A Radar and Model Based Synopsis of Surface Soil Moisture State
As It Relates to Back-Building Thunderstorm Behavior: Northern
Great Plains

Skye Leake
Western Michigan University, skleake96@gmail.com

Follow this and additional works at: https://scholarworks.wmich.edu/masters_theses

Part of the Geography Commons

Recommended Citation
https://scholarworks.wmich.edu/masters_theses/5182
A RADAR AND MODEL BASED SYNOPSIS OF SURFACE SOIL MOISTURE STATE AS IT RELATES TO BACK-BUILDING THUNDERSTORM BEHAVIOR: NORTHERN GREAT PLAINS

by

Skye Leake

A thesis submitted to the Graduate College in partial fulfillment of the requirements for the degree of Master of Science Geography Western Michigan University December 2020

Thesis Committee:

Lei Meng, Ph.D., Chair
Laiyin Zhu, Ph.D.
Kathleen Baker, Ph.D.
Severe thunderstorm behavior across the Great Plains of North America can result in negative economic impact, put infrastructure at risk, and pose a hazard to the lives of those calling the area home. Substantial research has been conducted throughout the Great Plains on a variety of severe weather conditions such as extreme wind, hail, tornadoes, and flash flooding, however such research is limited in the northern portion of the Great Plains. Some research has been conducted on these severe phenomena produced by thunderstorms in the northern Great Plains; however, little has been done in the realm of training/back-building of storms. This behavior has the potential to cause extreme flash flooding.

The multi-parameter interactions of local precipitation recycling feeding into thunderstorm development is still an open research question. From an observational standpoint this study identifies potential back-building behavior in NEXRAD radar imagery, relating that to high resolution Land Information System surface soil moisture data orientated to the background flow of the planetary boundary layer. Several weak correlations between soil moisture and NEXRAD reflectivity were observed, however these were not independent of the potential influence of the Great Plains Low Level Jet. Points of convective initiation were examined spatially, finding regions which warrant further analysis as a hotspot for back-building behavior.
DEDICATION

To all current and would-be graduate students unable or hindered in the completion of their studies by the emergence of SARS-CoV-2.

Skye Leake
TABLE OF CONTENTS

DEDICATION ........................................................................................................................................... ii
LIST OF TABLES .................................................................................................................................... vii
LIST OF FIGURES ................................................................................................................................... viii
LIST OF ABBREVIATIONS ....................................................................................................................... x

CHAPTER

I. INTRODUCTION AND BACKGROUND ............................................................................................... 1
   1.1 Scope of Phenomenon Under Study ............................................................................................... 1
   1.2 Study Area / Study Period ............................................................................................................. 2
   1.3 Physiography ............................................................................................................................... 4
       1.3.1 Climate .................................................................................................................................. 4
       1.3.2 Geology ............................................................................................................................... 5
       1.3.3 Land Use and Land Cover ................................................................................................. 9
   1.4 Meteorology ............................................................................................................................... 11
   1.5 Summary ..................................................................................................................................... 13
   1.6 Research Questions ..................................................................................................................... 13

II. REVIEW OF LITERATURE AND THEORETICAL FRAMEWORK ................................................ 15
   2.1 Introduction ............................................................................................................................... 15
   2.2 Meso-β-scale Circulations .......................................................................................................... 16
       2.2.1 Soil Moisture Feedback Mechanisms and Regional Precipitation Recycling ...................... 16
Table of Contents–Continued

CHAPTER

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2.2 Synoptic Influence on Convection Along Soil Moisture Boundaries</td>
<td>17</td>
</tr>
<tr>
<td>2.2.3 Surface Heterogeneities and the Generation of Nonclassical</td>
<td>20</td>
</tr>
<tr>
<td>Mesoscale Circulations</td>
<td></td>
</tr>
<tr>
<td>2.3 Convective Initiation – As Aligned to Underlying Boundary Layer</td>
<td>25</td>
</tr>
<tr>
<td>Flow</td>
<td></td>
</tr>
<tr>
<td>2.4 Regional Thunderstorm Climatology</td>
<td>26</td>
</tr>
<tr>
<td>2.5 Summary</td>
<td>27</td>
</tr>
<tr>
<td>III. METHODOLOGY</td>
<td>29</td>
</tr>
<tr>
<td>3.1 Data Selection and Filtering</td>
<td>29</td>
</tr>
<tr>
<td>3.1.1 NASA SPoRT Land Information System Data</td>
<td>30</td>
</tr>
<tr>
<td>3.1.2 Global Forecast System 4 Analysis Data</td>
<td>31</td>
</tr>
<tr>
<td>3.1.3 WSR-88D NEXRAD Level II Data</td>
<td>32</td>
</tr>
<tr>
<td>3.1.4 Haberlie and Ashlie MCS Database</td>
<td>32</td>
</tr>
<tr>
<td>3.1.5 Selection of Test Dates</td>
<td>33</td>
</tr>
<tr>
<td>3.1.6 Data Availability</td>
<td>36</td>
</tr>
<tr>
<td>3.1.7 Simplification Assumptions for Input Data</td>
<td>40</td>
</tr>
<tr>
<td>3.2 Data Processing and Region of Interest Extraction</td>
<td>40</td>
</tr>
<tr>
<td>3.2.1 Analysis of WSR-88D NEXRAD Data</td>
<td>41</td>
</tr>
<tr>
<td>3.2.2 Analysis of NASA SPoRT LIS Relative Soil Moisture Data</td>
<td>44</td>
</tr>
<tr>
<td>3.3 Data Relationships and Distributions</td>
<td>48</td>
</tr>
</tbody>
</table>
Table of Contents–Continued

CHAPTER

3.3.1 Intra-Variable Relationships ................................................. 48
3.3.2 Analysis of Covariance ......................................................... 49
3.3.3 Spatial Distribution of Cases ................................................. 49
3.4 Error Sources, Error Propagation, Error Management ..................... 51
  3.4.1 Error Sources ........................................................................ 52
  3.4.2 Error Propagation ................................................................ 52
  3.4.3 Error Management ............................................................... 52

IV. RESULTS .................................................................................. 56
  4.1 Correlations .............................................................................. 56
  4.2 Analysis of Covariance ............................................................... 58
  4.3 Spatial Distributions ................................................................. 60

V. DISCUSSION .............................................................................. 69
  5.1 Methods Employed ................................................................. 68
    5.1.1 Data Sources ..................................................................... 69
    5.1.2 Data Processing ............................................................... 69
    5.1.3 Statistical Measures .......................................................... 70
  5.2 Results .................................................................................... 70
    5.2.1 Correlation Within the Context of the GPLIJ ....................... 71
    5.2.2 Spatial Distributions .......................................................... 73
  5.3 Conclusions ............................................................................. 75
APPENDICIES

