formation of cumulus clouds can assist in convective nowcasting (Mueller and Wilson 1989; Wilson and Mueller 1993). With a propagating density current, updraft tilt is maximized when the propagation more or less opposes the low-level wind (Wilson et al. 1998).

Outflow collision over a favorable land use or orographic area has also been observed as a convective aide (Wilson and Schreiber 1986). In Lean et al. (2009), an isolated thunderstorm initiated in the south of England due to the combination of a convergence line induced by terrain, low-level moisture associated with a nearby front, and diurnal heating. Through the use of radar, satellite, and surface observation data, Goggins et al. (2009) identified various frontal boundaries that could act as CI mechanisms during one summer in the National Weather Service (NWS) Birmingham County Warning Area (CWA). In the study, convection was categorized into both autoconvection (a single boundary initiated without the aid of other boundaries) and convection initiated from the interaction of multiple boundaries. In Lima and Wilson (2008), outflow propagation from existing convection was found to be a more prevalent CI mechanism in the later afternoon hours, implying that different mechanisms prevail over different periods of the day.

Convergent boundaries can also result from a land-sea diurnal pressure gradient involving inland motion of the sea breeze. In the Florida summer season, the interaction
of the sea breeze with winds further inland has been found to cause the formation of north-south oriented convective bands that shift with the time of day as the sea-breeze pattern evolves (Gerrish 1971). These bands contain embedded updraft maxima associated with convective cumulus within the broader sea breeze frontal band (Wilson et al. 1994).

2.6 Machine Learning & Convective Echoes

Machine learning methods have been used in past studies for thunderstorm tracking. Convergent boundaries can be manually identified and entered by a forecaster into automatic forecasting systems such as the Auto-Nowcaster (ANC). When compared with the ANC, human involvement in convective nowcasting improves both detection rates and accuracy while simultaneously not drastically increasing the number of false identifications (Roberts et al. 2012). The ANC incorporates numerous predictors based on the surrounding atmospheric environment with boundaries able to be discovered and cumulus cloud properties derived. With the use of fuzzy logic functions these are converted into interest fields to rank in terms of importance in the formation of a convective nowcast. These importance values were then summed to form a convective nowcast field, with the regime with pulse convection showing the lowest skill due to its overall unpredictable nature (Roberts et al. 2012). Mueller et al. (2003) used the 35-dBZ threshold for an intercomparison of different nowcasting techniques, with a significant improvement in the performance of the ANC compared to extrapolation methods on the days when convection is not synoptically forced (e.g., when large-scale advection does not have a large role in convective forcing). The ANC also had a higher critical success
index (CSI) and probability of detection (POD) compared to the simpler storm extrapolation methods, albeit at the cost of a higher false alarm rate (Mueller et al. 2003). It is of note that this system does not account for static geographical predictors such as elevation and land use.

In convective nowcasting, there is the additional challenge of the coexistence of both newly initiated and existing storms. The issue is not exclusive to this study, so it is addressed during the CI identification process. Although CI along a convergent boundary is a common mechanism of ordinary cell CI, no boundary identification was performed in this study. This is not to say, however, that a gust front-type feature or other kind of boundary did not have an influence on the formation of observed CI. In Mecikalski et al. (2015), thermodynamic predictors (e.g., convective inhibition and surface-based CAPE) from data over 21 severe convective case days made up six of the seven highest importance rankings out of all NWP predictors in a CI nowcasting random forest model. Expanding upon the machine learning aspect, the use of a thunderstorm artificial neural network (TANN) model in Collins and Tissot (2015) actually performed worse than both a multiple linear regression model and human forecasters over several different cases in a multiple-skill score comparison for several Texas mesoscale regions. The studies here have shown that there are imperfections in automated forecasting and more research still needs to be done regarding CI detection, including that of the ordinary type (Wilson and Mueller 1993). Hence, the human forecaster is still an essential tool in convective forecasting.
2.7 Geographical CI Mechanisms

Other studies have looked at geographical CI mechanisms. For example, slight perturbations of the wind can influence convective probabilities. Hirt et al. (2019) found that by adding horizontal and vertical wind perturbations into 1-km cloud-resolving models a higher amount of CI can be induced over regions collocated with higher elevation gradients, with slower perturbations increasing the likelihood of CI further due to the maintenance of stronger updrafts (too intense of perturbations can destroy a convective updraft). Gambill and Mecikalski (2011) discovered that, on average, convective cloud (CC) frequencies are positively correlated with elevation gradient despite the low number of higher-gradient locations relative to flatter areas (Figure 2.3b), neglecting frontal boundary influences within 500 km.

Diurnal convective patterns can also stem from local ridge-valley variations in the terrain that are dominant during El Nino years, as opposed to larger-scale moisture circulation influences during La Nina years (Giovannettone and Barros 2008). Convective echo formation in parts of Mexico is dictated at the regional scale by topographic heterogeneity, along with minimal ocean-mountain range distance and the orientation of the mountain ranges (Giovannettone and Barros 2008). The study used GOES IR data at a temporal data resolution of three hours, which is sufficient for depicting general convective patterns but cannot effectively capture as many specific CI events as possible. Not accounting for topographical patterns over a region has also been found to lead to oversimplification of these precipitation patterns and thus needs to be considered (Brown et al. 2019). CC frequency over Alabama was discovered to be at a
maximum in the mid-afternoon around 3:00 CST in Gibson and Vonder Haar (1990). Just as with other regions, overall cloud minima and maxima tend to shift with the time of day over the Southeast U.S. with trends toward cloud minima in valleys (e.g., parts of northeast Alabama such as around Guntersville Lake) and maxima over higher-terrain features after 11:00 CST (Gibson and Vonder Haar 1990). Over the course of a summer radar case study in the Black Hills of western South Dakota, areas of higher convective echo frequencies that coincided with southeast upslope motions, differential heating, and either southwest or northwest flow at 500 hPa (corresponding to northeast and southeast echo maxima respectively) were discovered on the leeside of the Black Hills (Kuo and Orville 1973). Hence, when combined with other mechanisms synoptic patterns can have a profound influence on CC locations. In Brazil, westerly and easterly flow regimes can exhibit different convective characteristics; the former is associated with synoptic setups with shallower/less intense thunderstorms, while the latter is associated with a monsoon period with more intense thunderstorms (Lima and Wilson 2008). In the French Vosges Mountains, warm southeasterly upslope flow on the leeside of the mountain range can collide with westerlies from the windward side bringing in moister air, eventually spilling over to the leeside in the afternoon hours (Weckwerth et al. 2014). This convergence pattern helps initiate convection over the Vosges.

Not all boundaries with steep elevation gradients are guaranteed to form convection under the proper atmospheric conditions, however. Out of all identified convergence boundaries in Goggins et al. (2009), topographical “boundaries” were found to be convective ~49% of the time, a significantly lower percentage than the identified synoptic-scale fronts. This contrasts with the results from Koch and Ray (1997), which
found that topographic boundaries were convective ~80% of the time (likely due to sea breeze fronts near the coast, which were placed into the topographical category in the study). Steep terrain was found to be the dominant mechanism during the early afternoon hours in central Brazil, while outflow boundaries were the dominant mechanism overall (Lima and Wilson 2008). In addition, land type and heterogeneity can locally enhance thermal buoyancy and lifting condensation level (LCL) variability, in turn increasing the probability of CC formation (Gambill and Mecikalski 2011). Forest and savanna have been previously correlated with higher CC percentages relative to other classes in the Southeast U.S. in Figure 2.3a (Gambill and Mecikalski 2011). Through the use of land surface models, differential soil moisture and latent heat flux gradients over a region can be discovered. Thermal solenoidal circulations result from land use heterogeneity on scales of 10-20 km, dampened at smaller scales due to the greater amount of turbulent mixing as opposed to the pressure gradient force resulting from said circulation (e.g., Avissar and Liu 1996; Collins and Tissot 2015). Heating indices surrounding this phenomenon have been formed through the integration of land use and GOES-R data, not without drawbacks such as changes in vegetation, overestimation of latent heating through the assumption that precipitation only occurs in the mid-afternoon, and the exclusion of crop irrigation impacts (Walker et al. 2009).
Figure 2.3 In (a), average convective cloud (CC) scores of each MODIS land class over the Southeast U.S. for the July 8-14 2006 period between 1500-1900 UTC (left). A higher CC score equates to a greater number of CC occurrences for a particular land class. On the right, land classes versus average CC% (blue diamonds) and domain percentage (red circles) across all examined time periods. The only differing land classification here from the MODIS dataset used in the present study is water bodies (class 0 here, class 17 in present study). Same is done for elevation gradient (b). Figure adapted from Gambill and Mecikalski (2011).
2.8 Boundary Layer Processes

The importance of mesoscale and microscale diurnal boundary layer processes in weak synoptic environments cannot be overlooked. Beginning in the early morning hours, surface heating due to incoming solar radiation is an essential driver in the formation of a defined convective mixed layer (CBL). The CBL, which grows as the day progresses, has a vertical potential temperature (θ) profile that stays near-constant with height (pseudo-adiabatic). Solar insolation heats the surface and generates positive buoyancy and higher convective instability within the lower atmosphere because of an underlying unstable surface layer underneath where the frictional force prevails, promoting local-scale ascent of the parcels until they eventually reach their LCL and form clouds near the top of the CBL (Stull 1988). They can continue to rise as long as the parcels are warmer than their surrounding environment. This early convective cumulus formation is critical for the moistening of the boundary layer prior to CI (Weckwerth et al. 2014). Factors such as lingering stratus clouds overhead can hinder the instability and work to inhibit CBL growth. Another potential inhibitor is morning convection and excessively wet surfaces, but these two aspects are actually not a major issue in the summertime months with fast atmospheric recovery times given the high intensity of solar insolation.

Ordinary convection formation is guided by a favorable CBL and directly enhanced by low-level moisture and/or mesoscale circulations such as sea breezes (Rickenbach et al. 2015). Sea breezes, such as along the U.S. Gulf Coast, can travel up to 100 km inland and typically stall south of the red line shown in Figure 2.4, showing the
prevalence of midday daytime cumulus over land based on GOES-6 satellite analysis from the summer of 1986 (Gibson and Haar 1990). Convective echoes at that time are still confined to the Gulf Coast region within the influence region of the sea breeze. Since this phenomenon is not of focus in this study, the red line serves as the southern bounds of the study domain. Lake breezes can also be a local driver of CI. In the Tennessee River Valley, Lake Wheeler has been found to produce summertime thermal lake breeze circulations that result in precipitation downstream of the lake (Asefi et al. 2012).

Figure 2.4 Spatial distributions of satellite-derived cloud frequency (a) and convective cloud frequency (b, defined as cloud tops < -41°C) at 1200 CST for the June-August 1986 time period over part of the Deep South. Red lines depict southern bound of the domain used in the present study. Figure adapted from Gibson and Vonder Haar (1990).
Between the Rocky and Appalachian Mountains, semidiurnal forcing mechanisms tend to drive the formation of warm season precipitation systems (Carbone et al. 2002). With solar heating and continuous forced lifting in the daytime CBL in this region, propagation of existing precipitation systems is not the lone source of convection (Wilson and Schreiber 1986; Carbone et al. 2002). Once the evening hours commence and diurnal heating begins to wane, low-level atmospheric stability increases as existing pulse convection begins to dissipate. This occurs from an absence of larger-scale synoptic forcing, which can sustain individual storms for longer durations. The average lifetime of these pulse storms is roughly 30-60 minutes, based on the approximate duration sum of both the duration of the ascension of air to the anvil top and average time it takes precipitation to reach the ground (Markowski and Richardson 2010). However, when the pulse type is weakly supported by wind shear they can persist for ~2-4 hours.

2.9 Evapotranspiration

Evapotranspiration can have a significant impact on low-level moisture content and local weather patterns in ensuing days especially when surface soil moisture is high and vegetation cover is dense. Evapotranspiration itself is defined by the United States Geological Survey (USGS) as the process by which water is either lost to the atmosphere, evaporated from groundwater reservoirs, or transpired from plants on the land surface. A common meteorological phenomenon that helps raise evapotranspiration levels over an area is falling precipitation. When hydrometeors reach the ground, they either seep into the ground, evaporate off leaf surfaces, and/or become stored in plants that retrieve the water through their roots. On any given convective day, higher antecedent rainfall totals
are thus concurrent with greater available low-level (below ~900 hPa) moisture to fuel surface-based convection. Other factors that exhibit positive correlation with evapotranspiration levels include temperature (stomata open up more) and surface wind (increased flow of moisture off the surface and plants). Since these two factors are included in this study and tend to go hand-in-hand with evapotranspiration, each of their relationships with CI can be verified from the collected data. One variable that has a negative correlation with evapotranspiration is ambient relative humidity (lower relative humidity = better conditions for evaporation). Even though bodies of water (lakes, rivers, oceans) are a dominant producer of atmospheric moisture (90%) globally, transpiration over land contributes as much as 10% to the low-level moisture (USGS).

The remaining sections of this paper are defined as follows: Chapter 3 presents the data used, how it is prepared for the data processing and analysis, and the methods of the CI detection algorithm along with the statistical methods chosen for the analysis; Chapter 4 reveals and displays the results of the study; Chapter 5 contains a discussion, and Chapter 6 has the summary and conclusions of the study.
CHAPTER 3. DATA AND METHODS

3.1 Getting the Data & Study Period

In order to conduct the analysis, the proper convective days had to be selected. This involved the use of both synoptic weather maps and radar to determine whether a daily setup was conducive for widespread pulse storm CI and if pulse storm CI actually occurred, respectively. Using the NEXRAD mosaic of base radar reflectivity on the Iowa Environmental Mesonet, a daily search was conducted over the study domain over the summers of 2020 and 2021. The mosaic images, integrating a composite of nationwide WSR-88D radars at 1-km spatial resolution and 5-minute temporal resolution, were examined daily from mid-May to early September. On a convective day, scattered-to-widespread pulse convection appears as individual cells of higher radar reflectivity, indicative of high amounts of large hydrometeors and large rain rates characteristic of tropical-like environments. During these months, the primary upper-level jet stream shifts up into the northern U.S. and Canada which helps sustain the tranquil synoptic conditions [as defined in Brown and Arnold (1998)] for an extended period of time. Cooler-season days with stronger flow aloft (> 30 knots) provide advective motion to any existing storms and support longer-lived updrafts and interference between separate storms, especially when growing upscale along a boundary.

Stronger synoptic flow and frontal boundaries can facilitate a transition to an advective storm mode. Depending on the normal component of the bulk wind shear vector and cloud-layer mean wind relative to a synoptic front, frontal CI can lead to rapid upscale growth and/or longer-lasting discrete storms that tend to move with the front
(Dial et al. 2010). In addition, morning convection can lead to storm congealment and
delayed/limited morning solar heating which can overturn the atmosphere and reduce
instability. Other inhibitors are a lack of low-level moisture up to 850 hPa and
widespread cloud cover (limits instability and boundary layer heating).

During the months of May-August, Storm Prediction Center morning meso-
analysis was analyzed to confirm a calm synoptic pattern encompassing the entire depth
of the troposphere [in the typical environment examined in this study, the 200 hPa level
resides above the scale tropospheric height of ~10 km as referenced in Holton and Hakim
(2013)]. With minimal advective effects, false echoes are greatly reduced. A total of 36
case days were selected (Table 3.1). Given that summers 2020 and 2021 were not
unusually dry, this sample is representative of typical non-drought summers. The end
months (May and September) have the fewest days due to shorter duration of the tranquil
pattern, especially in their early and later parts respectively. All features are sub-
categorized into hierarchal clusters based on their inter-feature correlations (Figure 3.1).
The Ward linkage distance between clusters is calculated based on variance
minimization. Notable collinear feature clusters are elevation/elevation gradient, wind
speed/wind direction standard deviation, antecedent rainfall, and dewpoint/lifted
index/surface-based CAPE. The strong negative correlation of lifted index with dewpoint
and CAPE is due to its negative scale (stronger lifted indices are more negative). Other
broader clusters include land use/elevation/elevation gradient and wind direction/surface
temperature/950 hPa vertical velocity.
Figure 3.1  A hierarchical clustering dendrogram of similar features selected for the present study using the Ward’s linkage algorithm, with linkage distance (smaller distance at node = stronger clustering) on the y-axis (a) and corresponding Spearman correlation matrix (showing feature intercorrelation) of those same features (b). Common clusters are displayed as a single color in the hierarchy, with the y-axis as the linkage distances. Feature data from the 1600-1900 UTC time period across all case days was used in the making of these two visuals.
Table 3.1 All 36 case dates selected for the CI analysis. Days are sorted by month.