A. NEXRAD Data Extraction .................................................................................. 78
B. SPoRT LIS Data Extraction ............................................................................... 79
C. Variable Descriptions ......................................................................................... 80

BIBLIOGRAPHY ........................................................................................................ 82
LIST OF TABLES

3.1 Class breaks for PRISM Mean Annual Precipitation (liquid equivalent) ..................... 51

4.1 Multi-scale quadrat analysis of convective initiation points ........................................ 60

4.2 Statistical significance of PRISIM classifications ...................................................... 61
LIST OF FIGURES

1.1 Physiographic features of study area .................................................................3

1.2 Annual mean precipitation (mm) within study area .............................................6

1.3 Density map of irrigation throughout the contiguous United States, 1982 ..........10

3.1 Approximate spatiotemporal resolutions of data used and other common datasets ..........30

3.2 Alignment surface soil moisture (SPoRT LIS RSM 0-10cm) to the direction of the surface wind background flow (GFS4 10m components) ....................................................32

3.3 Refinement and identification of convective initiation locations ...............................35

3.4 Spatiotemporal distributions of identified test cases .............................................37

3.5 WSR-88D NEXRAD coverage of study area at a minimum of 10,000 ft above ground level ..................................................................................................................38

3.6 Region of Interest polygon masking underlying NEXRAD Level II Data .................42

3.7 Examples of build-up rates (green lines) of time series as calculated on NEXRAD Level II data from the ROI .................................................................44

3.8 Bivariate gaussian distribution applied in weighted calculations .............................46

3.9 Upwind versus crosswind means of a LIS SSM surface ........................................47

4.1 Correlation matrix between derived SPoRT LIS Soil Moisture (0-10cm), and presence of GPLLJ versus the derived values from NEXRAD data ($\alpha=0.05$) ..............57

4.2 Correlation matrix between derived SPoRT LIS Soil Moisture (0-10cm), and presence of GPLLJ versus the derived values from NEXRAD data ($\alpha=0.10$) ..............58
List of Figures – Continued

4.3 Statistical significance of ANCOVA between the GPLLJ and the dependent variable, controlling for the independent variables (α = 0.05) ............................................. 59

4.4 Statistical significance of ANCOVA between the GPLLJ and the dependent variable, controlling for the independent variables (α = 0.10) ............................................. 60

4.5 Precipitation polygon bins used for frequency distribution and observed convective initiations by bin ........................................................................................................ 63

4.6 Hotspot analysis (reg. rect. grid of 0.75° by 0.75°) ................................................................. 64

4.7 Hotspot analysis (reg. rect. grid of 1.0° by 1.0°) ..................................................................... 64

4.8 Hotspot analysis (reg. rect. grid of 1.5° by 1.5°) ..................................................................... 65

4.9 Mean LIS SSM relative to the convective initiation point, all cases ...................................... 66

4.10 Mean LIS SSM relative to the convective initiation point, 13.25-19.00 in precipitation, based on PRISM annual means ................................................................. 66

4.11 Mean LIS SSM relative to the convective initiation point, June ............................................. 67

4.12 Mean LIS SSM relative to the convective initiation point, July ............................................. 67

4.13 Mean LIS SSM relative to the convective initiation point, August ........................................ 68
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANCOVA</td>
<td>Analysis of covariance</td>
</tr>
<tr>
<td>ASCAT</td>
<td>Advanced Scatterometer</td>
</tr>
<tr>
<td>CO</td>
<td>Colorado</td>
</tr>
<tr>
<td>GFS</td>
<td>Global Forecast System</td>
</tr>
<tr>
<td>GFS3</td>
<td>Global Forecast System 3</td>
</tr>
<tr>
<td>GFS4</td>
<td>Global Forecast System 4</td>
</tr>
<tr>
<td>GPLLJ</td>
<td>Great Plains Low Level Jet</td>
</tr>
<tr>
<td>LIS</td>
<td>Land Information System</td>
</tr>
<tr>
<td>LLJ</td>
<td>Low Level Jet</td>
</tr>
<tr>
<td>MAUP</td>
<td>Modifiable Areal Unit Problem</td>
</tr>
<tr>
<td>MCS</td>
<td>Mesoscale Convective System</td>
</tr>
<tr>
<td>MN</td>
<td>Minnesota</td>
</tr>
<tr>
<td>MO</td>
<td>Missouri</td>
</tr>
<tr>
<td>NAM</td>
<td>North American Mesoscale Forecast System</td>
</tr>
<tr>
<td>NASA</td>
<td>National Aeronautics and Space Administration</td>
</tr>
<tr>
<td>NCMC</td>
<td>Nonclassical Mesoscale Circulations</td>
</tr>
<tr>
<td>ND</td>
<td>North Dakota</td>
</tr>
<tr>
<td>NEXRAD</td>
<td>Next-Generation Radar</td>
</tr>
<tr>
<td>NOAA</td>
<td>National Oceanic and Atmospheric Administration</td>
</tr>
<tr>
<td>PBL</td>
<td>Planetary Boundary Layer</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>PRISM</td>
<td>Parameter-elevation Regressions on Independent Slopes Model</td>
</tr>
<tr>
<td>ROI</td>
<td>Region of Interest</td>
</tr>
<tr>
<td>RSM</td>
<td>Relative Soil Moisture</td>
</tr>
<tr>
<td>SD</td>
<td>South Dakota</td>
</tr>
<tr>
<td>SPoRT</td>
<td>Short-term Prediction Research and Transition Center</td>
</tr>
<tr>
<td>SSM</td>
<td>Surface soil moisture</td>
</tr>
<tr>
<td>WRF</td>
<td>Weather Research and Forecasting</td>
</tr>
<tr>
<td>WSR-88D</td>
<td>Weather Surveillance Radar 1988 Doppler</td>
</tr>
<tr>
<td>Øe</td>
<td>Theta-E</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION AND BACKGROUND

Chapter I: Introduction and Background is organized into five sections. The broad context of the phenomenon under study is introduced, followed by the study area /period, physiography of the study area as well as the core meteorological phenomena. As this is an observational study, there will be limited coverage of additional meteorological phenomena that may be present. Finally, the impacts of and the degree to which they have been studied in the northern Great Plains will be discussed and summarized.