<table>
<thead>
<tr>
<th>May</th>
<th>June</th>
<th>July</th>
<th>August</th>
<th>September</th>
</tr>
</thead>
<tbody>
<tr>
<td>05/25/2020</td>
<td>06/18/2020</td>
<td>07/05/2020</td>
<td>08/05/2020</td>
<td>09/11/2020</td>
</tr>
<tr>
<td>05/27/2021</td>
<td>06/19/2020</td>
<td>07/06/2020</td>
<td>08/06/2020</td>
<td>09/12/2020</td>
</tr>
<tr>
<td>06/11/2021</td>
<td>06/24/2021</td>
<td>07/07/2020</td>
<td>08/07/2020</td>
<td></td>
</tr>
<tr>
<td>06/27/2021</td>
<td>06/30/2021</td>
<td>07/09/2020</td>
<td>08/08/2020</td>
<td></td>
</tr>
<tr>
<td>07/05/2020</td>
<td>07/18/2020</td>
<td>08/05/2021</td>
<td>08/11/2020</td>
<td></td>
</tr>
<tr>
<td>07/06/2020</td>
<td>07/19/2020</td>
<td>08/12/2020</td>
<td>08/11/2021</td>
<td></td>
</tr>
<tr>
<td>07/07/2020</td>
<td>07/20/2020</td>
<td>08/11/2020</td>
<td>08/11/2021</td>
<td></td>
</tr>
<tr>
<td>07/09/2020</td>
<td>07/21/2020</td>
<td>08/12/2020</td>
<td>08/11/2020</td>
<td></td>
</tr>
<tr>
<td>07/18/2020</td>
<td>07/22/2020</td>
<td>08/19/2020</td>
<td>08/11/2020</td>
<td></td>
</tr>
<tr>
<td>07/19/2020</td>
<td>07/25/2020</td>
<td>08/30/2020</td>
<td>08/12/2020</td>
<td></td>
</tr>
<tr>
<td>07/20/2020</td>
<td>07/26/2020</td>
<td>08/05/2021</td>
<td>08/11/2020</td>
<td></td>
</tr>
<tr>
<td>07/21/2020</td>
<td>07/27/2020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>07/15/2021</td>
<td>07/24/2021</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>07/26/2021</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.2 Radar Data for CI Detection

For CI identification, isothermal 2D radar reflectivity data at the –10°C level was obtained from the Multi-Radar and Multi-Sensor (MRMS) online archive (Zhang et al. 2011). The MRMS product suite, made operationally available in 2014, offers numerous 2D, 3D, and 4D dual-polarization radar mosaic products at a two-to-three minute temporal resolution. It integrates 146 different WSR-88D radars and 30 single-polarization C-band radars spread out across the continental U.S. The mosaic scheme that MRMS implements uses two different weighting functions in the horizontal and vertical directions, accounting for the horizontal grid point distance from the radar and the vertical height of the reflectivity bin respectively (Zhang et al. 2011). Its advantage over other methods (e.g., nearest neighbor) is that any discontinuities between the individual
scan regions are eliminated (Zhang et al. 2011). There are also other advantages as opposed to using data from separate radar sites individually. For one, the MRMS scheme avoids terrain interference and beam sampling issues found at longer radar ranges. Additionally, there is the ease-of-use advantage; only a single dataset is needed for each scan time instead of having to work with datasets from each radar site. The \(-10^\circ\text{C}\) isothermal surface is derived from a 3D Reflectivity Cube over each grid point using vertical temperature data from the Rapid Refresh (RAP) analysis. Further quality control eliminates non-hydrometeorological echoes from insects or frontal boundaries. Formatted in binary GRIB2 format, this MRMS data has a spatial resolution of \(~1.11 \times 1.01\ \text{km}\) (0.01° latitude/longitude) on a regular latitude-longitude grid projection. The data is only available within 24 hours of its scan time and obtained during the primary diurnal period of solar insolation between 1600-0000 UTC from the daily MRMS web archive.

For the file processing, a sufficient CI identification interval was defined as 15 minutes for this study. The files were grouped accordingly into these intervals for a total of 32 intervals per case day accounting for the eight-hour daily collection period. Over each 15-minute interval, a grid point is assigned a binary value of 1 if CI occurred at any point during that interval (e.g., exceeded 35 dBZ in at least one of the files). To minimize positive CI instances with existing echoes that span multiple grid points, a tally is only counted at a grid point if no adjacent points reach the 35-dBZ threshold. Once all data was collected, it was grouped into early (1600-1900 UTC), middle (1900-2200 UTC), and late (2200-0000 UTC) bins in order to effectively compare and contrast the dominant CI mechanisms in each of these timeframes, as well as the diurnal evolution of the boundary layer.
3.3 RAP Analysis

Archived hourly Rapid Refresh (RAP) analysis data was also obtained at a 13.54 km spatial resolution. The RAP model and analysis are run and produced, respectively, by the National Centers for Environmental Prediction (NCEP). It incorporates meteorological data from a variety of sources including satellite, aircraft, radar, and balloon soundings on a Lambert Conformal grid projection over the continental U.S. (Benjamin et al. 2016). The data are assimilated through the Gridpoint Statistical Interpolation analysis system (GSI) as opposed to the 3-dimensional variational assimilation method (3DVar) used in the previously-named Rapid Update Cycle (RUC) model (Benjamin et al. 2016). Output files are in binary GRIB2 format. Meteorological features relevant on convective days include moisture, instability, and other triggers such as vertical velocity (Collins and Tissot 2015). The following RAP feature variables were chosen along with their CI feature subcategory:

- 10-meter wind direction (wind feature)
- Local standard deviation of 10-meter wind direction (convergence feature)
- 10-meter wind speed (wind feature)
- Surface temperature (instability feature)
- 2-meter dewpoint temperature (moisture/instability feature)
- Surface-based lifted index (instability feature)
- Surface-based CAPE (instability feature)
- 950 hPa vertical velocity (instability feature)
For each selected case day, the hourly files in the 1600-0000 UTC timeframe were obtained totaling nine analyses per case day. With differing temporal resolutions between the RAP and MRMS, each hourly analysis (except for 0000 UTC) was assigned to the three 15-minute intervals after the analysis time at the top of each hour (00<sup>th</sup>-45<sup>th</sup> minute). The 45<sup>th</sup>-60<sup>th</sup> minute interval of each hour is assigned the analysis of the following hour due to the fact that this particular interval is closer to that hour than the previous hour. As an example, the 1915-1930 UTC interval is assigned the 1900 UTC analysis whereas the 1945-2000 UTC interval is assigned the 2000 UTC analysis. An important assumption made during this process is that each hourly analysis grid is steady state (static) over its assigned hourly interval. Temporal interpolation of the analyses to every 15 minutes was a possibility that was explored but would come with additional assumptions and complications with regards to feature temporal trends.

To re-scale the data to the MRMS spatial resolution, a cubic interpolation was applied using a piecewise Clough-Tocher triangulation curve minimization scheme (Renka and Cline 1984). The local standard deviation of the wind direction is not an included variable in the analysis files, so it was derived from the interpolated wind direction data using a 2D convolution with 3 x 3 kernels around each grid point. The biased standard deviation relation was applied to the convolved kernels and the convolution function was then executed using the following standard deviation relation in Equation (3.1):
\[ \sigma = \frac{1}{\sqrt{N}} \left( \sum_{i=1}^{N} (x_i - \bar{x})^2 \right). \]  

(3.1)

with “x” here being the interpolated wind direction. For the convolution boundary conditions, a default fill value of zero was used. A possible modification for future research is applying a different boundary condition (e.g., circular, symmetrical) to ensure non-zero standard deviations on the domain edges.

3.4 Elevation Data

The second feature dataset is elevation data from the 2001 version of the Coastal Relief Model (CRM; NOAA National Centers for Environmental Information – NCEI 2001), which is ran by NCEI at a spatial resolution of 3 arc-seconds (~0.0008333° latitude/longitude). The model combines water bathymetry and terrestrial topography data into a high-resolution geographical depiction of the coastal regions of the U.S., with the latter data coming from the USGS and Shuttle Radar Topography Mission (NOAA National Centers for Environmental Information 2001). Two different model regions were obtained, including the eastern Gulf of Mexico/Florida (Volume 3) and the central Gulf of Mexico (Volume 4). The two regions, which came in separate files, were combined into a single file for simpler processing. The data files are in netCDF format. Since this dataset comes at a higher spatial resolution than the MRMS data, it was divided into bins of the MRMS shape and averaging was performed over each bin to properly re-scale the two-dimensional grid to the MRMS resolution. Not only is elevation a desirable feature in this study, but elevation gradient as well (an indicator of the
“steepness” of terrain at a location). To calculate the local elevation gradient at each re-binned elevation grid point, a second-order finite difference gradient function was applied across the grid (central difference in the grid interior, one-sided difference at the grid edges) in Equation (3.2) below, valid for evenly spaced data in the zonal direction:

\[
\left( \frac{\partial (\text{elev})}{\partial x} \right)_i = \frac{\text{elev}_{i+1} - \text{elev}_{i-1}}{2\Delta x} + O(\Delta x^2) .
\]

(3.2)

where \( \Delta x \) is the spatial resolution and \( O \) is the truncation error. The same relation was applied to the meridional axis, and the magnitude of both was taken to get the final gradient value. This function was iterated over all binned elevation data points to get unique local gradient values at each point on the grid, with units of meters per unit pixel.

3.5 Land Use Data

The third feature dataset is 2019 Moderate Resolution Imaging Spectroradiometer (MODIS) MCD12Q1 land cover. This classification dataset contains multiple sub-classification sets, including six separate supervised classifications across the global domain from annual MODIS reflectance data (Friedl and Sulla-Menashe 2019). Within the MODIS data file, the Type 1 classification sub-dataset was selected (the annual IGBP classification) for this study, containing 17 different land class labels and an unclassified label (Table 3.2). Refining of the initial classifications was done with post-processing using supporting data information. The data is over a sinusoidal grid at ~463-meter
spatial resolution in HDF4 file format. The Regex geographic package was used to extract the latitude and longitudes from the original coordinate system, as well as isolate the relevant subset of the data that encompasses the study domain. As with the elevation data, the dataset was upscaled by re-binning to the MRMS resolution. However, unlike the elevation data the MODIS data cannot be simply interpolated as it is a classification dataset rather than a continuous one. A different re-binning technique was applied that finds the mode (most occurring land type) of each grid point and surrounding points. The resulting grid contains the assigned land cover modes at each re-binned grid point. Since there is no easy way to find the standard deviation of the MODIS data as an indicator of land variability, it was left out as a variable in this study.
Table 3.2 MODIS land use classifications and their respective numerical assignments.

<table>
<thead>
<tr>
<th>Land Class</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evergreen Needleleaf Forest</td>
<td>1</td>
</tr>
<tr>
<td>Evergreen Broadleaf Forest</td>
<td>2</td>
</tr>
<tr>
<td>Deciduous Needleleaf Forest</td>
<td>3</td>
</tr>
<tr>
<td>Deciduous Broadleaf Forest</td>
<td>4</td>
</tr>
<tr>
<td>Mixed Forest</td>
<td>5</td>
</tr>
<tr>
<td>Closed Shrubland</td>
<td>6</td>
</tr>
<tr>
<td>Open Shrubland</td>
<td>7</td>
</tr>
<tr>
<td>Woody Savanna</td>
<td>8</td>
</tr>
<tr>
<td>Savanna</td>
<td>9</td>
</tr>
<tr>
<td>Grassland</td>
<td>10</td>
</tr>
<tr>
<td>Permanent Wetland</td>
<td>11</td>
</tr>
<tr>
<td>Cropland</td>
<td>12</td>
</tr>
<tr>
<td>Urban Area</td>
<td>13</td>
</tr>
<tr>
<td>Cropland/Natural Vegetation</td>
<td>14</td>
</tr>
<tr>
<td>Permanent Snow &amp; Ice</td>
<td>15</td>
</tr>
<tr>
<td>Barren</td>
<td>16</td>
</tr>
<tr>
<td>Water Bodies</td>
<td>17</td>
</tr>
<tr>
<td>Unclassified</td>
<td>255</td>
</tr>
</tbody>
</table>

3.6 Antecedent Rainfall Data

The fourth feature dataset is daily antecedent rainfall from the NWS Advanced Hydrological Prediction Service (AHPS) (Lin and Mitchell 2005), as used in Walker et al. (2009). The precipitation data is available on the NWS AHPS web archive in netCDF format, obtained by the several NWS River Forecast Centers (RFCs) nationwide and assimilated/mosaicked by NCEP using a combination of radar, rain gauge, and satellite data sources at a spatial resolution of ~4 km on a Polar Stereographic grid projection. The daily files contain multiple sub-datasets: “observed” daily precipitation estimates (temporal sample interval from 1200 UTC on the previous day to 1200 UTC on the
current day), “normal” climatological precipitation values, the difference between the two variables, and the respective relative frequencies. In this study, only the “observed” rainfall is desired. For each selected case day, the daily antecedent rainfall data for the previous five days was obtained totaling five files per case day. In order to properly match each case day with the correct antecedent rainfall data, the files were retrieved using file path datetime similarity indexing between the MRMS and antecedent rainfall files. The rainfall data was then cubically interpolated to the MRMS grid resolution with the same Clough-Tocher scheme as with the RAP analysis data. To ensure the absence of any negative values in the interpolated data, any points on the interpolated grid that are unrealistic (< 0 inches) were set to a value of zero inches. From the five assigned files, three separate feature variables were formed: one-day, two-day, and five-day antecedent rainfall. These three rainfall totals were assumed to remain static on the grid over the course of their assigned days (for all daily 15-minute intervals), only changing once a new case day iteration began.

3.7 CI Algorithm Integration & Aspects

The spatial domain of the study over which the feature data was collected is shown in Figure 3.2. In the spatial analysis, only the Alabama portion of the domain was examined with Mississippi excluded. For the other statistical analyses, the entire domain grid was utilized for an increased sample size. A map of Alabama counties is provided in Figure 3.3 for future reference in the analysis. The domain encompasses central/northern Alabama and east Mississippi, optimized in order to capture elevation features associated with the Appalachian foothills, large urban areas, a diversity of land types, and avoid
major water bodies (e.g., Gulf of Mexico) with diurnal sea breeze patterns. The sampled elevation and land use data over the domain remain static over the grid for the duration of the study period, valid for these two geographical features as they tend to evolve on the time scale of decades or longer (in the case of elevation, centuries or longer).

Figure 3.2  The spatial domain over which radar (for determining a CI event) and feature data was collected for each case day, indicated by the red box. It encompasses eastern Mississippi and a majority of Alabama. Coordinate domain bounds are [90°W, 85.45°W] (longitude) and [31.50°N, 35.00°N] (latitude). Blue star shows Birmingham, red star shows Huntsville.
After the re-scaling of the feature data, their spatial resolutions then match the domain shape of the radar data (350 points along meridional/y-axis by 455 points along zonal/x-axis, approximately 389 km by 430 km respectively) for the purpose of effectively conducting the feature inter-comparison. The zonal distance across the grid varies slightly due to the Earth’s curvature but is still classified as regular in the latitude/longitude coordinate space. Thus, for the feature datasets that are coarser than the MRMS grid (less data points) an interpolation scheme was applied. On the contrary, for

Figure 3.3 Map of Alabama counties for future reference in the spatial analysis (Geology.com).
the datasets finer than the MRMS (more data points) a re-binning method was applied. Only the data points that are within the specified study domain are obtained from their respective files. When the interpolated datasets were extracted, a 0.2° extension of the domain bounds in Figure 3.2 in both the zonal and meridional directions was added in order to avoid the propagation of null values near the domain edges during the execution of the interpolation over the defined grid convex hull. This was done in order to ensure that every grid point over each time instance in the study has an assigned discrete numerical value from each feature dataset.

3.8 Spatial Statistical Methods

All re-scaled data, along with the coordinates and time information of the grid points over each 15-minute interval for all case days, were sent to a single comma-separated file. Each feature variable, the CI binary data, and the supporting information were put into separate columns, either in whole or in subsets, for the analysis. One row corresponded to one 15-minute interval sample with the feature values and supporting information for one CI event. Before analysis was conducted on the feature/CI data it was grouped accordingly into the three time groups (early, middle, late). Cumulative spatial distributions of the CI tallies over the sampled domain were formed over the grid for each of the time groups, summing all CI occurrences over the grid and forming a spatial frequency histogram at the MRMS grid resolution. Spatial CI density “heat maps” with coarser resolution than the point-by-point distributions are also made for side-by-side comparison with the static feature fields (elevation and land use). The main purpose of
these is to show visually where pulse CI tends to occur more frequently during the tranquil synoptic days for each time group.

3.9 Random Forest Model for Feature Importance Analysis

Another desired aspect of this study was a quantitative evaluation of importance of each feature set with respect to the binary CI class. From this, a machine learning approach was implemented. The two sub-datasets that go into a machine learning model are the sample predictors (or features) and the target class(es) (the variable to be predicted or classified). The idea behind a machine learning approach to classification is, given an initial input set of data to train the model, observe the predicted target outcome that the model produces with a separate set of test data. One such model is a random forest, which is made up of multiple decision trees. During the model training, each decision tree can be trained on various subsamples of the input training data rather than the entire dataset at once. In random forest, when each tree is formed, ~20% of the input training datasets is not included. This process is termed “bootstrapping” and it is used for the model runs in this study. With an initial specified set of model hyperparameters, the bootstrapped training prediction/classification sets of each subsample are then averaged out over all decision trees with the intention of optimizing the accuracy of the model, reducing the effects of overfitting, and canceling out prediction errors resulting from the variance of the different decision trees (Pedregosa et al. 2011).

Since the desired outcome here is a binary classification rather than a continuous variable as used for prediction (i.e., a percentage-based prediction), a classification random forest model was the chosen option. The Scikit-learn Python package offers
numerous built-in supervised and unsupervised machine learning modeling methods that can easily be inserted into Python programs, including random forests, support vector machines, and K-means clustering. One key advantage of this Python library over others is shorter runtimes on large datasets in a majority of examined machine learning libraries (Pedregosa et al. 2011). In the Scikit-learn library, a classification random forest is available through the RandomForestClassifier meta estimator function. The default feature importance method in this classifier is the Gini impurity importance, which ranks features based on the average decrease in node impurity over all decision trees and their relative node splitting contribution. Another method valid for classification, permutation-based importance, is an average importance obtained through shuffling of the features and re-running of the model over a particular number of iterations. This latter method avoids issues that come with the Gini method such as being strictly confined to a training set and a bias towards numerical features (many unique values). Both methods are implemented in this study for comparison.

To find the most optimal combination of hyperparameters for the model, an exhaustive grid search over a specified parameter space with a three-fold cross-validation scheme was performed. Fine-tuning the model through multiple exhaustive grid searches can be quite time-consuming and even unnecessary, especially if the mean cross-validated model scores do not show a clear positive correlation with the model complexity (e.g., greater tree depth, higher number of trees). With the very large sample size available for this analysis (> $10^8$ total samples across all three time groups/case days), the necessity of multiple rounds of hyperparameter optimization in order to maximize model performance is not fully justified here. Rather, a 3-fold stratified K-fold
cross-validation technique was applied to a single hyperparameter grid search on the training data of the late (2200-0000 UTC) day group to assess the degree to which the cross-validation scores improve with increases in the maximum tree depth and number of trees (test parameter ranges of 5-20 branches and 50-200 trees, respectively). The most optimal combination of hyperparameters was 200 decision trees and a maximum tree depth of 20. This combination was used for all model runs.

Before the fitting of the random forest model, the following adjustments were made to the features: (1) conversion of surface temperature and dewpoint to Celsius, (2) rounding of the interpolated lifted indices to the nearest integer, (3) the masking of negative (unrealistic) antecedent rainfall totals to a value of zero. The model runs for each time group are as follows: (1) all features included with four different bootstrapping random states over the study domain, (2) three most important and two least important features from the first four runs excluded at a single random state over study domain, (3) all features included with four bootstrapping random states over a sub-region of high CI density, (4) three most important and two least important features from the first four runs excluded at a single random state over the sub-region.
CHAPTER 4. RESULTS

4.1 Atmospheric Analysis of Two Case Days

As detailed before, all case days in this study were selected based on a specific set of meteorological criteria characteristic of tranquil synoptic convective days in the Southeast U.S during the summer months. Although the average atmospheric profile does not match exactly between different case days, key similarities do show up. To get a holistic view of the atmospheric profile on these types of days, various vertical levels were examined. The conditions at the surface and 500 hPa at 1800 UTC (early afternoon) along with the evening 0000 UTC Birmingham (BMX) sounding from two near-calm case days are shown as a quasi-microcosm of the synoptic situations from all sampled days over the study domain. MRMS reflectivity at ~10°C around ~1800 UTC is also shown for both days as a visual of the spatial distribution of convective echoes over the region by the early afternoon hours.

In Figure 4.1a pulse convection formed by 1800 UTC on 19 July 2020 over Huntsville, Birmingham, and other scattered areas in the Cumberland Plateau, while in Figure 4.1b widespread surface winds of 5 knots or less are present throughout the study domain. Wind direction patterns tend toward southwesterly to the north and southeasterly to the south (suggestive of a weak surface high pressure). Over central Alabama the winds are the lightest and most variable. These small wind magnitudes should have a rather insignificant impact on the motion of pop-up CI events that form. Overlaid on the surface map is 2-meter dewpoint temperature, showing widespread values > 19°C (~66°F) across most of the domain. Note the area of slightly lower dewpoints in south-central
Alabama. Ignoring other factors, the overlap in these two observations suggests a relationship between the two. In central and northern Alabama, there are CI event clusters present in and near the edge of where the higher dewpoint values reside. Looking beyond this single spatial distribution at a specific time, the dewpoint is not static partially due to local diurnal mixing processes and will more or less vary throughout the day, even only slightly. If this meteorological variable were indeed important in relation to where pulse storm CI appears, it would manifest in the forthcoming analysis.

The wind pattern must also be light further aloft in the atmosphere, such as at the 500 hPa level. On this day, not only are the surface conditions near-calm, but there is also a lack of definable flow further aloft (Figure 4.1c). Here, 500 hPa wind speeds across the region are no greater than 15 knots and are more geostrophic to the south, indicative of the presence of high pressure aloft as well over the region. The geopotential height gradient across Alabama (increase of ~30 meters from SW to NE) also points to the lack of a dynamic pattern and/or a defined surface low proximate to the domain. A defined shortwave trough propagated across the upper Midwest and Ontario on this day, with tranquil conditions prevalent over the southern U.S. Note the anti-cyclonic spatial pattern of these winds. From this synoptic situation, support for broad-scale atmospheric ascent is minimal and other factors (such as the features in this study) stand out better in influencing spatial CI event patterns.
Figure 4.1 Spatial plots (a-f) at 1800 UTC from two case days of: MRMS reflectivity at -10°C in units of dBZ (a, d), NCEP Reanalysis (NARR) 2-meter dewpoint in units of Celsius with 10-meter wind barbs in knots (b, e), and NARR 500 hPa (mb) geopotential height in units of meters with 500 hPa wind barbs in knots (c, f). NARR data provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, from their website at https://psl.noaa.gov/data/gridded/data.narr.html.
From Figure 4.1d the strongest radar echoes are seen in eastern Mississippi at 1800 UTC on 30 June 2021, with less intense echoes over Alabama compared to 19 July 2020. In Figure 4.1e, the surface wind patterns is quite similar to Figure 4.1b. Once again, note the light southeasterlies prevalent to the south and southwesterlies to the north. The near-westerlies up in Tennessee (north of 35°N) coincided with lower dewpoints at this time, which likely led to CI events being reduced/delayed in that area. On the contrary, pop-up echoes can be observed more abundantly further south. Similar to Figure 4.1b, two local dewpoint minima are present over the southern portion of the region, coinciding with a lack of echoes over those same areas. However, there appears to be less of a spatial overlap between occurring CI events and dewpoint from the lack of echoes over the local dewpoint maximum bullseye seen in W/NW Alabama. This
ambivalence requires a deeper investigation into the relationship between CI events and this feature. Regardless, an environment with widespread dewpoints > 66°F over the region can at least support pulse storm formation, but it is certainly not the sole mechanism.

With broad surface high pressure presiding over the Southeast U.S., the weak 500 hPa flow aloft (< 10 knots) in Figure 4.1f suggests a lack of an organized wave pattern over the region. Similar to 19 July 2020, the geopotential height gradient is oriented from southwest to northeast over the region (contrary to typical synoptic patterns in the Northern Hemisphere mid-latitudes). This height gradient is quite small in magnitude during both of the case days shown here, stemming from the tropical-like airmass occupying much of the Southeast U.S. and the poleward retraction of the dynamic wave pattern up into the northern U.S. and Canada during the summer months. Both of these selected case days are just two examples of the minimal synoptic forcing characteristics of all case days used in the analysis to come.

Atmospheric soundings are an effective gauge of the dynamic and thermodynamic atmospheric characteristics at a location during the time of radiosonde data collection, also acting as an indicator of weather evolution later on. At 0000 UTC on this day, there is modest instability (< 1000 J/kg) on the BMX sounding in Figure 4.1g as solar insolation begins to wane with the diurnal boundary layer still well in place. The strongest sign of a well-mixed CBL is seen in the near-adiabatic temperature decrease with height up to the approximate height of the LCL. Note the very shallow temperature inversion near the surface indicated by the small amount of convective inhibition (CIN)
listed to the right of the sounding, which will only expand into a deeper stable layer as the nighttime hours approach. In the morning hours, the stable layer vanishes as diurnal heating commences early in the day, the instability (CAPE) increasing as a result. In the vertical wind profile, wind speeds throughout much of the troposphere are tranquil (< 20 knots) with the wind speed at 500 hPa even calmer than that of the 1800 UTC analysis in Figure 4.1c. A “tall” atmospheric instability profile shows up on the 0000 UTC BMX sounding in Figure 4.1h, with just over 1000 J/kg. of CAPE and minimal CIN showing the remaining diurnal boundary layer still present (note the lack of a definable stable layer, which still has yet to form here). Low-level temperature and moisture profiles on this day do not diverge much from what is observed on 19 July 2020, with a “bottleneck” appearance at around the 800-850 hPa level. Despite the light winds, an evident convective mixed layer has persisted into the evening hours with the near-adiabatic environmental lapse rate below the LCL. The moistest conditions (nearest to saturation) in both soundings reside at around the 800-850 hPa level in proximity to the LCLs, with drier mid-levels above. This accompanies a tranquil wind profile below 250 hPa, with the mid-levels calmer than both the lower and upper levels. Wind speeds remain < 10 knots over this depth. It is essential to have this well-mixed profile with high moisture content (partially from surface evapotranspiration processes) and steady wind speeds below 800 hPa as it allows for easier lifting of unstable air parcels relative to the surrounding environment (Stull 1988). With the lack of available vertical momentum to mix down to the surface (related to light winds aloft), the winds remain calm near the surface and the environment can persist.
4.2 Spatial CI Event Distributions

Although Alabama does not have mountain ranges of the caliber seen in the western U.S., it does have notable elevation features scattered throughout the state. Figure 4.2a reveals such features in the northeast part of the state. A swath of ridges and valleys, denoted here as the valley-ridge region, extends from Shelby County (near Birmingham) into Cherokee County with the Piedmont region to the south. These regions are part of the Appalachian foothills, which extend northeastward paralleling the eastern seaboard. The mountain chain acts as a barrier between separate atmospheric air masses (e.g., cold air damming from high pressure to the east) and a source of orographic ascent (e.g., Barros and Kuligowski 1998; Nykanen 2008). North/northwest of the valley-ridge region lies the upper plains/Cumberland plateau. This encompasses much of north-central and into northwest Alabama, with higher elevations compared to areas south. Isolated higher-elevation features show up over Jackson and Madison counties (e.g., Monte Sano east of Huntsville). These features are collocated with a swath of high elevation gradients over Jackson County (Figure 4.2b). Pockets of high gradients are also seen across the valley-ridge and Piedmont regions. To the west is the Highland Rim region, covering northwest Alabama and the Tennessee River Valley. The last region covers the southern part of the domain, which from Figure 4.2a lies primarily south of 33°N. Closer to the Gulf Coastal Plain, elevation here is lower and no significant topography features exist.
Figure 4.2  Spatial plots of CRM elevation (a) and calculated elevation gradient (b). Units are in meters for elevation, and meters/pixel for elevation gradient. Circled regions are: Highland rim (red), Cumberland plateau (brown), valley-ridge (yellow), Piedmont (maroon), and coastal plain (purple).

Land use classifications over the domain are shown in Figure 4.3. A few distinct water bodies (yellow/light blue shading) are scattered throughout the state, including Wheeler Lake and Guntersville Lake on the Tennessee River, Lake Martin (Tallapoosa/Elmore/Coosa Counties), and Weiss Lake (Cherokee County). On the re-binned grid, this land class only makes up 0.528% of the 159250 total grid points. The light green shading represents developed urban areas (major cities include Huntsville, Birmingham, and Montgomery) and cropland/vegetation (especially prevalent in
northwest Alabama). Respectively, these make up just 0.907% and 2.61% of all grid points. Savannas and woody savannas (classes 8 and 9) are by far the most prevalent classes over the domain, together occupying 55.5% of all grid points. Needleleaf forests (evergreen and deciduous) make up just 0.399%, while broadleaf forests (evergreen and deciduous) make up 23.3% of all grid points. Mixed forest makes up 14.8% of all grid points. Each forest class has a unique shade of blue in Figure 4.3. As a result of the vastly greater frequencies of certain land classes relative to others, a potential bias arises with regards to total CI event count for each land class. Since the CI event counts for the savanna classes will be significantly greater than the other classes, the utilization of a relative CI percentage is more effective in this case (also applied to the other static datasets).
4.2.1 Early Day Group

A key advantage of the static datasets (elevation, land use) over the non-static datasets (RAP analysis, antecedent rainfall) is the ability to overlay the cumulative spatial CI event distributions. During the first hour of the daily sample period, the total CI event count is less in magnitude than later hours as the diurnal heating cycle is still increasing at this time. This trend is quite apparent in the spatial distributions of CI events over all selected case days. Figure 4.4a shows cumulative CI events for the first sample hour (1600-1700 UTC) over the domain. Near Monte Sano just east of Huntsville, a small

Figure 4.3 MODIS land use data over Alabama. Classifications are scaled according to the color bar and are grouped as follows: forestland (blue), shrubland/savanna (dark green), cropland/vegetation & urban area (light green), and water bodies (yellow).
group of high CI counts is present immediately north of this terrain feature. Several areas of interest stand out during this hour. One larger area of higher CI event concentration is located over higher terrain and patchy forestland in the valley-ridge region in southern Blount and western Saint Clair Counties. Another cluster of higher tallies shows up in the heart of the Piedmont region, home to the Talladega National Forest, coinciding with widespread elevation greater than 500 meters, high elevation gradients, and dense forest (as seen in Figure 4.3). Pockets of more frequent CI events also show up in the Cumberland Plateau region in far northeast Alabama over Jackson and Dekalb Counties, containing patches of forestland (Figure 4.4d) along with some of the higher elevation/elevation gradients in the domain. Note the lack of tallies over the non-forest land cover (e.g., savanna), particularly in Dekalb County. Another interesting cluster, somewhat more spread out relative to the others, is seen in the coastal plain region in Bibb and Perry Counties. Here, a lack of definable elevation features and dominant land type of forest exist.

Moving forward to the next hour, Figure 4.4b, e show a higher overall domain count of tallies in the 1700-1800 UTC timeframe compared to the previous hour. Higher CI event counts are now seen over northwest Alabama in the Highland Rim region, including west of the Florence/Muscle Shoals area (labeled accordingly in Figure 4.3). More diverse elevation levels and land types make up this first area of interest, with no clear spatial trend in either variable. There are a greater number of isolated points with CI event counts greater than one dispersed throughout north central Alabama, not associated with any CI event cluster. Higher CI event concentrations are also seen across central Alabama, one such area extending from western Jefferson County down into northern
Perry County. Insignificant elevation features and a mix of forest/savanna are found in this area. The higher CI event signal over Saint Clair County has now expanded further into Etowah and Blount Counties, and extending up through the valley-ridge region. Relative to surrounding regions, the initial cluster in the Piedmont region is not as stout as the previous hour. Interestingly, Talladega National Forest now possesses a lack of CI events compared with areas to the east and south. Jackson County continues to have higher CI event concentrations over higher elevation features in the northwest part of the county, along with eastern Dekalb County. Over western Madison/eastern Limestone Counties, an area dominated by cropland (as opposed to the more common forest and savanna types), a group of locations with greater CI event counts appears over higher elevation than areas to the south (near the Tennessee River).

During the next hour, the diurnal heating begins to approach its maximum. Consequently, Figure 4.4c, f show a higher overall number of CI events over the domain in the 1800-1900 UTC timeframe compared to the previous hour. One area of higher-count locations extends from Shelby County, south of both the Birmingham metro area, down to Autauga/Elmore Counties. Here, Shelby and northern Elmore Counties are especially of focus with spots of higher elevation gradients (e.g., northern Shelby County) and forest being a prevalent land type in these two areas. Another noticeable cluster now covers a majority of Dekalb County in the far northeast corner of the state, coinciding with its wide area of higher terrain and a mix of savanna/developed land types. Outside of this cluster, scattered points of higher CI event counts are seen throughout Dekalb, Marshall, and Jackson Counties. An interesting smaller-scale trend noted here is the lack of CI events over the Tennessee River in north Alabama north of
Lake Wheeler on the southern edge of Limestone County, potentially from local surface cooling by wind off the lake. A small cluster of events appear on the south side of the lake, where locally higher elevation exists and/or an offshore-bound lake breeze can collide with the background wind (Asefi et al. 2012). For the first two hours, other water bodies over the domain also lack evident clusters of high CI event counts. Only a few water bodies eventually have some or all their extent filled in by CI events (e.g., Lake Martin in Tallapoosa County).
Figure 4.4 Spatial CI event counts for the early (1600-1900 UTC) day group overlaid on elevation (a-c). Each “point” on the plots represents a grid point with a non-zero event count in that specific hourly interval across all case days. Some notable CI event clusters are circled in red.
Figure 4.4 (continued)  Spatial CI event counts for the early (1600-1900 UTC) day group overlaid on MODIS land use (d-f).
4.2.2 Middle Day Group

In the middle day group, the overall CI event count across the domain far exceeds that from the early group. Figure 4.5a, d show a continuation of this increasing trend into the mid-afternoon hours (1900-2000 UTC), most noticeably over the higher elevation/elevation gradients in northeast Alabama. Western Jackson County exhibits the largest cluster of CI events, collocated with a dominant forest land type. Central Dekalb County still has a defined grouping of CI events, albeit not to the extent present in Jackson County. Interestingly, the entirety of the Piedmont region now lacks a such grouping. Northern Shelby County still shows a small, dense cluster relative to surrounding areas just south of an area of significant elevation gradients. Huntsville and adjacent areas now have a higher concentration of CI events compared to previous hours (where urban and cropland land types are prevalent). Another fairly concentrated cluster is present over northern Pike County, which has a mix of savanna and forest land types as well as no significant elevation features. The absence of abundant CI events along the Tennessee River and its immediate adjacent areas, particularly in northwest Alabama, is once again noted.

Continuing on to the next hour, Figures 4.5b, e (2000-2100 UTC) show similar clustering trends to the previous hour, including a rise in high-count locations across the domain (e.g., in Blount, Jefferson, and Shelby Counties). Jackson County still has the largest cluster in the entire domain, now with a greater number of high-count locations in the southwest part of the county just north of the Tennessee River Valley. Dekalb County also contains scattered high-count locations. Central Jefferson County, which includes the
Birmingham metro area, begins to emerge as a hotspot for higher-frequency events. To the south, Shelby County also exhibits more widespread high-count locations over much of its extent (not just near the enhanced terrain in the northern part of the county). The Piedmont region still has a small amount of CI event, not excluding the higher-elevation area in Clay County. More widespread high-count locations are now present in the southeast corner of the domain, where there are no significant elevation features. The grouping of CI events around the Huntsville area seen in the previous hour is less evident.