1.1 Scope of Phenomenon Under Study

Thunderstorms throughout the Great Plains of North America can produce a variety of severe weather phenomenon. These range from tornadoes, straight line winds, hail, and flash flooding. The Great Plains contain a large portion of the United States agricultural industry, with a significant fraction of national yield of corn, wheat, and soybeans. Severe weather produced by thunderstorms can occur over the entirety of the Great Plains (from Texas extending into Manitoba), with occasional squall line thunderstorms nearly reaching across the entirety of the Great Plains. Such systems have been extensively studied throughout Texas, Oklahoma, and Kansas. Squall lines are not the only source of severe thunderstorm generation, portions of the Great Plains are subject to several mesoscale convective systems (MCS) over the course of a year.

The northern portion of the Great Plains have not undergone the extensive levels of study as the states further south, despite having large amounts of cropland and potentially vulnerable
infrastructure. This study aims to quantify spatial distributions and surface moisture characteristics of a certain thunderstorm behavior, back-building, which may result in flash flooding. This is accomplished primarily through the relation of land surface model data, radar reflectivity data, and general circulation model data.

1.2 Study Area / Study Period

The northern Great Plains as defined in this work, consist of the area bound between 40°N to 49°N and -95.31°W to -104.05°W. This includes the entirety of the states of North Dakota (ND), South Dakota (SD) and Nebraska (NE), as well as portions of Colorado (CO), Minnesota (MN), Iowa (IA), and Missouri (MO), as in Figure 1.1. This area was chosen based on two factors: 1) limited topographic features (thus limiting the influence of orthographic uplift), 2) few published studies of the region in relative to Kansas/Oklahoma/Texas studies or the entirety of the Great Plains. Dataset details and the selection process will be further discussed in Chapter III - Methodology. Numerous small areas are masked out of underlying data, as these coincide with water features or urban areas larger than ~9km². Analyses within this study were performed with a variety of datasets: 1) Short-term Prediction Research and Transition Center (SPoRT) Land Information System (LIS) model-based soil moisture, Next-Generation Radar (NEXRAD) reflectivity data, Global Forecast System 4 (GFS4) analysis data of surface wind conditions, and Parameter-elevation Regressions on Independent Slopes Model (PRISM) climatological means. Due to limitations imposed by data availability 2007 – 2017 (inclusive) defined the study period. The months of March – October were examined, however, soil moisture memory from previous years through the winter months limit the independence of each year.
Figure 1.1 – Physiographic features of study area
1.3 Physiography

The study area is relatively uniform. Small topographic features exist, chiefly the Black Hills of SD and the Sandhills of NE. The higher plains containing the Missouri River Plateau within the Dakotas slowly give way to broad swaths of glacial till encompassing most of the eastern portion of the study area (Figure 1.1). The elevation of the region is generally higher in the west trending lower towards the east. The major waterbodies in the area consist of manmade reservoirs on the Missouri River. Most of the study area is encompassed by the Missouri River watershed, ultimately discharging at the Gulf of Mexico.

1.3.1 Climate

The portion of the northern Great Plains contained by the study area primarily consists of continental climate types with small portions classified as a semi-arid climate. Through the warm months evaluated in this study (March through October) the surface routinely switches from an energy-limited evapotranspiration regime to a moisture-limited one.

*Seasonality*

Most of the study area has seasonal changes in temperature and precipitation consistent with a humid continental climate. The northern Great Plains experience harsh winters with the influence of the Rocky Mountains, prairie provinces of Canada and northern latitudes creating and maintaining cold dry air masses over the region. This freezes soil moisture, particularly in ND, limiting the availability of moisture lower in the soil column for several weeks after the surface has thawed. Frozen layers of soil also influence the permeability of the soil column to moisture at the surface following the winter months. The presence of snow influences the albedo and insulates the surface, further influencing rates of soil thaw.
Moisture Gradient

The mean annual precipitation ranges from ~330mm to ~900mm (available online at http://www.prisim.oregonstate.edu). The precipitation gradient across the study area extends from northeast Montana (least) to southeast Nebraska (most). These PRISM data utilized for annual mean precipitation over the study area can be seen in Figure 1.2. The transition between these two extremes is approximately linear. There exists an anomaly around Rapid City, SD, a result of orthographically forced precipitation from the Black Hills. Rapid City, SD receives 465mm (~18.32in) of mean annual precipitation (liquid equivalent), whereas Dickinson, ND receives 400mm (~15.75in.), Crosby, ND receives 375mm (~14.95 in.), and Alliance, NE receives 380mm (~14.96 in). The eastern portion of the study area receives more mean annual precipitation. Grand Forks, ND receives 545mm (21.46in.), Sioux Falls, SD receives 672 (26.46in.), Omaha, NE receives 786 (30.94in) (https://prism.oregonstate.edu/normals).

1.3.2 Geology

Topography

The elevation profile of the northern Great Plains is characterized by a slight elevation change with the maximum elevation occurring in the western portion (Black Hills), and a minimum elevation in the southeast. There are few anomalies above the prairie’s surface. In the northwestern portion of the study area the major topographic features are buttes and the badlands. Rivers and creeks cut into the prairie surface in the western and central portions of the study area, with valleys a few tens of miles wide. The dominant physiographic features are the Missouri River, Platte River, Nebraska Sand Hills, and the Black Hills. The rest of the area was prairie a century or two ago, now primarily under agricultural use.
Figure 1.2 – Annual mean precipitation (mm) within study area. Climatological period 1981-2010. Derived from: PRISM (https://prism.oregonstate.edu/normals/ (2015))
Black Hills

The Black Hills of western SD is the one major anomaly in the study area occurring along the South Dakota/Wyoming border (see Figure 1.1). The Black Hills have a maximum elevation of 2,207m above ground level and cover an area of ~12,000 km². This has resulted in abnormal precipitation patterns within the local area due to orthographic uplift.

Drift Prairie

The central portion of the Dakotas contains two predominant hydrological features: 1) the Missouri River, which bisects the region from the northwest to the southeast, 2) the drift prairie to the east of the Missouri River. The hydrologic features of the drift prairie are characterized through the presence of poorly drained glacial till with depressions formed through the uneven deposition of material from retreating continental glacial activity (Wisconsinan glacial period, 90,000 – 11,500 years ago) (Joshi et al., 2017). The glacial dill has slowly shifted throughout the millennia leaving shallow depressions, locally referred to as “sloughs”. They collect local precipitation and snowmelt, resulting in a dense system of venal ponds. Fieldwork in similar dune formations in the southern Great Plains found the inclusion of fine-grained particles, particularly clay, in the upper horizons of the soil resulted in better overall soil cohesion and moisture retention capacity (Leland H. Gile, 1979; Muhs & Wolfe, 1999). Recent work has been undertaken in the examination of the water height within sloughs to better understand the cyclic nature and coupling with local soil moisture (Huang et al., 2013). The moisture availability in the surface soil coupled with the numerous shallow depressions left in the landscape and the generally low surface albedo of heavy tillage soils in the region provide a large capacity for rapid changes to the soil and latent heat fluxes, enabling rapid evapotranspiration. This provides a
source of high Theta-E $\theta_e$ air; mixed into the planetary boundary layer (PBL) in April through June.