Over the course of the next hour (2100-2200 UTC), the diurnal heating begins to wane. As a result, the overall CI event count stopped increasing. Regardless, Figure 4.5c, f still show scattered high-count locations in Dekalb and Jackson Counties, along with increased clustering over Jefferson and northern Shelby Counties. The gradual amplification of this latter cluster throughout the late afternoon hours serves as a potential indicator of an additional underlying mechanism that takes effect in the later hours of the day, as it is not present during peak heating hours. Unlike previous hours, a small grouping of high CI event counts has formed over the area in and around Lake Guntersville (Marshall County). In Madison County a small cluster of events has occurred just north of the Tennessee River, which contrasts with earlier hours. This area has non-negligible elevation gradients. This cluster extends southward into eastern Morgan County. Dallas County now has a higher concentration of CI events, including around the Alabama River. Another major cluster at this time is seen over Autauga County, a stronger signal than previous hours. Lastly, the CI event cluster seen over northern Shelby County in previous hours is now absent.
Figure 4.5 As in Figure 4.4a-c, but for the middle day time group (1900-2200 UTC).
Figure 4.5 (continued) As in Figure 4.4d-f, but for the middle day time group (1900-2200 UTC).
4.2.3 Late Day Group

During the course of the last two hours of the daily sampling period when data was collected, a clear diminishing of the overall CI event count occurs across the domain. The CI event cluster in Jefferson County, in and around the Birmingham metro area, really begins to stand out in Figure 4.6a, c. To the south, scattered high-count locations continue to linger in Shelby and Chilton Counties. The major cluster in Jackson County is now diminished, replaced instead with scattered CI events primarily near higher elevation features. Despite shrunken in size, the smaller grouping in southern Madison County is still present. Here, a broader trend is realized from visual inspection. In the early time group, the southern part of the domain (below 32.5⁰N) has a lower number of CI events relative to regions to the north, where the significant elevation features are located. The spatial distribution between the two is more balanced in the middle time group and even favors the southern portion in the late time group. Another cluster of tallies shows up in the southwestern part of the domain over southern Hale and northern Marengo Counties at this time, over which the dominant land type is savanna.

One central feature is seen in Figure 4.6b, d, the last hour of the daily sampling period, induced by an underlying mechanism which was not initially regarded during the proposal of this study. This CI event cluster is likely associated with urban heat island (UHI) effects over Jefferson County from the Birmingham metropolitan area (the largest in Alabama in terms of both population and spatial extent). The focus area of higher CI event counts overlaps both the metro (denoted by the green urban land type) and higher elevation features in the east/southeast part of Jefferson County. Through the UHI effect,
air is heated through the release of sensible heat by a dense urban area with many structures, promoting localized lift and eventually convective clouds. Given the time of day here when the Jefferson County CI event cluster is present, this effect is certainly plausible. UHI effects are sometimes not the sole factor in local CI event enhancement over an area. Rather, a combination of factors have been shown before to induce CI along higher-elevation features as well as intensify existing convective echoes through localized expansion of the boundary layer mixing and localized convergence effects, such as that observed in the Beijing metropolitan area (Li et al. 2017a; Wang and Sun 2008). Thus, it is plausible that is the case here as well. A second cluster shows up in Chilton County, collocated with higher elevation (> 200 meters) and savanna land type.
Figure 4.6  As in Figure 4.4, but for the late time group (2200-0000 UTC).
4.3 **Cumulative CI Event Heat Maps**

The point-by-point spatial CI event distributions analyzed above have CI events summed at each grid point (at the 350 x 455 MRMS domain resolution) in hourly intervals over the 1600-0000 UTC daily sampling period. While these give a precise look into smaller-scale spatial discrepancies and temporal trends, a more holistic and comprehensive view of the spatial variations serves as a comparison tool between the three time groups. Figure 4.7 shows a set of “heat maps” of CI events for each time group with the grid resolution decreased by 75% (sums were calculated in ~4 x 4 km bins over the domain). In Figure 4.7a, several focus areas of higher event density appear for the early time group (27917 early CI events, 2326 events per 15-minute interval on average). The first spans from Bibb County up to western Jefferson County. Although no significant elevation features are present in this area, it does have a higher proportion of forest land type than surrounding areas (Figure 4.3). Another focus area makes up parts of the valley-ridge and Cumberland Plateau regions, extending from Shelby County up into Dekalb County. From Figure 4.2 it is clear that this higher event density swath roughly parallels a southwest-northeast path of higher elevation/elevation gradients. Eastern Dekalb County, which has some of the highest elevation in the state, also holds the densest cluster of events for the early time group across the entire domain. Lesser clustering is seen in the Piedmont region. A third focus area of higher event density is in far northwest Alabama, to the west and south of Florence/Muscle Shoals. No prevalent land type or large elevation features exist here, with the exception of a few minor spots of enhanced elevation gradient no higher than 120 meters over Colbert County. As identified before, the southern third of the domain lacks CI events relative to the northern
parts for this time group (4913 early CI events below 32.5°N, only 17.6% of all early domain CI events). To get a better sense of small-scale spatial trends around Huntsville and the Tennessee Valley area, a zoomed-in visualization of this area is also displayed in Figure 4.7a. This was done for all three of the time groups. Indeed, one strip of higher event density is seen in central Madison and eastern Limestone Counties, cutting west-east through Huntsville to near Monte Sano. Another area of higher density in northeastern Jackson County, where significant elevation gradients and heavy forest cover exist. A small cluster just south of Lake Wheeler, west of Decatur, can also be seen. Here, the dominant land type is cropland with no significant elevation features.

The middle time group consists of a vastly greater number of overall bin counts compared to the early time group in Figure 4.7b (50002 middle CI events, 4167 events per 15-minute interval on average). It should be addressed that a chunk of these CI events likely stemmed from outflow boundary interactions, rather than solely any of the potential mechanisms discussed in this study. In addition, the north-south event distribution is more balanced than the other two time groups (25808 and 24194 middle CI events above and below 33.25°N, respectively). The acceleration of pulse storm CI formation is well underway during the mid-afternoon hours here, one such focus area located over Jefferson, Shelby, and Blount Counties, at the southwest end of the Appalachian foothills. A notable trend from the early time group is that areas which stood out in terms of event density during the early group now exhibit lower counts relative to surrounding areas (e.g., the eastern Piedmont region, Bibb County, and Coosa/Elmore Counties). Atmospheric recovery time from previous convection likely had an influence in this spatial trend, with neighboring areas that had less early convection now
experiencing increased middle events. Western Jackson County has the highest-count cluster of the domain, where several bins have more than 20 CI event counts. This cluster composes much of Jackson County and southern Madison County. Out of 4241 middle CI events over the Tennessee Valley inset region in Figure 4.7, the cluster consists of 1530 of those events (making up 36.1% of all events inside this inset region, where the cluster is defined within the latitude/longitude bounds [34.5⁰N, 34.9⁰N], [−86.7⁰N, −86.0⁰N]). Another higher event density area is present down in the southeast corner of the domain, where weak clustering had already occurred in the early time group. A third area of higher event density in northwest Alabama, which also had relatively higher event densities in the early time group, has shifted eastward to overlay the Florence/Muscle Shoals area. Homing in on the Tennessee River Valley, the higher event density within Madison County shifts southward while the signal south of Lake Wheeler is not as evident. Marshall County has noticeably less CI events than adjacent counties and not much change from the early day group. Meanwhile, southern Limestone and eastern Lauderdale Counties (north of the Tennessee River) continue to have extremely low CI events relative to surrounding areas.

A significant drop in overall event density across the domain comes with the late time group as a result of the accompanying instability drop and increase in solar zenith angle, just as what was observed in Figure 4.6. This is seen in Figure 4.7c, with the stout drop in event counts evident over the entire domain (24469 late CI events, 3059 events per 15-minute interval on average). Compared to the early time group, a proportion of CI events nearly twice as high is found in the southern part of the domain (7660 late CI events below 32.5⁰N, making up 31.3% of all late domain CI events). The most notable
feature here is the high event density over Jefferson and Shelby County and extending down to Shelby County, including the Birmingham metropolitan area. This is again indicative of a possible UHI effect likely beginning to take hold here during the course of the early evening hours (988 late CI events in this cluster area within the latitude/longitude coordinate bounds [33.25°N, 33.75°N], [–87.20°E, –86.60°E], making up 4.04% of all late domain CI events). This is certainly a non-negligible fraction of all events, considering the size of the cluster relative to the domain. Although consideration of this domain fraction alone is not a strong indicator of event clustering, the higher spatial density of the cluster area relative to surrounding areas is very apparent. Higher event density is also seen to the south in Chilton County. Focusing in on the Tennessee River Valley, a marked decrease in CI events from the earlier timeframes has occurred in Jackson County. Two areas with very low event counts are Etowah/eastern Blount and Limestone Counties. Other than a small area of higher event density south of the Huntsville area (near the Tennessee River), counts are down over essentially the whole domain. This trend coincides with the decrease in the diurnal forcing mechanisms through the evening hours, which work to render the static features (elevation and land use) insignificant.
Figure 4.7 Spatial heat maps of CI events tallied in ~4 x 4 km bins for the early (1600-1900 UTC) (a), middle (1900-2200 UTC) (b), and late (2200-0000 UTC) (c) time groups over all case days. The region shown is the Alabama portion of the domain. Left plots show this region, while the right plots focus on the Tennessee River Valley (includes Huntsville). Color scale represents the total CI events per histogram bin (each of which contain multiple grid points) over all 36 case days in that specific time group.
4.4 CI Event Frequency Relationships of Static Features & Wind Direction

From the spatial distributions it is clear that non-random (spatial clustering of CI event tallies) features are indeed present over the domain across all three daily time groups: the valley-ridge region/Sand Mountain (early), southwestern Jackson County (middle), and the Birmingham area (late). In order to analyze the causality behind the locations of these main clusters, the most important mechanisms for each of these time periods are assessed. Each grid point in the domain has a unique combination of re-scaled static and non-static feature values for every 15-minute interval. The static features never change at the grid points throughout the study period, antecedent rainfall totals only change once per case day, and the RAP analysis features evolve at each sample interval (temporal resolution of 15 minutes). Given the favorable tranquil background environment and large sample size, if a specific feature is indeed more strongly correlated with CI events than other features then an interval grid point that has a higher value of that feature should have a raised probability of isolated CI events. To avoid the inherent bias resulting from higher sample sizes of certain bins (e.g., many grid points with savanna land type versus few with barren land type), a relative CI frequency was calculated on each bin with Equation (4.1):

\[
CI\% = \frac{CI \ bin \ count}{CI \ count \ across \ all \ bins}.
\]  

(4.1)
The discrete-interval bins were formed with appropriate bin sizes depending on the range of the feature sets. In the case of the two classification features (land use and wind direction), the individual classifications were used in place of bins.

4.4.1 Elevation and Elevation Gradient

As was found in various past studies, topography can be a source of orographic parcel ascent as long as at least a component of the prevailing wind is oriented upslope. If an area possesses pockets of higher elevations, an increased potential for orographic lift would exist in and around the higher elevation. Even though the elevation features over the study domain are small relative to other parts of the U.S. (mountain ranges in the West), they can still act as localized areas of orographic ascent through differential heating in weak flow regimes. This lift can aid parcels in reaching their LCL. In the early time group for the whole domain over all case days (Figure 4.8a), a steady upward trend in CI event frequency is indeed seen up to 500 meters, leveling off near 0.12% at that point and actually decreasing for the rest of the higher-elevation bins up to 700 meters (the extent of the rightmost bin). The fit takes the form of an upside down parabolic trend, with a majority of the variance between CI% and elevation explained (as seen in the $R^2$ value of 0.6114).

The upside down parabolic trend seen in the early time group is less apparent in Figure 4.8b, with no visible positive trend between CI% and elevation up to 350 meters. Past this point higher CI% is much more evident, peaking near 0.135% for the 550-600 meter bin. The 600-650 meter bin, perhaps surprisingly, has a considerably lower CI% than the surrounding bins (~0.07%). Although lower than the early time group, a sizable
correlation is still seen in the middle time group which depicts the leveling off of the upward trend over the higher-elevation bins (R² of 0.4531, almost half of the variance explained in this time group).

Contrary to both of the earlier time groups, a parabolic trend is seen in the late time group (Figure 4.8c). The CI% decreases with elevation up to 300 meters and increases exponentially past that point. The CI% peaks near 0.1% at the 600-650-meter bin, while the CI% of the highest elevation bin is significantly lower by about half of the peak bin (~0.05%). The percent of variance explained is the lowest of the three time groups, albeit still not completely insignificant (R² of 0.3545). This overall decrease in CI% versus elevation explained variance from the beginning to the end of the daily sampling period is supportive of the notion that the terrain mechanism is more important in CI event formation earlier in the day, consistent with Lima and Wilson (2008).

Not only is the highest elevation bin correlation with CI events seen in the early day group, but the discrepancy between extracted CI and non-CI samples over the entirety of the domain is also largest in the early group (Figure 4.8d). In the early group, the CI sample has an average elevation of 163.45 meters while the non-CI sample has a mean of 141.69 meters. The spread (standard deviation) of the early day group CI sample is also larger (93.72 meters versus 78.70 meters for non-CI). In the mid-day CI group, the difference in the means narrows (145.83 meters for CI versus 141.68 meters for non-CI) along with the difference in the spreads (83.87 meters for CI versus 78.53 meters for non-CI). The mean difference actually reverses in the late day group, with 138.13 meters in
the CI sample as opposed to 142.21 meters in the non-CI sample at similar spreads (77.71 meters versus 78.51 meters, respectively).

Figure 4.8 Binned elevation (x-axis) versus relative CI event frequency (CI%) of each bin (y-axis) with second-degree polynomial regression curves overlaid. The model fit equations defining the curves were formed using the median elevation values from each bin and the corresponding bin CI%. Fit correlation scores are displayed in the top left of each plot. Elevation is binned in 50-meter intervals. Time groups are ordered as early (a), middle (b), and late (c).
Elevation gradient is an indicator of terrain steepness at a point. Thus, it should show a similar or even stronger positive trend with isolated CI event formation as in Figure 4.8a. The early time group does show a steady positive trend across most of the bins, primarily up to 140 meters/pixel, in Figure 4.9a. A local maximum in CI% is seen at the 160-180 meter/pixel bin (slightly higher than 0.12%), while the 140-160 and 180-200 meter/pixel bins have lower CI% values than surrounding bins. The most notable feature here is the very significant increase in CI% at the second-highest elevation gradient bin (220-240 meters/pixel), spiking to just over 0.3%. This spike is the prime contributor to the non-negligible $R^2$ of this CI% distribution (0.1321). Despite this stand-out distribution feature, the highest bin (240-260 meters/pixel) has a CI% of 0% (no CI
events) over the duration of all case days. The reason for this is that only one domain grid point fit in the highest bin, where even one CI event strongly skews the percentage as opposed to a bin with many grid points. In future work, a region with more widespread higher elevation gradients can be isolated in order to attain a better bin count balance of this feature. Regardless, the overall positive trend here should not be ignored.

The steady net positive trend seen in the early time group also shows up for the middle time group in Figure 4.9b, although the 200-220 and 220-240 meters/pixel bins have considerably higher CI% than in the early group. Just as seen in the elevation bin distributions, another difference in the CI% distribution from the previous time group is the higher overall CI% values of the middle day group. This ties back to the greater number of overall CI events across the entire domain, causing an inflation of CI event frequencies over many grid point bin values. Once again, the second-highest bin has the highest CI% over all case days (~0.38%). Aside from the highest bin, the bin with the lowest CI% is 180-200 meters/pixel (~0.06%), running counter to the net positive trend in the other bins. The second and third-highest bins towering above the rest act as the primary driver in the slightly more significant explained variance compared to the early day group (0.1438).

In the late time group cumulative domain CI% values are down across all elevation gradient bins (Figure 4.9c), resulting from the overall drop in domain CI events compared to earlier time groups. Oscillation between 0.05-0.06% occurs in the lower bins up to 140 meters/pixel, a major factor in the very insignificant explained variance (0.0055). Like the middle group, the 180-200 meters/pixel bin has the lowest CI% of all
bins and the second/third-highest bins have the two highest CI% values (~0.115% and ~0.085% respectively, a vast drop from the middle group). However, the absence of a definable net positive trend in the late day group suggests, as with the elevation CI% distributions, a heightened importance of this feature in the earlier periods of the summertime diurnal convective cycle as opposed to later hours when other mechanisms emerge (e.g., outflow gust fronts).

Just as with elevation, the early day group shows the largest discrepancy in the means between the CI and non-CI samples (Figure 4.9d). In the early group, the mean of the CI sample is 13.37 meters/pixel while the non-CI sample mean is 11.38 meters/pixel. The spread of the early CI sample is also larger (16.77 meters/pixel versus 13.28 meters/pixel for non-CI). In the mid group, the difference in the means is smaller at 12.22 meters/pixel for the CI sample and 11.48 meters/pixel for the non-CI sample. The spreads of the CI sample here is still wider (15.56 meters/pixel versus 13.69 meters/pixel for non-CI). In the late day group, the means are nearly equal (11.55 meters/pixel for CI versus 11.54 meters/pixel for non-CI) with similar spreads as well (13.46 meters/pixel for CI versus 13.85 meters/pixel for non-CI). The temporal trend between samples seen here is consistent with the elevation distributions in Figure 4.8d.
Figure 4.9  As in Figure 4.8.a-c, but for elevation gradient (binned every 20 meters/pixel).
4.4.2 Land Use

Higher-density CI events matched fairly well with areas that contained scattered to widespread forestland [as also found in Gambill and Mecikalski (2011)] and secondarily savanna/cropland based on the spatial CI event distributions (e.g., Jackson County). These classifications should thus have the higher cumulative CI% than the other land types. Land classes that are completely absent from the re-scaled domain grid are excluded from the ensuing analysis. They include classes 3 (deciduous needleleaf forest), 6 (closed shrubs), 7 (open shrubs), and 15 (permanent snow/ice).

In the early time group, land class 9 (savanna) has the highest cumulative CI% (0.0475%) out of all classes (Figure 4.10a). This is followed by classes 4 (deciduous broadleaf forest), 13 (urban area), and 1 (evergreen needleleaf forest). Forest classes make up two of the top four classes in terms of CI%, somewhat consistent with the
enhanced likelihood of CCs over forestland found in Gambill and Mecikalski (2011). Urban area ranks as the top land class in the middle and late groups, coinciding with the prevalence of the UHI effect in the later heat maps as opposed to the earlier times (from Figure 4.7e). Land class 16 (barren) ranks the lowest, with a CI% at/near 0%. This class is neglected as it only takes up one grid point over the entire domain. Other lower-ranking classes include 2 (evergreen broadleaf forest) and 11 (wetlands). None of the land classes overly stand out from the rest as with the elevation and elevation gradient trends. The difference between the highest and lowest ranking classes is less than 0.02%, with a standard deviation of just 0.0111%.

As seen with elevation and elevation gradient, middle group CI% values are higher overall than the early time group due to the higher domain CI event count (Figure 4.10b). Land class 13 (urban area) is now the top-ranking class, higher than in the early time group. Following behind are classes 1 (evergreen needleleaf forest), 8 (woody savanna), and 4 (deciduous broadleaf forest). Here, forested land types compose three of the top four land classes for CI%. The lowest-ranking classes are 16 (barren), 11 (wetlands), and 17 (water bodies). The former two also ranked near the bottom for the early time group, whereas the water body class ranked higher for the early day group. Overall CI% discrepancies between classes is over twice as wide as the early group with a distribution standard deviation of 0.0205%.

In the late time group, CI% numbers drop back down from the middle-timeframe in Figure 4.10c (which was also seen with elevation and elevation gradient). Class 13 (urban area) still prevails as the top-ranking class, by an even wider margin than in the
two previous time groups (~0.09%). It drops in CI% by less than 0.01% from the middle time group, while the other classes experience sharper drops. This is an evident effect of the increasing UHI influence in the early evening hours (e.g., Jefferson County) as opposed to the less-significant CI event frequencies in the other land classes. The second to fourth-ranked land classes are identical to the middle day group. As for the lowest-ranking classes in the late time group, they are 16 (barren), 14 (cropland/natural vegetation), and 9 (savanna). The cropland/vegetation class ranking low for the late time group is contrary to the conclusions of Gambill and Mecikalski (2011), and the low ranking of the savanna class starkly contrasts with its standing in the early day group. With the former, the decrease in importance of evapotranspiration processes stemming from areas of denser vegetation in the later part of the diurnal cycle is a potentially key factor in its ranking decrease. The spread of the distribution is wider than the early group but narrower than the middle group, with a standard deviation of 0.0172%.
Figure 4.10  MODIS land class versus relative CI event frequency (CI%) of each class over all case days. Time groups are ordered as early (a), middle (b), and late (c).
4.4.3 Wind Direction

The low and upper-level wind patterns tend to have a more disorganized and random spatial and temporal nature on synoptically tranquil days than in the presence of a defined surface pressure system. If the wind has proper orientation in the low levels up a terrain feature, albeit light in magnitude, conditions can arise for orographic support in convective formation. Additionally, weak transport of higher moisture content from nearby areas (e.g., evapotranspiration from previous rainfall and/or dense vegetation) can occur. In the entire study sample, there is a complete absence of the northerly (N) wind direction category over the sampling period and is excluded from the ensuing analysis.

The four most prevalent wind directions in the early day group over all case days are ESE (14.67%), SE (14.59%), WSW (12.96%), and SSE (12.54%). A key commonality between these is a southerly component, with three of them also having an easterly component. The two least frequent wind directions are NNW and NNE, both of which have a dominant northerly component (< 0.01%). A standard deviation of 0.0184% points to a wider variation in CI% than what is seen in the early and late land use CI% distributions. In Figure 4.11a, the NW direction prevails as the top-ranking early CI% at ~0.07% and the second to fourth-ranked categories SSW, WNW, and SW respectively. All four of these directions share a similarity: a westerly component. The three lowest-ranking categories are NNE, NNW, and ESE. The former two are near or at 0% as they only make up a miniscule portion of the sample. Although ESE is the most prevalent direction, it has the third-lowest CI%.
The four most prevalent wind directions in the middle day group are SE (14.48%), SSE (14.32%), ESE (12.45%), and S (12.41%). The two least frequent wind directions are again NNW and NNE (< 0.01%). As with the static feature CI% distributions, this timeframe comes with an overall increase in CI% from the higher tally count over the domain. The spread is wider than the early group with a standard deviation of 0.0259%, a likely consequence of the higher CI event counts (which can increase spatial as well as the CI% distribution discrepancies). The top-ranking category for this time group is again NW at ~0.09%, followed by S, SSW, and W (Figure 4.11b). It is noted that out of these four, three have a westerly component. The E, ENE, and ESE directions all experience notable CI% increases relative to other directions. The three lowest-ranking categories are NNE, NNW, and NE. Only the latter changed from the early group, with the other two staying the same. All three lowest-ranked directions have a northerly component.

The four most prevalent wind directions in the late day group are SSE (14%), SE (13.67%), SSW (13.49%), and S (13.45%). The two least frequent wind directions, similar to the other two time groups, are NNE and NNW. Consistent with the static feature distributions, overall CI% numbers are down for all directions from the middle time group (Figure 4.11c). The late group has less variability than the middle group and slightly more variability than the early group, with a standard deviation of 0.019%. The top-ranking direction is now WNW near 0.065%, followed by SE, S, and ENE. All four of these categories are quite unique from one another on the wind direction scale, alluding to higher levels of spatial randomness in wind direction in relationship to more frequent CI events brought about by mechanisms such as outflow interaction. Of
particular note is the sizable decrease in ranking of the NW direction, which had the highest CI% for the early and middle day groups. The three lowest-ranking categories NNE, NNW, and NW, all of which have a northerly component. Alongside the other categorized study feature (land use), all three of the wind direction CI% distributions have lower peaks than what is seen in the elevation and elevation gradient distributions.

Figure 4.11  As in Figure 4.10, but for 10-meter wind direction. All direction categories are shown on the x-axis, with the northerly category (N) absent from the dataset altogether.
4.5 CI versus Non-CI Antecedent Rainfall & RAP Analysis Field Distributions

The RAP model field analyses and antecedent rainfall grid distributions evolve continuously over time (they do not remain static during the course of the study period). Cumulative CI versus non-CI event (1 versus 0) distributions are a way to assess differences between event and non-event instances. To effectively compare them through avoidance of quartile extent shrinking in the non-CI distribution, this strong imbalance in the dataset where the number of non-CI instances is far greater was realized. For each of the three separate time groups, a random sample of the non-CI events with the same sample size as the CI events was taken to account for this sample size imbalance and the resulting quartile range differences of the two samples (CI and non-CI). This way, a fair comparisons are enabled between both the distributions and time groups.

4.5.1 Antecedent Rainfall

Early, middle, and late CI versus non-CI distributions for one-day antecedent rainfall over the whole domain are shown in Figure 4.12. The lower quartile ranges are quite compressed for all samples relative to the higher quartiles due to the large number of sample points that fall into the lower quartiles. The larger quartile range for the early CI sample comes with a slightly higher mean (0.14 in. versus 0.12 in.) and median than the early non-CI sample, the latter of which is near zero for non-CI. Despite the lesser quartile extents, the standard deviation of the early non-CI sample (0.283 in.) is slightly higher than the early CI sample (0.274 in.). Compared to both sample distributions from the early day group, the CI/non-CI quartile extent differences of the middle samples are similar to the early distribution with elongated upper and middle quartile ranges. A higher
mean/median discrepancy is seen in the middle distribution relative to the early distribution (0.15 in. versus 0.12 in.). The one-day CI sample mean exceeds the non-CI sample in all three time groups, while the non-CI sample means reside in the upper quartile of their distributions (indicating sizable skewness from high-rainfall sample points). Additionally, the overall range of the CI sample is larger than the non-CI sample. In the middle day group, the standard deviation of the CI sample (0.303 in.) is slightly greater than the non-CI sample (0.288 in.). Both of these spreads are higher than their counterparts in the early time group. The trends from the previous time groups continue into the late time group, with elongated upper quartiles (large chunk of the data in the lower quartile) and a higher mean for the CI sample (0.16 in. versus 0.12 in.). This difference is the largest of the three time groups. The standard deviation of the late CI sample (0.331 in.) is greater than the late non-CI sample (0.286 in.), with the former being the largest of the three time groups and the latter near its middle counterpart. Since the antecedent rainfall data grid values only changes once for each case day, the increase in average one-day rainfall of the CI sample from the early to the late group signals CI events occurring at a greater number higher-rainfall points in the later hours.
Similar to the one-day distribution, in Figure 4.13 the two-day quartile range of the early CI sample is higher than the early non-CI sample with a stark difference between the mean distribution values (0.34 in. versus 0.27 in.). The early distribution also has a sharp rise in spreads relative to the early one-day distribution, with standard deviations of 0.543 in. for the CI sample and 0.51 in. for the non-CI sample (an effect of the higher rain totals). The CI/non-CI quartile extent differences are most alike between the middle and late distributions, and the difference between the middle CI/non-CI means is slightly higher than the early distribution (0.35 in. versus 0.27 in.). The standard deviation of the middle CI sample (0.559 in.) is greater than the non-CI sample (0.509 in.), a higher spread discrepancy than the early samples. Akin to the one-day data from the early and middle groups, the quartile ranges of the late CI sample are larger than the
non-CI sample. The confidence interval around the median is wider than the previous timeframe, but closer to that seen in the early time group. The CI mean in the late day group is still higher than the non-CI mean at a difference lesser than the other two time groups (0.33 in. versus 0.27 in.), contrary to the one-day data. The standard deviations of the late distribution are very similar, with the non-CI sample having a slight edge to the CI sample (0.518 in. versus 0.516 in.). Lower quartile ranges are not as compressed as with the one-day data, although still not as elongated as the upper quartile ranges. All means of the two-day samples also lie within the upper and lower quartile bounds (within the box), unlike the one-day data.

![Cumulative Two-Day Antecedent Rain Distributions Over Whole Domain](image)

**Figure 4.13** As in Figure 4.12, but for two-day antecedent rainfall.

As is seen in the one-day and two-day data, the overall range of the CI sample is larger than the non-CI sample for all three time groups in the five-day antecedent rainfall
data (Figure 4.14). In the early distribution, the mean of the CI sample is greater than that of the non-CI sample (0.85 in. versus 0.76 in.). As seen in the two-day middle distribution the standard deviation of the early CI sample (1.192 in.) is higher than the early non-CI sample (1.12 in.), supported by the longer quartile ranges. Discrepancy between the two middle samples is greater than between the early samples, evidenced by further separation between the middle CI and non-CI means (0.92 in. versus 0.76 in.). This is also accompanied by larger discrepancy between the standard deviations of the middle CI (1.288 in.) and non-CI (1.139 in.) samples as well, both of which are larger than their counterparts in the early distribution. The mean difference between the late CI and non-CI samples is nearly identical to that of the early distribution (0.85 in. versus 0.76 in.). The standard deviations of the late samples (1.173 in. for CI, 1.155 in. for non-CI) are closer than in the other time groups. Regardless of the decrease in contrast of the two samples compared to the middle distribution, the CI sample prevails as having higher antecedent rainfall on average relative to the non-CI sample for all three time groups (true across the one-day, two-day, and five-day antecedent rainfall distributions). In all but two of the antecedent rainfall time group distributions, the CI sample having a wider spread than its corresponding non-CI sample stands out as a main trend. A leading factor, as mentioned, is the higher rainfall bounds in the former sample. Although the CI sample mean is greater for all three time groups, there is no single time group that emerges for any of the three antecedent rainfall features in terms of CI versus non-CI mean difference.
4.5.2 Surface-Based CAPE

On average, times where CI events occurred at points had higher antecedent rainfall and evapotranspiration levels. Unlike the antecedent rainfall distributions, the means of the surface-based CAPE distributions in Figure 4.15 reside slightly near their respective medians. The variation of this instability feature over all case days can vary quite significantly, both spatially and temporally. This stems from the balanced nature of the distributions themselves, as opposed to the majority of sample points clustering in the bottom quartiles. The early time group has a pronounced discrepancy between the CI and non-CI means, even more so than the middle and late day groups (2694 J kg$^{-1}$ versus 2128 J kg$^{-1}$). The standard deviation of the early CI sample (690 J kg$^{-1}$) is smaller than the non-CI sample (901 J kg$^{-1}$), evidenced by the smaller lower bound in the latter. A key difference between the early and middle CI samples is that the upper quartile of the early
CI sample extends further than the non-CI sample, up to near 4500 J kg\(^{-1}\). The means of both middle samples are lower than their early counterparts, with a potential factor being scattered low-level evaporative cooling from existing convection (works to temporarily raise atmospheric stability). Regardless, the same CI/non-CI trend persists between the two middle samples (2336 J kg\(^{-1}\) versus 1963 J kg\(^{-1}\)). The standard deviation difference between the two is also lesser than the early time group where the spread of the non-CI sample still prevails (694 J kg\(^{-1}\) for CI versus 847 J kg\(^{-1}\) for non-CI), indicated by the convergence of the extents of the two distributions. In the late CI sample, the extent of the lower quartile is narrower than in the other two time groups while the upper quartile is slightly lower than the middle group. The means of both late samples retract further during this timeframe, which can be attributed to both existing convection and, more importantly, the waning of the diurnal heating cycle over the early evening hours. On average, the late CI sample still has higher surface-based CAPE than the non-CI sample at this time, the smallest discrepancy of the three time groups (2089 J kg\(^{-1}\) versus 1806 J kg\(^{-1}\)). The standard deviation of the non-CI sample is also still greater than the CI sample (885 J kg\(^{-1}\) versus 761 J kg\(^{-1}\)), with the difference between spreads the least of the time groups.
4.5.3  2-Meter Dewpoint Temperature

The dewpoint temperature can act as both an instability and moisture indicator. It is defined as the temperature to which the air needs to cool to in order to achieve saturation, with higher values signaling moister conditions and shorter time to attain saturation. Since dewpoint temperature and surface-based CAPE tend to be interrelated, it should come as no surprise that their CI versus non-CI trends are similar (Figure 4.16). Firstly, the early CI sample mean is higher than the early non-CI sample (23.2°C versus 22.1°C). As with the surface-based CAPE distribution, a narrower range of values is observed for the CI sample. With the range extent of the CI sample narrower than that of the non-CI sample, the standard deviation of the early non-CI sample (1.97°C) is noticeably higher than the early CI sample (1.34°C). Dewpoint distributions in the middle time group are similar to their respective counterparts from the early day group. The
difference between the middle CI and non-CI means is slightly higher (22.8°C versus 21.6°C), where these two mean values are slightly lower than the means from the early time group. Also similar to the early group, the standard deviation of the middle non-CI sample (2.07°C) is still greater than the corresponding middle CI sample (1.41°C). Both spreads are higher than the early group. The non-CI sample having a wider distribution and higher mean dewpoint continues into the late time group, with the mean of the late CI sample higher than the late non-CI sample (22.7°C versus 21.8°C). This difference is the least of the three time groups. In the late sample spreads, the standard deviation of the non-CI sample (2.12°C) is still greater than the CI sample (1.49°C). These spreads are the highest of the time groups. Wider extent of the non-CI samples in the dewpoint distributions is even more apparent than in the surface-based CAPE distributions. The higher mean antecedent rainfall of the CI samples could partially correlate with this trend (future analysis). The maximum mean differences of CAPE and dewpoint in the early time group suggests a heightened importance of atmospheric instability in this timeframe.

Figure 4.16  As in Figure 4.12, but for 2-meter dewpoint.
4.5.4 Surface-Based Lifted Index

As an instability indicator, a lower (more negative) surface-based lifted index (LI) implies more unstable low-level atmospheric conditions conducive to convection. This stability parameter is defined as the temperature difference between the 500 mb level and the surface: $LI = T_{500} - T_{sfc}$. The relationship with CI events is quite evident in the early time group (Figure 4.17), where the CI sample exhibits a more negative mean LI than the non-CI sample ($-7^\circ C$ versus $-5^\circ C$). Of note also in this time group is the non-CI sample having a further extent into higher LI values, closer to zero (indicative of higher stability). Due to this extent into higher LI values, the standard deviation of the early non-CI sample ($2.1^\circ C$) is greater than that of the early CI sample ($1.4^\circ C$). The middle day group also has clear dissimilarities between the CI and non-CI samples, with the mean of the former still lower (more unstable) than the latter ($-6^\circ C$ versus $-5^\circ C$). The difference in the middle group is less extreme than the early group. Overall extent of the LI in the middle non-CI sample is narrower than the early non-CI sample, resulting in a slightly lower standard deviation ($2.0^\circ C$) while still having higher spread than that of the middle CI sample ($1.2^\circ C$). As for the late time group, it shares more commonality with the middle group as opposed to the early day group with the quartile extents in the late CI sample more closely resembling that of the middle CI sample. The LI mean, along with the median, of the CI sample remains noticeably lower than the non-CI sample in the late group ($-6^\circ C$ versus $-5^\circ C$). Visually, the overall box spread of the late non-CI sample is similar to the CI sample as seen in the middle group. The standard deviations, however, tell a different story, with the late non-CI sample considerably higher ($2.0^\circ C$ versus
1.3°C). It is the smallest standard deviation difference of the three time groups, albeit not insignificant.