**Soils**

The soil column interaction with the PBL is limited to the first few millimeters. The potential available moisture within the column extends deeper into the column. This deeper moisture lags that of the surface. The layers of soil over time act as a low pass filter, smoothing the variance of excess (deficit) surface moisture. Whereas there is some capillarity action between soil particles to move moisture from lower layers towards the surface, the primary driver for lower level moisture transport is through vegetation. The evaporative regimes transition between that of an energy-limited regime and a moisture-limited regime. In an energy-limited regime moisture is in abundance and energy to drive evaporation is in a deficit, temperatures at the surface stay relatively constant. Moisture-limited regimes have an abundance of energy at the surface, but the lack of moisture and low specific heat of the soil constitutes high variability in surface temperature. Vegetation at the surface buffers the steep contrasts seen over bare soil surfaces in different energy regimes, however due to the access vegetation has to moisture deeper in the soil column, contrasts can still be prevalent. The contrast of surface conditions over mesoscale domains can drive shallow convection, as can contrasts in the vegetation over the surface (Segal et al., 1988).

The study area does have some contrast in soils spatially. The western edge experiences a mild semi-arid climate, modifying the vegetation type/coverage and soil properties. However, the western edge of the study area has different land use land cover characteristics than that of the eastern. Soil types/distributions could be an influencing factor in the development of shallow
convection. However, these potential interactions will not be considered within the scope of this study.

**Drainage**

Much of the region is dominated by the Missouri River drainage basin which eventually discharges at the Gulf of Mexico. The Platte River, part of the Missouri River Basin, drains large portions of NE. Portions of eastern ND, northeast SD, and western MN are contained within the Red River watershed which discharges at Hudson’s Bay.

Drainage, across portions of the study area is poor in comparison to other areas of the continent, with some areas comprised of multiple sloughs per square mile. The large fraction of lands in agricultural use also influences the permeability of the surface, lessening runoff potential. Shallow grades across swathes of the northern Great Plains limit runoff as well. The portions of the study area have the spring thaw intersect the period of early season storms, further influencing surface runoff characteristics.

1.3.3 Land Use and Land Cover

**Agricultural**

Agricultural activity in the study area dominates land use, with rangelands in the west, transitioning to no-till/ partial tillage in the central region and finally towards heavy tillage agriculture in the east. There are two exceptions to this trend: 1) the Sandhills of NE are often classified as grasslands, 2) the Black Hills of SD are a major topographic feature in the area, with heterogeneous land use and land cover not suitable for widespread agriculture.

The use of irrigation equipment varies across the study area. center-pivot irrigation systems are common; their presence is directly tied to the dominant crops in the area. Center-
pivot irrigation systems have different degrees of water use efficiency, this is dependent on the design of the sprinkler heads, type/stage of growth of the crops, and time of irrigation during the diurnal cycle. Convection over large areas of irrigation has been under study for decades, see Figure 1.3 (McGuire, 2009; Segal & Arritt, 1992; Yang et al., 2019).

Figure 1.3 – Density map of irrigation throughout the contiguous United States, 1982. Adapted from (Segal & Arritt, 1992). Area “7” in this map falls within southern Nebraska.

Note: this figure is from after the significant uptick in irrigated crops across the Great Plains (1949 - 1974), when the groundwater withdrawals increased by ~475%, remaining relatively constant thereafter (McGuire, 2009).
The type of crops grown have a large role in the amount of water usage. Corn has been documented to significantly alter the PBL moisture characteristics through evapotranspiration and trigger convection in conjunction with evening temperature inversions (Brown & Arnold, 1998; Segal et al., 1988). This phenomenon is in part tied to moisture availability near the surface. Other crops will produce similar alterations to the PBL, however, of commonly planted crops in the study area, corn has the highest per-day water usage, however due to crop rotations and the differences in arable land across the study area, the magnitude of water usage varies (Lund et al., 1993; Schlegel et al., 2016).

1.4 Meteorology

Scales of weather phenomena over the northern Great Plains fall under two main categories, synoptic and mesoscale. Synoptic phenomena operate at a scale of hundreds to thousands of kilometers and can last for multiple hours to days. Mesoscale features are smaller, operating at spatial scales of kilometers and temporal scales from minutes to a few hours. A single thunderstorm cell, back-building storms, MCS, and larger mesoscale convective complexes are all mesoscale features. Midlatitude cyclones and similar frontal based systems operate on the synoptic scale, providing forcing for more localized events.

Prevailing atmospheric flow over the northern Great Plains is from the west, however topography dictates most moisture that may originate from the Pacific is mainly lost due to orthographic forcing as it passes the Rocky Mountains. During the summer months an upper level ridge may divert this flow such that it comes from the northwest (air subsidence due to negative vorticity) (Koster et al., 2004). The Great Plains Low Level Jet (GPLLJ) is responsible for the advection of moisture from the Gulf of Mexico northward, parallel to the Rocky Mountains. The GPLLJ is an important source of high θ_e air, which when coupled with favorable
air profile characteristics, provides energy for several types of thunderstorm systems. Not all moisture for these systems is provided by the GPLLJ, another source is regional/local precipitation recycling, where moisture that falls from one storm may fuel a subsequent storm developing over the region days or weeks later. Some portions of the study area are more likely to exhibit this recycling behavior; however, identification of these areas is highly dependent on the scale of analysis, soil properties of the region, and surface energy balance (i.e. albedo, presence of vegetation) (Meng & Quiring, 2010).

Back-building thunderstorms are a type of training thunderstorm behavior, where the development of new cells in a multicell cluster occurs on the upwind segment. This results in the storm appearing stationary (or even propagating up-wind) on radar imagery. As the cells develop, they move downwind, and a new cell develops over the same location on the Earth’s surface. This behavior may be partially tied to the surface conditions, in particular the available surface moisture may locally feedback into a system. However, this relies on favorable atmospheric (in)stability and windshear to facilitate moisture transport from the surface (Hitchcock & Schumacher, 2020). If training/back-building does occur, it has the potential to create extreme levels of local rainfall, causing flash flooding, if the soil was already saturated. Such extreme outcomes of these events are rare, however one such devastating event occurred near Rapid City, SD on June 9-10, 1972. Over a small area ~380 mm (~15in) of rainfall was recorded, which when coupled with high levels of soil moisture and supporting topography resulted in the deaths of 238 individuals and over 3,000 injuries, with a property damage cost of over $160 million in 1972 dollars (~$995 million in 2020) (Carter et al., 2002).
1.5 Summary

The northern Great Plains are a relatively homogeneous study area in terms of topography. Several factors dictate the gradual temperature and moisture gradients across the region. The differences in mean annual precipitation and the continental climate drive the land use land cover of the study area, which in turn has an impact on the chances for local circulations to manifest as convective rainfall.