![Cumulative Lifted Index Distributions Over Whole Domain](image)

**Figure 4.17** As in Figure 4.12, but for surface-based LI.

### 4.5.5 Surface Temperature

In a tranquil, unstable environment characteristic of the selected case days in this study, a higher surface temperature ($T_s$) with maintained colder temperatures aloft can aid in raising the instability through heightened low-level lapse rates. Hence, a similar CI versus non-CI relationship to the previous RAP model variables should be present, even across the entirety of the study domain (higher average $T_s$ for the former). As seen in Figure 4.18, though, that is not the case. The extents of the two early samples are quite similar, with the extent of the CI sample being wider by a narrow margin. On average, the early CI sample is actually slightly cooler than the early non-CI sample (34.9°C versus 35.4°C) along with their respective medians. Influenced by more lower quartile
sample points, the mean lies slightly below the median in the early CI sample. The standard deviation of the early CI sample (3.7°C) is only slightly higher than the early non-CI sample (3.5°C), this difference being less than 50% than in the dewpoint distributions, indicating less spread of this feature relative to 2-meter dewpoint. The extent of the middle CI event distribution is still slightly larger than its middle non-CI counterpart, albeit not significantly. Compared to the early group, a wider discrepancy is seen between the two middle samples (33.5°C for CI versus 34.9°C for non-CI). Both of these mean Tₙ values are lower than their respective counterparts from the early group. Difference in the standard deviations of the samples is 0.1°C, indicating near-matching sample variabilities (3.8°C for CI versus 3.7°C for non-CI). Despite this commonality, there is a clear general shift of the CI sample towards cooler temperatures. Both late Tₙ samples trend cooler from the previous time groups, likely stemming from both the overall drop in diurnal heating and/or evaporative cooling from existing/previous convection during the late timeframe. The extents of the two late samples are quite similar, and the difference between the mean values more closely matches that of the early day group (30.3°C for CI versus 30.9°C for non-CI). Unlike the previous two time groups, the standard deviation of the late non-CI sample (3.2°C) is now slightly greater than the late CI sample (3.1°C). In all three time groups the CI sample is cooler on average than the non-CI sample, running contrary to the heightened buoyancy response from an increased Tₙ. A probable explanation for the actual trend here comes from the following concept: Over a point where CI is occurring at any given instance there exists opaque cloud cover from an intensifying convective cloud overhead that induces deceleration of surface solar heating, potentially working to even lower the surface
temperature. Based on this premise, the wider spreads of the CI samples in the early and middle day groups can be at least partially attributed to this phenomenon.

4.5.6 950 hPa Vertical Velocity

Vertical velocity at the 950 hPa level ($\omega_{950}$) is a direct indicator of low-level air ascent or descent on the meso and microscales. In the RAP analysis dataset, this variable is in units of Pascals per second (a vertical pressure rate of change with time as opposed to the other form in meters per second). In the vertical pressure coordinate system, negative values imply a net upward motion for a grid point at the 950 hPa level as the vertical pressure decreases with time and rises in the vertical direction. Inversely, a positive value implies net downward motion as the vertical pressure rate of change is positive. As opposed to the surface-based LI, which assumes a stability state in the surface-500 hPa vertical layer, this feature variable characterizes the instantaneous

Figure 4.18  As in Figure 4.12, but for surface temperature.
vertical motion at a single point. Figure 4.19 shows the CI versus non-CI box $\omega_{950}$ distributions across all three time groups. The extent of the early CI sample is slightly wider than the early non-CI sample, with all samples spanning both positive and negative values. On average, the early CI sample has a slightly more negative $\omega_{950}$ compared to the early non-CI sample ($-0.03$ Pa s$^{-1}$ versus $-0.02$ Pa s$^{-1}$). Going with its wider extent, the standard deviation of the early CI sample ($0.137$ Pa s$^{-1}$) is greater than the early non-CI sample ($0.116$ Pa s$^{-1}$). Compared to the early day group, the discrepancy between the middle distributions is even more evident. As a result, the difference between the middle means is significantly greater ($-0.090$ Pa s$^{-1}$ for CI versus $-0.010$ Pa s$^{-1}$ for non-CI).

Trends from the early group are unalike as well, with the non-CI sample mean increasing by $\sim 0.01$ Pa s$^{-1}$. The difference in spread between the middle standard deviations is approximately on par with the early group ($0.227$ Pa s$^{-1}$ for CI versus $0.205$ Pa s$^{-1}$ for non-CI), however both of these are nearly double their early counterparts. The CI versus non-CI gap widens further in the late day group, shown in their respective means ($-0.13$ Pa s$^{-1}$ versus $-0.01$ Pa s$^{-1}$). While the late CI mean drops by $\sim 0.04$ Pa s$^{-1}$ from the middle group, the late non-CI mean holds relatively steady. Increases in both of the standard deviations are seen as well in the late group ($0.303$ Pa s$^{-1}$ for CI versus $0.26$ Pa s$^{-1}$ for non-CI). The difference between these two is the largest of the three time groups. From these trends, it is clear that the CI samples typically have lower values (stronger upward motion) and a wider range compared to the non-CI samples. This coincides with the higher average surface-based CAPE for the CI samples as depicted in their proportionality from the following approximation between maximum updraft velocity ($w_{\text{max}}$) and CAPE, applied here to non-updraft velocities as well in Equation (4.2):
\[ w_{\text{max}} = \sqrt{2 \ast CAPE} \]  \hspace{1cm} (4.2)

On a broader scale more negative \( \omega_{950} \) can also imply a small contribution from synoptic-scale lift, aiding in the initial formation of convective cumulus (whereas positive values indicate greater subsidence). As for the general increase in sample spread/extent with time, two possible explanations arise: More extreme \( \omega_{950} \) fluctuations characteristic of the existing convection and/or more local variability in the low-level wind. The latter is examined next.

Figure 4.19  As in Figure 4.12, but for 950 hPa (mb) vertical velocity (\( \omega_{950} \)).
4.5.7 Wind Direction Standard Deviation

Wind variability can be deduced through two different vector properties: speed and direction. Even on the scale of a few kilometers, spatial variation in the wind direction can exist. A higher degree of horizontal wind variation can produce either localized areas of convergence or divergence, which in turn can raise the degree of vertical motion (upward or downward). Figure 4.20 shows CI versus non-CI wind direction standard deviation ($\Theta_{\text{std}}$) samples for all three time groups. In the early day group, the CI sample has a higher mean $\Theta_{\text{std}}$ (2.43°) than the non-CI sample (2.19°), as well as a higher spread/extent (standard deviation of 3.27° for CI versus 3.22° for non-CI). In the middle day group, the $\Theta_{\text{std}}$ means and distribution extents are higher for both CI and non-CI than their early counterparts. The CI sample here still has a higher mean than the non-CI sample (2.84° versus 2.56°), along with a higher standard deviation (3.45° for CI versus 3.34° for non-CI). The increase in average CI and non-CI $\Theta_{\text{std}}$ continues into the late day group, where the CI sample still possesses a higher mean $\Theta_{\text{std}}$ than the non-CI sample (2.90° versus 2.62°). As for the sample spreads, they hold fairly steady from the middle group (3.45° for CI versus 3.32° for non-CI). From these statistics and the sample distributions themselves, a trend emerges of samples with the CI events having more local wind direction variability on average than the corresponding non-CI samples. This connects with the typical light wind pattern present on the selected case days in this study, where higher variability (a characteristic of features such as differential heating and outflow boundaries) stems from a lack a low-level dynamic response to a nearby synoptic-scale system.
4.5.8 Wind Speed

Depending on the spatial geographic composition surrounding a location (e.g., elevation and land use), it is possible for a variation in near-surface wind speed to work in tandem with the geographic features. Neglecting the effects of precipitation drifting and evaporation, orographic precipitation intensity can indeed be sensitive to horizontal wind speed when upstream of an elevation feature (Kunz and Kottmeier 2006). In Figure 4.21, the calm surface wind pattern associated with the characteristic case day in this study is captured well with the vast majority of all time group samples < 10 knots. Since sample points from the entire domain are represented here, there lacks a stark distinction between CI and non-CI samples. In the early group the CI sample has a lower average 10-meter wind than the non-CI sample (1.7 knots versus 1.8 knots), along with a lower standard deviation (0.92 knots versus 0.97 knots). Shifting to the middle group, average
wind speeds increase (2.2 knots for CI versus 2.1 knots for non-CI) as well as the spreads (1.07 knots for CI versus 1.06 knots for non-CI). As opposed to the early group, the middle CI sample has a higher mean and extent than the non-CI sample. In the late day group, mean winds continue to increase relative to the early/middle day groups (2.4 knots for CI versus 2.2 knots for non-CI) along with the spread of the CI sample (1.15 knots for CI versus 1.06 knots for non-CI). The CI sample only exceeds the non-CI sample in the average/median 10-meter wind for the middle and late groups. One explanation is the more widespread nature of mature convection in the later afternoon hours, which can produce isolated instances of stronger surface winds (sustained and gusts). However, higher point wind speeds are indeed represented in the RAP model analysis despite interpolation. While a widening gap between the CI and non-CI samples is seen here from early to late, stronger feature trends may exist within a smaller area of higher CI event density. The increase in mean wind speeds with time here is a probable result of the deepening of the CBL throughout the day, as mixing of winds aloft toward the surface produces heightened turbulence.
4.6 Higher Event Density Subregion

Previously seen in Figure 4.7a, b, eastern Dekalb County is one area of high CI event density in the early day group. In the middle group, southwestern Jackson County in far northeast Alabama contains a very well-defined area of high-frequency points with additional higher-density areas in Dekalb and Madison Counties. To search for any stronger feature trends here, the data in this subregion (which includes the eastern edge of Huntsville proper) was isolated from the rest of the domain dataset for additional analysis (Figure 4.22). Cumulative CI events within the subregion are as follows: 2402 (early), 3343 (middle), 1117 (late). Respectively, the average amount of CI events per 15-minute interval over all case days are approximately 200 instances/interval, 279 instances/interval, and 140 instances/interval. This matches the spatial patterns on the
heat maps in Figure 4.7. The late day group is excluded from the ensuing subregion analysis, as there exists a lack of a definable tally cluster over the area in Figure 4.7c.

4.6.1 Subregion Elevation

Over the subregion, which contains many significant elevation features, the positive correlation in the early day group between elevation and CI% (~0.88) is stronger than over the entire spatial domain (Figure 4.23a). In particular, the 525-550 and 575-600
meter bins have the highest CI% out of all other early bins. As with Figure 4.8, the middle day group has a smaller elevation versus CI% correlation (~0.42) than the early group, possessing an upside-down parabolic nature rather than a near-linear fit relationship (Figure 4.23b). This middle-early difference is more significant than over the entire domain, although this could be partially due to inflation of middle CI% in the lower-elevation bins from a higher overall event count. Regardless, the starker trend in the early group here furtherly suggests that elevation is more important in the formation of pulse storm CI during the early hours as opposed to the two later time groups.

In Figure 4.8d, the early group also shows a noticeable difference in mean sample elevation between the CI and non-CI samples taken from the subregion despite its abundance of high elevation. In the early day group, the CI sample has a mean elevation of 341.3 meters whereas the mean elevation of the non-CI sample is 319.2 meters, a difference of over 20 meters. The early spreads are fairly even (102.2 meters for CI versus 99.0 meters for non-CI). In the middle group, the difference in the means is narrower just as with the samples taken over the entire domain (321.3 meters for CI versus 316.9 meters for non-CI). Like the early group, the spreads are still similar as well (99.6 meters for CI versus 98.1 meters for non-CI).
Figure 4.23  As in Figure 4.8, but only with elevation data at grid points over the subregion. Elevation is binned every 25 meters. Shown are the CI% distributions for the early (top left) and middle (top right) time groups, along with box plot distributions for these two groups (bottom).

4.6.2 Subregion Elevation Gradient

Over the larger domain, points with high elevation also tend to have higher elevation gradients. For grid points in the study domain with an elevation gradient $> 130$. 

107
meters/pixel, the mean elevation is 370.3 meters as opposed to a mean of 141.8 meters across all domain grid points. This feature correlation appears in Figure 4.8 and 4.9. One noted difference in the subregion bin distributions (Figure 4.24a, b) from the distributions for the entire domain (Figure 4.9a, b) is that the fit correlations are significantly lower over the subregion with correlations < 0.1. One contributor to this is the lack of counts in the 180-210 meter/pixel range, particularly in the early day group. For all four bin distributions, little upward trend is seen in the 0-200 meter/pixel range. Additionally, the CI% of the second-highest early elevation gradient bin is lower over the subregion than the whole domain, opposite of the middle day group. CI% numbers are higher overall in the middle group due to the greater event prevalence in that time group. The middle group here has the stronger relationship than the early group, contrary to the other distributions. While true that the highest bins have over a threefold higher CI% of the lower bins in both groups, the low correlations suggest rather insignificant relationships. With elevation, the bins have a more balanced distribution of grid points which ensures a positive trend is better captured.

With all previous analysis pointing to early significance of topography, a countertrend shows up between the subregion elevation gradient samples. In the early day group, the non-CI sample has a higher mean elevation gradient (36.1 meters/pixel) than the CI sample (33.2 meters/pixel). This could be a result of CI occurring on top of higher-elevation features (where the elevation gradient is minimal) as opposed to on a slope. The spread of the non-CI sample is also slightly greater (36.5 meters/pixel for non-CI versus 34.6 meters/pixel for CI). This trend is reversed in the middle group, where the CI sample mean (38.6 meters/pixel) is now higher than the non-CI sample (35.0 meters/pixel) along
with a reversal of the spread difference (37.8 meters/pixel for CI versus 36.1 meters/pixel for non-CI). All means here are greater than their counterparts in Figure 4.9d, consistent with the higher mean subdomain elevations (larger proportion of grid points representing significant elevation).
Figure 4.24  As in Figure 4.23, but for elevation gradient (binned every 10 meters).
4.6.3 Subregion Select RAP Model Features

As shown in the clear trends in samples from the entire domain between the CI and non-CI RAP model feature distributions, higher mean magnitudes of the CI samples from many of the features exist across all three time groups. The case for heightened importance of features such as surface-based CAPE and 2-meter dewpoint would be furthered if similar sample trends are present in the subregion. Indeed, a comparable subregion CI versus non-CI trend appears for the surface-based CAPE feature in Figure 4.25. In the early group, the CI sample has a higher mean (2838 J/kg) than the non-CI sample (2253 J/kg). This mean difference is very similar to the early sample difference in Figure 4.15. The non-CI sample, however, has a larger spread (909 J/kg versus 603 J/kg). For the middle group, the difference in the means is lesser (2446 J/kg for CI versus 2075 J/kg for non-CI), very alike in magnitude to the middle sample difference in Figure 4.15. The non-CI sample still has a greater standard deviation (734 J/kg versus 510 J/kg). As in Figure 4.15, the CI sample has a higher mean surface-based CAPE along with a smaller spread for all time groups. All subregion sample means here are also higher than their entire-domain sample counterparts.
Dewpoint temperature and other instability features (e.g., surface-based CAPE) can go hand-in-hand when it comes to atmospheric convective support. This premise is upheld by the results in Figure 4.15, 4.16, and 4.25. Thus, it should come as no surprise that similar CI versus non-CI trends show up between the subregion dewpoint samples in Figure 4.26. In the early group, the CI sample (22.4°C) has a higher mean than the non-CI sample (21.6°C) while the non-CI sample has the greater spread, as with surface-based CAPE (1.8°C versus 1.2°C). The early mean difference here is less than the difference seen in Figure 4.16. A continuation of this trend translates to the middle day group, with the CI sample having a mean of 22.0°C as opposed to the non-CI sample (21.3°C). The difference between the spreads is less than the early group (1.7°C for non-CI versus 1.2°C for CI). The subregion mean dewpoint differences are smaller than in Figure 4.16 (unlike with surface-based CAPE).
In Figure 4.19, the CI samples in all three time groups had a more negative sample mean $\omega_{950}$ compared to the non-CI samples (i.e., stronger vertical motion). This trend translates to the subregion as well (Figure 4.27). In the early day group, the mean of the CI sample ($-0.056 \text{ Pa s}^{-1}$) is nearly double that of the non-CI sample ($-0.027 \text{ Pa s}^{-1}$). Note how both of these means are more negative than their counterparts in Figure 4.19. With the numerous elevation features present in this subregion, these mean increases potentially point to the role of orographic lift in early event formation (as also deduced from Figure 4.23a). The spreads are similar ($0.080 \text{ Pa s}^{-1}$ for CI versus $0.086 \text{ Pa s}^{-1}$ for non-CI). In the middle day group, the mean $\omega_{950}$ difference is considerably wider ($-0.090 \text{ Pa s}^{-1}$ for CI versus $-0.025 \text{ Pa s}^{-1}$ for non-CI) while the spreads are similar as in the early group ($0.128 \text{ Pa s}^{-1}$ for CI versus $0.120 \text{ Pa s}^{-1}$ for non-CI). Here, only the middle non-CI sample has a notable difference in the mean $\omega_{950}$ from its corresponding sample in Figure 4.19.

Figure 4.26  As in Figure 4.25, but for 2-meter dewpoint.
Through the implemented exhaustive grid search, 36 different iterations were produced over a hyperparameter “space” (3 validation folds * 4 random forest sizes * 3 maximum tree depths). Out of all hyperparameter combinations, it was found that the highest max depth/estimator hyperparameter pair in the specified range (20 nodes and 200 decision trees) had the highest validation score within the specified parameter space (0.9916). The hyperparameter combinations with a maximum tree depth of 10 were considerably lower than the other combinations (~0.77) while the combinations with a maximum tree depth of 15 scored considerably higher (~0.94). It was noted during this process that the validation scores were much more sensitive to changes in the maximum tree depth as opposed to changes in the random forest size (number of trees). Due to the performance of the highest hyperparameter combination, it was used in the ensuing analysis for all model runs. Since only the importance rankings are desired in this study, 100% of the available data was used for training in their respective time groups.

Figure 4.27  As in Figure 4.25, but for $\omega_{950}$.

4.7  Random Forest Importance


When incorporated into a random forest classification model, each feature (or predictor) has a unique degree of importance in the determination of the class labels of that particular model. Instead of a spatial analysis, CI frequency distribution, or a CI versus non-CI feature sample comparison, all predictors here are utilized into training of the model with a constant set of hyperparameters.

4.7.1 Feature Importance over Whole Domain

A higher Gini importance value (a unitless scale ranged from 0-1) in a feature implies higher average node purity and thus an evener feature split during the class labeling. Due to the bias of the Gini importance toward high cardinality values (e.g., RAP model analysis fields), separate categorical and non-categorical feature comparisons are necessary for the Gini importance metric. Tables 3-5 show the feature importance values of all three time groups using model runs with training data from all 36 case days over the entire grid domain. The model random state was perturbed four times, along with two additional runs with the three most important/two least important features from the four random states excluded and a run that calculated the permutation importance (which reduces the bias towards high cardinality features) at a singular random state. For the latter, one feature from each of the five hierarchal clusters in Figure 3.1 was selected to incorporate into the model in order to avoid collinearity between similar features. Across all three time groups, the highest-ranking features share some commonalities.

By order of ranking in the early time group (Table 4.1), the three highest-ranking Gini importance features across all four bootstrapping states are surface-based CAPE, 2-meter dewpoint, and surface temperature. Here it is noted that dewpoint is primarily a
moisture indicator while the other two are instability indicators. Although surface-based LI is also an instability indicator, it ranks lower as a result of its limited scale range compared to the top three features. Looking at the two topography features, elevation consistently ranks nearly double in importance magnitude than elevation gradient. The two least important features, in order of lowest Gini importance, are land use and one-day antecedent rainfall across all four random states. With land use, the main driver behind its bottom ranking is likely its limited categorical scale range. In the run excluding the top and bottom-ranking features (set to a random bootstrap state of 42 in all three time groups), the three top-ranking features are LI, 10-meter wind speed, and elevation. The two bottom-ranking features are 10-meter wind direction and elevation gradient. From this run, the top-ranking features are a mix of indicator types including instability, wind, and topography (static). Out of the five selected features for the early permutation run, the top two early features are one-day antecedent rainfall and $\Theta_{std}$. This contrasts with the early Gini runs, where these two features are ranked in the bottom and middle tiers respectively.
In the middle time group, the Gini importance rankings do not stray greatly from the early group (Table 4.2). One difference does appear; the third and fourth rankings are swapped ($\omega_{950}$ and surface temperature). Again, all of the top-three features are some form of an instability/buoyancy indicator. The bottom two rankings continue to be held by land use and one-day antecedent rainfall. For the antecedent rainfall features, one-day

<table>
<thead>
<tr>
<th>Model Run:</th>
<th>Random State = 42 (Gini)</th>
<th>Random State = 43 (Gini)</th>
<th>Random State = 44 (Gini)</th>
<th>Random State = 45 (Gini)</th>
<th>Exclude (Gini)</th>
<th>Random State = 42 (Perm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-Day Antecedent Rain</td>
<td>.0324</td>
<td>.0330</td>
<td>.0326</td>
<td>.0328</td>
<td>.00114</td>
<td>.00114</td>
</tr>
<tr>
<td>Two-Day Antecedent Rain</td>
<td>.0501</td>
<td>.0489</td>
<td>.0504</td>
<td>.0491</td>
<td>.0848</td>
<td>.0848</td>
</tr>
<tr>
<td>Five-Day Antecedent Rain</td>
<td>.0622</td>
<td>.0615</td>
<td>.0622</td>
<td>.0616</td>
<td>.0996</td>
<td>.0996</td>
</tr>
<tr>
<td>Wind Direction</td>
<td>.0450</td>
<td>.0450</td>
<td>.0453</td>
<td>.0449</td>
<td>.0649</td>
<td>-.00445</td>
</tr>
<tr>
<td>$\Theta_{std}$</td>
<td>.0683</td>
<td>.0686</td>
<td>.0685</td>
<td>.0681</td>
<td>.113</td>
<td>.000202</td>
</tr>
<tr>
<td>Wind Speed</td>
<td>.0886</td>
<td>.0882</td>
<td>.0892</td>
<td>.0888</td>
<td>.141</td>
<td>.141</td>
</tr>
<tr>
<td>Dewpoint</td>
<td>.126</td>
<td>.124</td>
<td>.132</td>
<td>.123</td>
<td>-.000849</td>
<td>.175</td>
</tr>
<tr>
<td>Sfc-Based LI</td>
<td>.0731</td>
<td>.0779</td>
<td>.0748</td>
<td>.0806</td>
<td>.175</td>
<td>.175</td>
</tr>
<tr>
<td>Sfc-Based CAPE</td>
<td>.151</td>
<td>.150</td>
<td>.143</td>
<td>.149</td>
<td>.149</td>
<td>.149</td>
</tr>
<tr>
<td>Sfc. Temp.</td>
<td>.0906</td>
<td>.0909</td>
<td>.0906</td>
<td>.0911</td>
<td>.0911</td>
<td>.0911</td>
</tr>
<tr>
<td>$\omega_{950}$</td>
<td>.0733</td>
<td>.0726</td>
<td>.0723</td>
<td>.0728</td>
<td>.110</td>
<td>.110</td>
</tr>
<tr>
<td>Elevation</td>
<td>.0806</td>
<td>.0813</td>
<td>.0805</td>
<td>.0805</td>
<td>.130</td>
<td>-.00765</td>
</tr>
<tr>
<td>Elevation Gradient</td>
<td>.0428</td>
<td>.0427</td>
<td>.0428</td>
<td>.0424</td>
<td>.0821</td>
<td>.0821</td>
</tr>
<tr>
<td>Land Use</td>
<td>.0154</td>
<td>.0154</td>
<td>.0153</td>
<td>.0153</td>
<td>.0153</td>
<td>.0153</td>
</tr>
</tbody>
</table>
and two-day antecedent rainfall importance values are lower than the early group, however the five-day antecedent rainfall importance values are higher. Gini importance of elevation is also lower than in the early group, supporting the notion that this feature is a more dominant mechanism earlier on in the day. In the middle run excluding the top and bottom-ranking features from the left four runs, the three most important features are surface-based LI, surface temperature, and 10-meter wind speed (two of which are instability indicators) while the bottom two are 10-meter wind direction and two-day antecedent rainfall. Compared to the exclusion run of the early day group, two of the top three features remain the same and only the bottom-ranking feature in the excluded run remains the same. Out of the five features included in the middle day permutation run, the top two middle day features are one-day antecedent rainfall (as in the early day group) and 10-meter wind direction. Elevation ranks in the bottom two for both the early and middle day groups, contrary to its strong early spatial and statistical correlations in this study.
Table 4.2  As in Table 4.1, but for the middle day group over the entire domain.

<table>
<thead>
<tr>
<th>Model Run:</th>
<th>Random State = 42 (Gini)</th>
<th>Random State = 43 (Gini)</th>
<th>Random State = 44 (Gini)</th>
<th>Random State = 45 (Gini)</th>
<th>Exclude (Gini)</th>
<th>Random State = 42 (Perm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-Day Antecedent Rain</td>
<td>.0304</td>
<td>.0304</td>
<td>.0301</td>
<td>.0306</td>
<td>.000558</td>
<td></td>
</tr>
<tr>
<td>Two-Day Antecedent Rain</td>
<td>.0442</td>
<td>.0448</td>
<td>.0462</td>
<td>.0445</td>
<td>.0728</td>
<td></td>
</tr>
<tr>
<td>Five-Day Antecedent Rain</td>
<td>.0648</td>
<td>.0639</td>
<td>.0640</td>
<td>.0640</td>
<td>.0977</td>
<td></td>
</tr>
<tr>
<td>Wind Direction</td>
<td>.0390</td>
<td>.0386</td>
<td>.0386</td>
<td>.0387</td>
<td>.0588</td>
<td>-.00188</td>
</tr>
<tr>
<td>$\Theta_{\text{std}}$</td>
<td>.0645</td>
<td>.0644</td>
<td>.0643</td>
<td>.0646</td>
<td>.108</td>
<td>-.00215</td>
</tr>
<tr>
<td>Wind Speed</td>
<td>.0810</td>
<td>.0814</td>
<td>.0813</td>
<td>.0813</td>
<td>.130</td>
<td></td>
</tr>
<tr>
<td>Dewpoint</td>
<td>.140</td>
<td>.139</td>
<td>.137</td>
<td>.135</td>
<td>-.0111</td>
<td></td>
</tr>
<tr>
<td>Sfc-Based LI</td>
<td>.0731</td>
<td>.0727</td>
<td>.0731</td>
<td>.0794</td>
<td>.163</td>
<td></td>
</tr>
<tr>
<td>Sfc-Based CAPE</td>
<td>.120</td>
<td>.122</td>
<td>.121</td>
<td>.118</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sfc. Temp.</td>
<td>.101</td>
<td>.102</td>
<td>.102</td>
<td>.103</td>
<td>.161</td>
<td></td>
</tr>
<tr>
<td>$\omega_{950}$</td>
<td>.107</td>
<td>.107</td>
<td>.107</td>
<td>.107</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elevation</td>
<td>.0728</td>
<td>.0728</td>
<td>.0728</td>
<td>.0728</td>
<td>.123</td>
<td>-.00564</td>
</tr>
<tr>
<td>Elevation Gradient</td>
<td>.0456</td>
<td>.0456</td>
<td>.0457</td>
<td>.0457</td>
<td>.0853</td>
<td></td>
</tr>
<tr>
<td>Land Use</td>
<td>.0163</td>
<td>.0162</td>
<td>.0163</td>
<td>.0162</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Random forest results over the entire domain for the late day group are shown in Table 4.3. In order of Gini importance, the top three features over all four bootstrapping states are 2-meter dewpoint, $\omega_{950}$, and surface-based CAPE. These are the same top-ranking features as the middle day group, only now $\omega_{950}$ has ascended to second-most important. One notable trend is a marked decrease in importance of the surface-based LI (an instability indicator) from the earlier time groups. Elevation is more important than in the middle group but less important than the early day group, while elevation gradient
importance is actually higher than in both early day and middle day groups. As with the other two time groups, the two lowest-ranking features are land use and antecedent rainfall. For the late day run with the three most important and two least important features from the other late runs excluded, the three most important features are wind speed, elevation, and surface-based LI. The two least important features here are 10-meter wind direction and two-day antecedent rainfall. Out of the five features included in the late permutation run, the top two late features are wind direction and $\Theta_{\text{std}}$. The most notable trend here from the early and middle day groups is the decreased importance of one-day antecedent rainfall, and the top two being wind direction features suggests an increased significance of thunderstorm outflow for this later time period. It also runs counter to the trend in Figure 4.12, where the mean discrepancy in one-day antecedent rainfall is largest in the late day group. Moisture (2-meter dewpoint) is also less important than in the earlier time groups.
Table 4.3  As in Table 4.1, but for the late day group over the entire domain.

<table>
<thead>
<tr>
<th>Model Run:</th>
<th>Random State = 42 (Gini)</th>
<th>Random State = 43 (Gini)</th>
<th>Random State = 44 (Gini)</th>
<th>Random State = 45 (Gini)</th>
<th>Exclude (Gini)</th>
<th>Random State = 42 (Perm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-Day Antecedent Rain</td>
<td>.0341</td>
<td>.0342</td>
<td>.0343</td>
<td>.0349</td>
<td>-.00108</td>
<td></td>
</tr>
<tr>
<td>Two-Day Antecedent Rain</td>
<td>.0468</td>
<td>.0461</td>
<td>.0464</td>
<td>.0457</td>
<td>.0745</td>
<td></td>
</tr>
<tr>
<td>Five-Day Antecedent Rain</td>
<td>.0695</td>
<td>.0690</td>
<td>.0692</td>
<td>.0691</td>
<td>.106</td>
<td></td>
</tr>
<tr>
<td>Wind Direction</td>
<td>.0404</td>
<td>.0407</td>
<td>.0402</td>
<td>.0404</td>
<td>.0635</td>
<td>.000738</td>
</tr>
<tr>
<td>$\Theta_{std}$</td>
<td>.0716</td>
<td>.0711</td>
<td>.0714</td>
<td>.0718</td>
<td>.119</td>
<td>.000617</td>
</tr>
<tr>
<td>Wind Speed</td>
<td>.0925</td>
<td>.0921</td>
<td>.0926</td>
<td>.0922</td>
<td>.143</td>
<td></td>
</tr>
<tr>
<td>Dewpoint</td>
<td>.122</td>
<td>.120</td>
<td>.122</td>
<td>.119</td>
<td>-.00840</td>
<td></td>
</tr>
<tr>
<td>Sfc-Based LI</td>
<td>.0546</td>
<td>.0589</td>
<td>.0572</td>
<td>.0613</td>
<td>.124</td>
<td></td>
</tr>
<tr>
<td>Sfc-Based CAPE</td>
<td>.114</td>
<td>.113</td>
<td>.112</td>
<td>.112</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sfc. Temp.</td>
<td>.0938</td>
<td>.0939</td>
<td>.0930</td>
<td>.0932</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\omega_{950}$</td>
<td>.118</td>
<td>.117</td>
<td>.118</td>
<td>.118</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elevation</td>
<td>.0779</td>
<td>.0779</td>
<td>.0780</td>
<td>.0775</td>
<td>.132</td>
<td>-.00138</td>
</tr>
<tr>
<td>Elevation Gradient</td>
<td>.0477</td>
<td>.0481</td>
<td>.0477</td>
<td>.0479</td>
<td>.0887</td>
<td></td>
</tr>
<tr>
<td>Land Use</td>
<td>.0176</td>
<td>.0177</td>
<td>.0176</td>
<td>.0176</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.7.2  Feature Importance Over Subdomain

Random forest importance results over the subdomain of higher event density (see Figure 4.22) are displayed for the early and middle day groups in Tables 6-7. In the early time group (Table 4.4), the three most important (Gini) features across all four bootstrapping states are surface-based CAPE, 10-meter wind speed, and 2-meter
dewpoint. Compared to the data in Table 4.1, elevation has lower model importance and elevation gradient has a higher importance. The high Gini ranking of wind speed here suggests a potential significance of orographic flow over the subregion. The bottom two features remain as land use and one-day antecedent rainfall, consistent with the whole-domain model runs. In the run with the three most important/two least important features excluded, the top three features are surface-based LI, surface temperature, and $\Theta_{\text{std}}$. The bottom two features are 10-meter wind direction and two-day antecedent rainfall. Out of the five features included in the permutation run, the top two early features are one-day antecedent rainfall and 10-meter wind direction. Here, only the latter differs from the permutation rankings in Table 4.1 (albeit still a wind direction indicator).
Table 4.4  As in Table 4.1, but for the early day group over the subregion.

<table>
<thead>
<tr>
<th>Model Run:</th>
<th>Random State = 42 (Gini)</th>
<th>Random State = 43 (Gini)</th>
<th>Random State = 44 (Gini)</th>
<th>Random State = 45 (Gini)</th>
<th>Exclude (Gini)</th>
<th>Random State = 42 (Perm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-Day Antecedent Rain</td>
<td>.0339</td>
<td>.0339</td>
<td>.0336</td>
<td>.0336</td>
<td></td>
<td>-.000815</td>
</tr>
<tr>
<td>2-Day Antecedent Rain</td>
<td>.0435</td>
<td>.0428</td>
<td>.0430</td>
<td>.0428</td>
<td>.0751</td>
<td></td>
</tr>
<tr>
<td>5-Day Antecedent Rain</td>
<td>.0577</td>
<td>.0596</td>
<td>.0595</td>
<td>.0584</td>
<td>.0969</td>
<td></td>
</tr>
<tr>
<td>Wind Direction</td>
<td>.0460</td>
<td>.0460</td>
<td>.0459</td>
<td>.0462</td>
<td>.0712</td>
<td>-.00111</td>
</tr>
<tr>
<td>Wind Dir. S.T.D.</td>
<td>.0736</td>
<td>.0748</td>
<td>.0753</td>
<td>.0728</td>
<td>.133</td>
<td>-.00177</td>
</tr>
<tr>
<td>Wind Speed</td>
<td>.107</td>
<td>.104</td>
<td>.106</td>
<td>.107</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dewpoint</td>
<td>.104</td>
<td>.104</td>
<td>.107</td>
<td>.103</td>
<td>-.00169</td>
<td></td>
</tr>
<tr>
<td>Lifted Index</td>
<td>.0713</td>
<td>.0785</td>
<td>.0789</td>
<td>.0820</td>
<td>.156</td>
<td></td>
</tr>
<tr>
<td>Sfc-Based CAPE</td>
<td>.168</td>
<td>.164</td>
<td>.158</td>
<td>.160</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sfc. Temp.</td>
<td>.0892</td>
<td>.0879</td>
<td>.0886</td>
<td>.0876</td>
<td>.155</td>
<td></td>
</tr>
<tr>
<td>ω950</td>
<td>.0752</td>
<td>.0748</td>
<td>.0737</td>
<td>.0758</td>
<td>.111</td>
<td></td>
</tr>
<tr>
<td>Elevation</td>
<td>.0687</td>
<td>.0693</td>
<td>.0697</td>
<td>.0697</td>
<td>.113</td>
<td>-.00140</td>
</tr>
<tr>
<td>Elevation Gradient</td>
<td>.0472</td>
<td>.0463</td>
<td>.0455</td>
<td>.0461</td>
<td>.0897</td>
<td></td>
</tr>
<tr>
<td>Land Use</td>
<td>.0151</td>
<td>.0147</td>
<td>.0150</td>
<td>.0150</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the middle time group subregion runs of four bootstrapping states (Table 4.5), the top three Gini features are surface-based CAPE, 2-meter dewpoint, and ω950 while the bottom two are still land use and one-day antecedent rainfall. ω950 still ranks higher in Gini importance than in the early day group. The bottom two features continue to hold constant (land use and one-day antecedent rainfall). In the run with the three most
important/two least important features excluded, the top three are surface-based LI, surface temperature, and 10-meter wind speed (the former two are instability indicators) with the bottom two as 10-meter wind direction and two-day antecedent rainfall. As in Table 4.4 for the early day group, the top two features in the middle permutation run are one-day antecedent rainfall and 10-meter wind direction. Also of note, elevation ranks higher over these two subregion groups (third) than over the entire domain.