The northern Great Plains region, despite its location in the center of the North American continent, receives a large amount of moisture from the Gulf of Mexico via the GPLLJ. This moisture, held aloft, provides a source of energy for convection through latent heat release across the entirety of the Great Plains. Although convection organizes in several ways, one of the most devastating in terms of rainfall for a local area can be through the back-building of thunderstorms. Moisture available for local precipitation recycling may be forced aloft by local circulations at the surface, driven by heterogeneities of the surface such as moisture and vegetation differences.

1.6 Research Questions

This study sets out to answer questions under two main themes. Firstly, as a means of identifying the broad patterns soil moisture surface and the observed convection. Secondly, the identification of geospatial characteristics of these storms. These themes are embodied in two research questions.

1) What are the general characteristics of surface soil moisture local to points of convective initiation exhibiting back-building/training behavior within the selected test cases?
2) Is the spatial distribution of these events the result of random clustering? Do hotspots of convective initiation exist?
CHAPTER II

REVIEW OF LITERATURE AND THEORETICAL FRAMEWORK

The review of the literature and the theoretical framework is organized into five sections: 1) a short introduction to the topics, 2) the generation of meso-β-scale circulations relative to surface conditions, 3) convective initiation – as viewed relative to the background flow of the PBL, 4) research into the climatology of thunderstorms within the study area, 5) a short summary of the former topics highlighting the most pertinent concepts from the literature.

2.1 Introduction

The field of atmospheric sciences is of a broad nature in its own right; further confounded by its interdependence on other Earth systems. With regard to the multitude of ways the land surface of the planet and its atmosphere interact, the focus for this research was based around two primary components: 1) the development of deep convection leading to enhanced storm systems, 2) the formation of a methodology suitable for analyzing the former with limited bias. To recognize these goals in the timeframe allotted, a multidimensional approach was employed. This leveraged the theoretical groundwork for the development of mesoscale circulations capable of spurring deep convection, previous research bounding the spatiotemporal aspects of the study area, and analysis methods agnostic to the underlying data structure, collection, and storage.

Meso-β-scale (domains of horizontal scale of 20-200km) circulations can develop under a wide array of surface and PBL characteristics with deep convection. The background flow and physical properties of the mid-upper atmosphere (≤850hPa) can energetically alter the behavior of storms originating from these circulations. Most meso-β-scale activity occurs without coupling to the lowest levels of the atmosphere/surface. Under the right conditions deep
convection can persist from a given location and produce back-building/training storms. Under certain circumstances the product of this behavior results in some geographic locations experiencing extended periods under strong convection. This can have profound effects on the local hydrologic cycle, as the rate and spatial coverage of precipitation affects infiltration and runoff ratios. Localized circulations may also be responsible for entraining of large amounts of high $\theta_e$ air to the mid-levels of the atmosphere when the synoptic conditions are not providing large amounts of mid-level moisture. This may aid in further convection when the atmosphere is in a synoptically primed state, contributing to precipitation over a larger region. However, the atmosphere must be in a stability state favorable for the incorporation of any surface moisture entrained in near surface circulations.

Chapter II aims to amass the existing concepts of the development of meso-β-scale circulations and place them in the physiographic context of the north plains study area. Deep convection in Weather Surveillance Radar 1988 Doppler (WSR-88D) NEXRAD system (and its precursors) are examined along with the development of the SPoRT LIS soil moisture model dataset. Finally, this chapter aims to address general thunderstorm climatology within the study area to establish intra-annual temporal considerations utilized in Chapter III - Methods.

2.2 Meso-β-scale Circulations

2.2.1 Soil Moisture Feedback Mechanisms and Regional Precipitation Recycling

The Great Plains of North America exhibit soil moisture feedback characteristics juxtaposed to those neighboring the region or elsewhere on the globe. Northern portions of the Great Plains, extending from near the Gulf of Mexico to the southern plains of Manitoba, display areas of prolonged soil moisture memory >90 days due to frozen soil during the winter months (Dirmeyer et al., 2009). The soil moisture memory (lag associated with anomalies within the soil
column) has been found in numerous studies to be an important component of weekly, monthly, and seasonal variability of atmospheric interactions controlling the potential evapotranspiration (Beljaars et al., 1996; Dirmeyer, 2000; Douville, 2002; Fennessy & Shukla, 1999; Koster & Suarez, 1999; Schlosser & Milly, 2002). Large portions of North America, including seasonally frozen areas of the Great Plains, maintain a strong positive correlation with the soil moisture memory, although the lag term decreases to ~40 days in spring months and 15-20 days in winter months (Dirmeyer et al., 2009).

Precipitation recycling is moisture that originated from within a defined region and the atmosphere above then subsequently falls again onto the land surface of that same region, is often expressed as a ratio of local moisture to moisture that has been transported through advection into the region. The lag associated with this recycling operates on a scale of hours to months and is tied to the soil moisture memory and evapotranspiration characteristics. Dependent on the scale of analysis, the recycling mechanism may be triggered by synoptically favorable conditions or mesoscale circulations. Determination of the local contribution of moisture regarding storm frequency and intensity has resulted in the development of techniques to trace moisture sources (Feng et al., 2012; Yang et al., 2019). The moisture source for Great Plains is often the GPLLJ, which transports moisture from the Gulf of Mexico northwards. This source is not always present, which in turn affects the distribution of convection over the Great Plains, both seasonally and diurnally (Frye & Mote, 2010).

2.2.2 Synoptic Influence on Convection Along Soil Moisture Boundaries

The oceans have a vast thermal energy capacity and circulate slowly in comparison to the atmosphere. Whereas knowledge of the connection between sea surface temperatures and precipitation recycling remains limited, there exist positive correlations between the two for
precipitation over the Great Plains. Dirmeyer et al. (2009) found strong agreement between several land surface models and regional precipitation feedback. Over the Great Plains it was found ~12% of moisture underwent recycling between the two land surface models employed. Daily evapotranspiration was found to be significant with respect to precipitation recycling and the recycling ratio across large portions of the Great Plains in the spring and summer months. Longer term analysis was performed by Meng and Quiring (2010), where precipitation totals from the years spanning 1915-2007 were used as a proxy for soil moisture data. In years where the Niño pattern was present, it was found that higher sea surface temperatures corresponded with greater precipitation. The positive correlation between these variables was most prevalent in southern Oklahoma and the eastern edge of the north plains study area.

Several factors can provide (un)favorable conditions for convective initiation to occur, such as surface heating, vegetation distribution, and wind velocity within different layers of the troposphere. Under ideal conditions the PBL/surface coupling will provide high \( \theta_e \) air, favoring the development of deep convection under certain atmospheric stability conditions. However, the PBL is not always coupled to the surface, and the atmospheric profile is not always ideal for deep convection to occur. The surface properties of the soil (i.e. vegetation, specific heat, moisture capacity) may also inhibit convection.

Analysis on the sensitivity of convection developing over soil moisture heterogeneities was performed by Frye and Mote (2010). Over a seven year period (1998-2004) the southern Great Plains were categorized on two factors: Synoptically primed (benign) days and the presence (lack) of the Low Level Jet (LLJ). The study area was centered on the Weather sensing Radar 1988 Doppler (WSR-88D) site at Oklahoma City/Twin Lakes (KTLX) and comprised of a 3° latitude x 3° longitude region. This location was chosen as representative of the southern
Great Plains for the warm season (April – September), as well completely contained within the soil moisture feedback hotspot identified by numerous studies (Koster et al., 2004; Meng & Quiring, 2010). Frye and Mont (2010) found a greater probability of convective initiation occurring over Oklahoma when the conditions of low soil moisture areas (~15%) and the presence of the LLJ were met.

Frye and Mote (2010) identified Synoptic conditions via morning soundings (1200 UTC) for several stations in the surrounding region. The presence of the LLJ was determined via 850hPa charts. Synoptically benign (SB) days were classified as days where the lapse rate from 850hPa – 700hPa had a lapse rate of < 6.0°C km⁻¹, whereas synoptically primed (SP) days were classified as the lapse rate ≥ 6.0°C km⁻¹. The presence of the LLJ was determined under the following criteria: 1) windspeeds ≥ 25 kt (12 m s⁻¹) at 850hPa, 2) winds originate from the Gulf of Mexico, 3) winds cover a majority of the southern Great Plains. Combining the synoptic and LLJ conditions gives 4 different types of synoptic days: 1) NoLLJ/SB, 2) NoLLJ/SP, 3) LLJ/SB, 4) LLJ/SP (Bonner, 1968).

Soil moisture data for the Fyre and Mote (2010) study were sourced from the Tropical Rainfall Measuring Mission Microwave Imager. The daily soil moisture data product was derived from the Microwave Imager (active X-band radar at 10.65 GHz) and auxiliary data. The final data product offered ~0.5cm penetration of the surface on a ⅛° grid (Gao et al., 2006). These soil moisture data are volumetric, expressed as a percentage between plant wilt and saturation (relative soil moisture). These data closely correlate with directly sensed soil moisture data from the Oklahoma Mesonet, however this is subject to differing data collection spatial scales (Gao et al., 2006). The soil moisture gradient surface was also calculated in this analysis.
Frye and Mote (2010) identified convection initiation locations in the study area for all days with available data where reflectivity $\geq 35$dBZ. Wilson and Schreiber (1986) and Wilson and Mueller (1993) employed thresholds of $\geq 30$dBZ, Parker and Johnson (2000) utilized $\geq 40$dBZ with a core $\geq 50$dBZ. Mecikalski and Bedka (2006) used reflectivity threshold of $\geq 35$dBZ to reduce the chances of ground clutter interference near the sensor location. Convective events had to: 1) occur solely within the study region, 2) initiation of the event had to be persistent across three sequential radar time periods, 3) convective cell had to be isolated from other high reflectivity areas. The former two criteria were used in the filtering of events to remove intensifying events, rather than points of new convection (Frye and Mote, 2006). Each convective initiation point was then linked to a Tropical Rainfall Measuring Mission Microwave Imager soil moisture grid cell center with the nearest neighbor.

Frye and Mote (2010) found LLJ/SP days to be the most common for the study area, with NoLLJ/SP being the second most common. When probabilities of convective initiation were calculated for the soil moisture values it was found that, as expected, LLJ/SP days had higher a likelihood of occurring over any point on the surface (mean of $\sim 15\%$ volumetric soil moisture), whereas NoLLJ/SB days were found to have convective initiation occur over soil moisture cells with higher values (mean of $\sim 25\%$ volumetric soil moisture).

2.2.3 Surface Heterogeneities and the Generation of Nonclassical Mesoscale Circulations

Segal and Arritt (1992) examined the relationship between surface heat flux gradients and the presence of Nonclassical Mesoscale Circulations (NCMC). This study employed both a numerical modeling component as well as addressing observation data of PBL flow. The presence of perturbed areas that occurred from surface heterogeneities of evapotranspiration, albedo, and heat flux differentials (Bowen Ratio) were examined (Segal et al., 1988). These
circulations were termed NCMCs - where circulation occurred or disrupted the background flow between a perturbed area and the surroundings (Pielke & Segal, 1986; Segal et al., 1984). This mechanism operates on the difference between sensible heat fluxes of the and the surroundings, the size and strength are determined by the magnitude of the difference (Entekhabi et al., 1996). Perturbed areas were determined to develop based on a multitude of surface heterogeneities (over different areas of the planet), vegetation, surface albedo, surface moisture, atmospheric optical thickness, sea surface temperatures, and differences in polar ice distribution. The scale at which NCMCs operate varies by the source of the PA, however all are within the meso-β-scale regime. NCMCs originating from a soil moisture perturbed area was found to occur at the highest frequency with a temporal persistence scale of days, these were most likely generated at the meso-β-scale from spatial variation in rainfall. A soil moisture perturbed area could develop from only marginal uniformity of the area (some patchiness of surface characteristics tolerated), and with a moderate difference in sensible heat flux versus the surroundings (in comparison to other perturbed area types). A NCMC may not be strong enough to overcome the background flow, as these circulations are found to be ~5ms⁻¹ under ideal conditions. If the circulation does not reach the threshold necessary to overcome the background flow, then it may be able to be detected as a deviation from the background flow across a homogenous area.