Table 4.5  As in Table 4.1, but for the middle day group over the subregion.

<table>
<thead>
<tr>
<th>Model Run:</th>
<th>Random State = 42 (Gini)</th>
<th>Random State = 43 (Gini)</th>
<th>Random State = 44 (Gini)</th>
<th>Random State = 45 (Gini)</th>
<th>Exclude (Gini)</th>
<th>Random State = 42 (Perm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-Day Antecedent Rain</td>
<td>.0365</td>
<td>.0362</td>
<td>.0353</td>
<td>.0357</td>
<td>-.000824</td>
<td></td>
</tr>
<tr>
<td>2-Day Antecedent Rain</td>
<td>.0427</td>
<td>.0443</td>
<td>.0437</td>
<td>.0440</td>
<td>.0735</td>
<td></td>
</tr>
<tr>
<td>5-Day Antecedent Rain</td>
<td>.0598</td>
<td>.0593</td>
<td>.0610</td>
<td>.0605</td>
<td>.0984</td>
<td></td>
</tr>
<tr>
<td>Wind Direction</td>
<td>.0396</td>
<td>.0400</td>
<td>.0394</td>
<td>.0387</td>
<td>.0598</td>
<td>-.00132</td>
</tr>
<tr>
<td>Wind Dir. S.T.D.</td>
<td>.0712</td>
<td>.0708</td>
<td>.0706</td>
<td>.0707</td>
<td>.120</td>
<td>-.00195</td>
</tr>
<tr>
<td>Wind Speed</td>
<td>.0881</td>
<td>.0869</td>
<td>.0879</td>
<td>.0884</td>
<td>.144</td>
<td></td>
</tr>
<tr>
<td>Dewpoint</td>
<td>.114</td>
<td>.114</td>
<td>.120</td>
<td>.115</td>
<td>-.00207</td>
<td></td>
</tr>
<tr>
<td>Lifted Index</td>
<td>.0743</td>
<td>.0772</td>
<td>.0806</td>
<td>.0780</td>
<td>.165</td>
<td></td>
</tr>
<tr>
<td>Sfc-Based CAPE</td>
<td>.151</td>
<td>.151</td>
<td>.141</td>
<td>.148</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sfc. Temp.</td>
<td>.0899</td>
<td>.0882</td>
<td>.0892</td>
<td>.0902</td>
<td>.136</td>
<td></td>
</tr>
<tr>
<td>0990</td>
<td>.108</td>
<td>.107</td>
<td>.107</td>
<td>.107</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elevation</td>
<td>.0609</td>
<td>.0618</td>
<td>.0605</td>
<td>.0608</td>
<td>.108</td>
<td>-.00156</td>
</tr>
<tr>
<td>Elevation Gradient</td>
<td>.0498</td>
<td>.0496</td>
<td>.0499</td>
<td>.0495</td>
<td>.0936</td>
<td></td>
</tr>
<tr>
<td>Land Use</td>
<td>.0142</td>
<td>.0142</td>
<td>.0140</td>
<td>.0142</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 5. DISCUSSION

In the typical synoptic environment on the case days selected for this study, several key characteristics stand out. For one, widespread surface dewpoint temperatures > 65°F (18°C) are common over the region. There also tends to be minimal low-level (< 5 knots) and 500 hPa (< 15 knots) wind speeds leading to suppressed amounts of wind shear. At the surface, one strong indicator of synoptic-scale subsidence is the presence of high pressure in the vicinity of the study domain. A complement to these conditions is a weak 500 hPa geopotential height gradient over the region, coinciding with the lack of a well-defined synoptic-scale horizontal temperature gradient. In the thermodynamic profile, a well-mixed CBL is evident up to near the LCL (usually between 800-900 hPa in a Southeast U.S. summertime environment) with drier mid-levels further aloft. This is a characteristic thermodynamic profile on days conducive to severe weather (multicell and supercell storm modes), excluding the dynamic support necessary to sustain such storms.

The typical background environment on the collected case days is also similar to that observed in previous studies (e.g., Brown and Arnold 1998; Gambill and Mecikalski 2011).

In the spatial CI event distributions, various features appear over the three time groups. There is a disproportionality of land classes and elevation in the re-scaled grid, with the savanna and broadleaf forest land classes together making up 78.8% of all domain grid points. In addition, flatter elevation areas are far more abundant than higher-terrain features, which are mainly found in northeast Alabama. Given these feature distribution imbalances, the need arises for the analysis of relative CI% frequencies in
assessing relationship with CI of the selected features. In northeast Alabama, the Cumberland Plateau and valley-ridge regions (home to numerous higher-elevation features) exhibit the highest concentration of CI events over the early and middle day groups. The highest overall CI event counts are found in the middle day group (50002), while the lowest overall counts occur in the late group (24469). Smaller areas of higher CI event concentration also show up in each of the time groups. For instance, the Piedmont region and Bibb County have higher event concentrations in the early group relative to surrounding areas with reduced event counts later in the day. Jefferson County, which includes the Birmingham metropolitan area, shows a localized CI event concentration in the late day group. This is suggestive of two local mechanisms: urban heat island and topography in the southeast part of the county.

Correlation bar plots of the static (elevation and land use) and other categorical (wind direction) data were then assessed. A prominent early CI event signal exists over the study domain with respect to both the elevation and elevation gradient features, pointing to the importance of differential heating and orographic lift processed during this timeframe. This result is similar to that in Lima and Wilson (2008), where the majority of CI events were found to be clustered around higher terrain until ~2 p.m. with a lack of other present mechanisms (e.g., gust fronts). The correlation of these two features with relative CI% decreases from the early to late day CI group (0.61 versus 0.35 for elevation, 0.13 versus 0.0055 for elevation gradient). The net positive difference in mean elevation and elevation gradient between the CI and non-CI samples is also most obvious in the early group (~22 meters for elevation, ~2 meters/pixel for elevation gradient).
A few trends between the different land class groups and CI% also show up. The evergreen needleleaf, deciduous broadleaf, and savanna land classes have the highest CI% in the early day group (the latter leading the way at 0.0475%), with evergreen needleleaf more commonly found in higher-elevation areas. Urban area prevails above the other land classes in the middle and late day groups, most significantly so in the latter (closely related to the inferred urban heat island effect over Jefferson County). The wetland class ranks near the bottom in the early and middle groups while the cropland class exhibits a marked drop in ranking in the late group. As with the elevation features, overall CI% bin values are highest in the middle group due to its greater CI event counts (peak CI% of ~0.09% versus 0.05% in the early group). In the 10-meter RAP wind direction distributions versus CI%, directions that possess a westerly component rank higher than those that lack such component. This is most apparent in the early day group, where directions with a dominant easterly component (E, ENE, ESE) rank in the bottom three (excluding the absent NNE and NNW directions).

There is a net positive difference in the three antecedent rainfall features (one-day, two-day, and five-day rainfall) between the CI and non-CI samples in all three time groups, suggestive of a positive relationship between CI and evapotranspiration levels (minimum discrepancies of 0.02 in. for one-day, 0.06 in. for two-day, and 0.09 in. for five-day). On average, samples with CI events also had more moist and unstable conditions than the corresponding non-CI samples in all three time groups. This is reflected in all three of the surface-based CAPE, 2-meter dewpoint, and surface-based LI distributions, consistent with the findings in Mecikalski et al. (2015). It also highlights the impact of high quantities of moisture on both cloud LCLs and instability. On average,
more local variability in the 10-meter wind exists in samples of CI events as opposed to non-CI events (average positive difference of 0.267⁰ across all time groups). This suggests the significance of features such as differential heating boundaries on where CI events occur at a given time.

Next, a subregion of higher CI event density in the early and middle day groups was isolated and examined further. Early subregion elevation correlation with CI% was found to be even stronger than in the results over the entire study domain (0.88 versus 0.61). This relation is not as apparent with elevation gradient, however (potentially due to a lack of high-gradient locations overall within the region). Mean differences in moisture and instability features, along with $\omega_{950}$, are just as sharp if not sharper than over the entire domain (e.g., mean CI versus non-CI positive difference of 478 J/kg for surface-based CAPE and 0.8⁰C for 2-meter dewpoint).

Lastly, feature importance rankings were formed using a random forest classification model of 200 decision trees, a tree depth of 20, and several bootstrapping states. Moisture and instability indicators (specifically, surface-based CAPE, 2-meter dewpoint, and surface temperature) have the highest Gini importance out of all other features in the early day group. The CI versus non-CI box plot distributions support these results, albeit the cardinality bias from the Gini method. Elevation has higher importance in the early day group than in the other two groups, connecting to its strong early CI% correlations. $\omega_{950}$ is the opposite case, instead of increased importance in the later periods. Another metric examined was the permutation importance with the selection of one feature from each indicator cluster in Figure 3.1, where the 10-meter wind direction
and antecedent rainfall features prevail as the most important. Here, the role of localized differential heating boundaries in the formation of pulse CI storms is indirectly observed.

The results discussed here, specifically the top three ranking features in permutation importance, are synthesized into a single conceptual model that depicts a typical summertime pulse convective pattern (Figure 5.1). First note the horizontal south-to-north veering of the wind direction from a dominant southerly component to a dominant westerly component, acting as a representation of the synoptic-scale surface high pressure pattern over the Southeast U.S. observed in Figure 4.1. Of course, this pattern does not hold exactly over all sampled days, which will shift the favorable CI areas on each day to a certain degree as it has been shown that the static features are not the sole mechanism. Accounting for this pattern variation, certain feature combinations are inherently more favorable for CI than others based on the outcomes of this study and previous ones. Figure 5.1 conveys one such combination of areas containing locally higher amounts of antecedent rainfall having an enhanced likelihood of CI occurring if it is collocated with or located near higher terrain and favorable low-level wind direction, also depicted with enhanced detail in Figure 5.2. This combination of enhanced latent heat flux from evapotranspiration and orographic lift is what can result in mesoscale differential heating circulations favorable for CI, depicted in Figure 5.3 (e.g., Segal and Arritt 1992; Walker et al. 2009). On the other hand, a location can have the locally higher antecedent rainfall but be situated in an unfavorable spot when it comes to whether orographic lift can occur (Figure 5.2).
Figure 5.1  A conceptual model of summertime pulse CI over USGS topography (University of Texas at Austin) in north Alabama based on the results of this study. Areas circled in red have locally higher antecedent rainfall than surrounding areas, with the two hatched areas showing where CI is more probable. Blue triangles denote where high-elevation features are located. Huntsville is represented by the purple star, and Florence/Muscle Shoals is represented by the yellow star. Dark red arrows indicate the prevalent near-surface wind directions.
Figure 5.2  A 3D version of the conceptual model in Figure 5.1 over a smaller spatial scale for the early day group, with additional features identified. Red circles represent areas of locally high antecedent rain while the blue arrows show the direction of the background 10-meter wind. The north (“N”) direction is indicated by the gray arrow. Formation of storms in the most favorable area is shown by the thunderclouds.
This study does come with sources of error and limitations, both in the methodology and the datasets that should be recognized. First, there exists the presence of mesoscale convergent boundaries (e.g., thunderstorm outflow) that can locally enhance CI event frequency, acting as the dominant mechanism (especially in the later hours). This kind of mechanism was not directly accounted for in any CI feature analyzed in this study, although the derived wind direction standard deviation can be an indicator of these boundaries even in situations where the wind speeds are small (< 5 knots). Another limitation is missing CI counts at a grid point over a 15-minute interval. Since the

Figure 5.3 A schematic of a mesoscale differential heating circulation over land. The different components are labeled accordingly, with the green slope representing an elevation feature. $L$ is the length scale, $T_1$ and $T_2$ are the layer-mean temperatures, and $p_0$ and $p_1$ are vertical pressure levels. Original figure adapted from Walker et al. (2009).
methodology was designed such that a grid point can only have a maximum of one tally per interval and CI can occur at any instance (even within the MRMS interval of two minutes), a missing CI count causes an underestimation of the true CI count. On the other end, there were also likely scattered instances where a tally was counted that was part of an existing echo. Therefore, these two effects would tend to balance each other out, however both limitations still need to be considered.

In the MRMS data there are inherent error biases in the individual radar components (e.g., transmitter, antenna, receiver) that have an effect on the reflectivity return data despite the avoidance of other radar issues such as bright banding, beam broadening, and the “cone of silence” (area near/above a single radar where the beam cannot reach when sampling data during scans) (Zhang et al. 2011). There are also discrepancies between the observed and RAP model temperature profiles which can affect, even slightly, the vertical level at which the isotherm is found. In the MODIS data, the wetland classification tends to be underrepresented in the dataset, croplands are underrepresented in tropical areas where the grid pixel sizes are much larger than average crop field sizes, and some grassland areas were accidentally classified as savannas (Friedl and Sulla-Menashe 2019). This latter issue can be of importance to the study domain as the woody savanna classification is the most prominent land class over Alabama. The AHPS antecedent rainfall has the following errors involving the assimilated radar data: presence of frozen hydrometeors, radar calibration error, varying validity of implemented Z-R relationship, and the presence of beam obstructions (DOC/NOAA/National Weather Service 2005). The RAP analysis data has had its share of error sources including issues with an overly warm and dry boundary layer in version 2 of the model that have since
been addressed in version 3, which was used in this study (Benjamin et al. 2016). This led to lesser amounts of simulated clouds and excessive downward solar radiation fluxes, addressed in the more recent versions with boundary layer temperature pseudo-innovations and improved surface observation-forecast matching methods leading to lower forecast errors in surface temperature, wind, and dewpoint (Benjamin et al. 2016). For the re-scaled feature data, it is probable that the re-scaled grids missed some smaller-scale spatial trends in the data that could have a significant impact on the discovered spatial/CI relationships.
CHAPTER 6. CONCLUSIONS

This study was centered around the hypothesis that summertime pulse CI events in the Southeast U.S. occur non-randomly and static features are most important in dictating pulse CI occurrence in the early afternoon. Meteorological features would show a lesser correlation with CI partially due to their non-static nature. If no features are significantly correlated with these CI events, then they are generally expected to be randomly distributed across the study region.

With the entirety of the results discussed, the pillar components of this study (see Chapter 1.2) are now addressed. There is an evident early CI signal with respect to elevation, as apparent in the CI% bar plots and its feature importance ranking in the early day group. From this, the significance of elevated differential heating and orographic lift is seen in the early afternoon hours. Given the subtle positive average difference in antecedent rain between CI and non-CI samples across all time groups, which also ranks high in permutation importance, it can be reasonably concluded that areas where rainfall occurred over previous days are positively correlated with the likelihood of pulse CI event formation over the study region, assuming steady state background conditions. More on this component is found in the future work section below.

On average, pulse CI events occurred in moister and more unstable conditions than in non-CI instances across all three time groups, which can lead to a lowering of cloud LCLs and/or increased instability making it easier for ascending air parcels to reach their LFC. In the Gini importance rankings, where the moisture and instability indicators ranked high among all features. Despite the clear discrepancies seen in the box plot
distributions, this indicator category ranks lower in permutation importance relative to other feature categories such as antecedent rainfall and wind variability (wind direction and $\Theta_{\text{std}}$).

Several non-random spatial patterns also emerge in the cumulative CI event plots and heat maps, with a few standing out in particular. Northeast Alabama shows CI event clusters in several different areas in the early and middle day groups. In the late day group, Jefferson County, including Birmingham proper, contains the largest cluster (again, likely attributed to an urban heat island effect). Other minor clusters are seen elsewhere, indicating more of a non-random nature when it comes to the spatial distribution of CI events. Both static and non-static features also show discernible statistical discrepancies between instances with and without CI events (e.g., elevation and antecedent rainfall). In finality, based on this along with the other results of this study, summertime pulse CI events are indeed not completely random in the Southeast U.S. and this finding should be examined further.

Additional future work building off this present study could comprise of any of the following. More features can be incorporated/derived, along with further analysis of eastern Mississippi. The relation of 10-meter wind direction to elevation gradient through the implementation of a single normalized dot product could provide better indications of low-level orographic forcing, increasing the likelihood of the LCL being attained (Nair et al. 2008). The land-surface variability (LSV) index from Gambill and Mecikalski (2011) could also be implemented. This would require not only MODIS and elevation data, but a vegetation dataset as well (e.g., NDVI). Another possible future step is the making of
average atmospheric soundings over areas of higher CI event density versus lower event density and/or CI events versus non-CI events. Further correlation analysis between different features (e.g., 2-meter dewpoint and antecedent rainfall) would help confirm the trends observed in the box plot distributions. A final idea is further machine learning analysis through additional random forest runs (e.g., incorporating different feature sets) and a shift in focus toward prediction of pulse CI events, involving the calculation of different skill scores (probability of detection, false alarm rate, critical success index). The use of other machine learning methods (e.g., support vector machines, stepwise logistical regression) to calculate feature importance could be compared with the importance rankings in this study, with agreement on the ranking of certain features signaling higher confidence of their importance. Deterministic convection-allowing NWP model runs (e.g., WRF-ARW) can be performed for intercomparisons as well with the MRMS radar observations (similarity of CI event locations).
REFERENCES