NCMCs developing from soil moisture heterogeneities can be difficult to detect with current remote sensing techniques due to relatively short lifespans, however there are locations in the contiguous United States that are favorable for their formation, thusly there is a higher chance for detection. Areas of heavy tillage agriculture, particularly those containing center pivot irrigation systems, provide two common modes for perturbed area development: soil moisture and vegetation heterogeneities (Segal et al., 1988). The development of cooler temperature
irrigated islands in the southern Great Plains had been observed from the early 1980s when transects were flown across them at multiple altitudes (Barnston & Schickedanz, 1984). These islands had up to 2°C cooler maximum air temperature when compared to the surrounding environment on days that were synoptically undisturbed.

In vegetation perturbed areas the conditions dictating the closure of the stomata on the leaves of the irrigated crops and the resultant loss of evapotranspiration are reduced. This is due to the land surface operating within a dominantly energy-limited regime (Koster et al., 2009; Segal & Arritt, 1992). The presence of irrigation hotspots, often following natural hydrological features such as rivers or aquifers, is similar to that proposed by (Anthes, 1984). Given that human activities can alter convective precipitation patterns, Anthes (1984) proposed bands of vegetation could be planted in semiarid regions to alter the location and frequency of precipitation in these regions. It was hypothesized these bands of vegetation could increase convective precipitation through the generation of NCMCs, increase of low-level moist static energy, and through an increase in atmospheric water vapor. Irrigated farmland acreage increased throughout the 1980’s, and thus groundwater demand. Extended irrigation usage can also modify the soil moisture memory characteristics in the first several meters of the soil column, which in itself may alter precipitation patterns due to local moisture recycling (Orth & Seneviratne, 2012). Yang et al. 2019 ran a Weather Research and Forecasting model based study of the southern Great Plains focusing on the irrigation scheme in the region. The temporal distribution of irrigation rates was found to be peak in July. This is consistent with the findings of DeAnglis et al. (2010). The largest impact of irrigation within the Great Plains was found to be around the areas of intense irrigation agriculture, particularly of the center pivot type. The union of the Great Plains study area utilized by Yang et al. (2019) and the north plains study area
encompasses all of NE. Latent and sensible heat fluxes were found to increase by 46.7 Wm\(^{-2}\) and -33.4 Wm\(^{-2}\) respectively (average over 2 years). This was found to lead to a decrease in just the surface skin temperatures, limiting outgoing longwave energy, and an overall rise in surface available energy throughout the entire surface.

Changes in precipitation over the Great Plains have been investigated against the distribution of irrigated agriculture over time (Deangelis et al., 2010). For the years spanning 1949-1995 it was found the increase in irrigated land from the Ogallala Aquifer in the Great Plains had limited effects on larger-scale regional precipitation downwind of the irrigation during the month(s) with the highest irrigation rates (July and August). This was not found to be significant for all areas or months in which irrigation was occurring. Whereas there was a significant modification to precipitation rates over the larger areas of Illinois and Indiana, this did not manifest itself in feedback on smaller regional scales. Natural deviations in precipitation throughout the spatiotemporal limits of the study could also be responsible for the observed sign of the precipitation feedback mechanism (precipitation favored over dry areas versus precipitation favored over wet areas). This is only a hypothesis as Deangelis et al. (2010) were not able to validate results across several statistical tests. Irrigated lands may provide strong regional moisture feedback, but this is under mesoscale and synoptic forcing, which limit which months this phenomenon may flip to a significant positive signal (Deangelis et al., 2010).

Findell & Eltahiar (1997) leveraged the Illinois Climate Network, the most complete set of soil moisture data available for the Midwest at the time. Correlations were established between surface soil moisture and subsequent rainfall for short climatological periods (21 days). The resulting positive correlations between the surface moisture and rainfall potential were consistent with previous research (Pan et al., 1996), where surface soil moisture maintains wet
(dry) conditions for a region. Synoptic/global atmospheric forcing (i.e. sea surface temperatures) and local feedback was considered as the primary driver for these observed effects. Synoptic/global atmospheric forcing was found to correlate with precipitation more closely over the autumn and winter months, whereas local soil moisture/precipitation feedback more closely aligned with the spring and summer months. However, the interactions of atmospheric forcing and local feedback were not ruled out in the above studies.

Entekhabi et al. (1996) examined the soil moisture controls on the thermal inertia of the surface in the Great Plains. Moisture contained within the soil column breaks down some of the dynamics of weather (climate). They established phase lag correlation within the soil column and the impedance of high frequency fluctuations as a function of depth in the soil column. Furthermore, horizontal soil moisture anomalies create sharp gradients in sensible/latent heat fluxes. This can generate a PA, which under favorable micro and synoptic circumstances can manifest as a NCMC. Entekhabi et al. (1996) established spatial variation of this phenomenon may trigger the growth of squall lines and baroclinic disruptions. This research is performed with data of limited spatiotemporal resolution. The authors acknowledge this as a limitation of the data collection technology at the time of writing. The effect of soil moisture anomalies may have at differing scales are consistent with the earlier findings (Segal & Arritt, 1992). Data collection schemes and analysis need to account for potential mismatch in spatiotemporal resolution (Robock et al., 2000).

Sun and Ortega (1979) observed pre-storm convergence is affected by the presence of a horizontal heat gradient. Fast & McCorcle (1991) examined the surface soil moisture effects on the structure of the planetary boundary layer. Ziegler et al. (1995) correlated the presence of gradients of sensible heat flux formation along drylines (magnitude of 100 W m$^{-2}$ 50 km$^{-1}$).
Lanicci et al. (1987) found the spatial gradients in soil moisture as contributing factors to storm formation in the Great Plains.

2.3 Convective Initiation - As Aligned to Underlying Boundary Layer Flow

Convective initiation and the relation to underlying soil moisture is often not aligned to the mean flow of the boundary layer. Often the spatial scale of the data is coarse enough that the orientation is of little concern. However, the increase in spatiotemporal resolution in the years since many of these studies were conducted now allows for such alignment. Taylor (2015) examined points of convection aligned with the underlying low level flow (Dee et al., 2011; Taylor et al., 2011). Infrared imagery was used for identification of convective initiation from cold cloud tops over Europe for 2004-2013 (months of April-September). Soil temperature and the derived soil moisture data products from the Advanced Scatterometer (ASCAT) were utilized. From the points identified as perturbed areas, underlying soil temperature and moisture data were queried, gradients computed, and composite means were taken. It was found that deep convection formed preferentially over areas with positive along-wind soil moisture gradients. Convective initiations occurred to an upper bound of 78mm rainfall in the previous month (Taylor, 2015). The resulting convective initiation points were split into high windspeed (≥3.0 m s\(^{-1}\)) and low wind speeds (<3.0 m s\(^{-1}\)); higher wind speeds have a greater effect on the disruption of PBL temperature profiles and disrupt NCMC. As expected, within the cases of the strongest background flow, the center of convection moved downwind and over a wetter surface. This occurred at a faster rate than with the lighter winds, and the lightest winds (<1.5 m s\(^{-1}\)) saw instances of upwind propagation from the point of convective initiation.
2.4 Regional Thunderstorm Climatology

Regional climatology of thunderstorms exclusive to the north plains are rather limited. Mohee and Miller (2010) established a generalized climatology of ND thunderstorms though the analysis WSR-88D NEXRAD data and surface observations for the years 2002-2010. This research was performed to identify at-risk times from a public safety perspective and that of infrastructure vulnerability. For this time period the temporal bounds of cell generation/propagation for ND were computed. These results were similar to other spatial-temporal analyses of thunderstorms over the Great Plains. No differentiation between types of deep convection were made. Mohee and Miller (2010) did not incorporate any surface characteristics, instead the focus was on the vulnerability of critical underlying infrastructure to the diurnal and seasonal distribution of thunderstorms.

Studies on the distribution of storms over large portions of the Great Plains are more common. These studies tend to focus on great plains bound by northern Texas to Nebraska. The reasons for this seem to be twofold: 1) higher frequency, especially of intense, storm systems—such as those developing along Oklahoma/Texas drylines 2) more favorable distribution of sensing platforms – the Oklahoma Mesonet is particularly dense. Larger storm systems are more common in this region as well, with meso-α-scale (domains of horizontal scale of 200-2000km) systems, such as mesoscale convective complexes occurring throughout warm season months. Regional soil moisture feedback and source tracing studies typically take place within the southern Great Plains domain as well. This may be due to sensing coverage (state mesonets, Tropical Rainfall Measuring Mission products) and the minimum area required to confidently perform this research.
The north plains region is covered in part as a sub region for decades of storm system distribution and tracking studies, both with NEXRAD data and infrared data (Haberlie & Ashley, 2019; Jirak et al., 2003; Maddox, 1980). Whereas these studies do not normally tie the north plains physiography into the discussion of storm spatiotemporal distribution, they are often linked to prevalent phenomena in the region (i.e. distribution of center-pivot irrigation, precipitation feedback regimes). The lack of studies addressing the north plains region exclusively remains a gap in the current research.

2.5 Summary

Feedback of soil moisture operates on a variety of spatiotemporal scales within the Great Plains of North America. This can occur for a variety of reasons from synoptic forcing to smaller meso-β-scale circulations. The surface flux (sensible and latent heat) distribution play an important role in the development of NCMCs, however spatial heterogeneity of the surface does not guarantee convective initiation from these circulations. The synoptic conditions dictate the chances of convective initiation occurring, as the background flow may disrupt the development of NCMCs. In cases with limited background flow these circulations can produce training/back-building convection over a local area, releasing large amounts of moisture into unstable atmospheric environments. This may lead to severe local weather hazards, such as hail, non-tornadic winds, and flash floods.

This study focused on the alignment of soil moisture heterogeneities to the background flow and adds to the body research of storm localization. This was performed within a study area that has few thunderstorm climatological studies specific to the physiography of the region. This analysis was performed at a finer spatial scale than research from prior decades. Spatial and temporal domains from previous studies were used to establish general bounds of the study area.
and provided a baseline for the examination of spatial distributions. Existing techniques in the relation of convective initiation to the soil moisture surface were leveraged, as well as the application of some concepts outside of the current literature of this subject area.
CHAPTER III
METHODOLOGY

The methodology employed in this study is organized into four sections: 1) Data selection and filtering – which encompasses availability, spatiotemporal resolution, and quality control, 2) processing and region of interest extraction – covering the localization and calculation of descriptive statistics for each case 3) data relationships and distributions – methods employed for the quantification of spatial and temporal patterns 4) error sources, error propagation, error management – which endeavors identify and partially quantify both inherited error from the data sources as well as the error introduced by the filtering methods herein.

The methodology developed for this research was highly integrated into the codebase for data processing. To aid in the speed of interpretation of the codebase several flowcharts outlining high level process of the codebase are provided. These flowcharts are not intended as an outline for the development of the algorithms employed nor do they provide insight to internal class structure. These are available in Appendices A and B.

3.1 Data Selection and Filtering

This study examines several inputs of physical mechanisms acting on the PBL and extending into the upper layers of the troposphere. The temporal scale of these variables is of great importance, as the varied characteristics of the surface, PBL, and mid-layers of the upper troposphere are subject to rapid change. This study does not attempt to model these interactions, instead only to establish the magnitude of the relationship. This is accomplished through the correlation of surface model data (National Aeronautics and Space Administration (NASA) SPoRT LIS - relative surface soil moisture (0-10cm)), standard 10 m wind analysis data (Global Forecast System 4 (GFS4) wind components), and reflectivity to identify points of strong
convection (Weather Sensing Radar, 1988, Doppler - Next Generation Radar, Level II (hereafter referred to as (WSR-88D) NEXRAD Level II)), and climatological means (PRISM). Dates of interest are a subset of that generated by Haberlie & Ashley (2019). Spatiotemporal availability of these datasets is displayed in Figure 3.1 Each of these datasets utilized in this study will be further discussed in the following subsections.

<table>
<thead>
<tr>
<th>datset</th>
<th>Spatial Resolution (m)</th>
<th>Temporal Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentinel</td>
<td>1x10^3</td>
<td>Minute</td>
</tr>
<tr>
<td>Landsat</td>
<td>1x10^3</td>
<td>Hour</td>
</tr>
<tr>
<td>SPoRT LIS</td>
<td>1x10^3</td>
<td>Day</td>
</tr>
<tr>
<td>ASCAT</td>
<td>1x10^3</td>
<td>Week</td>
</tr>
<tr>
<td>WSR-88D</td>
<td>1x10^3</td>
<td>Month</td>
</tr>
<tr>
<td>NEXRAD</td>
<td>1x10^3</td>
<td></td>
</tr>
<tr>
<td>GOES IR</td>
<td>1x10^3</td>
<td></td>
</tr>
<tr>
<td>GFS 4</td>
<td>1x10^6</td>
<td></td>
</tr>
</tbody>
</table>

![Figure 3.1](image)

*Figure 3.1 – Approximate spatiotemporal resolutions of data used and other common datasets*

3.1.1 NASA SPoRT Land Information System Data

The SPoRT LIS is a continuous integration of the Noah land surface model at the 3km resolution. It was initialized with uniform conditions over the contiguous United States where it was run with forcing from atmospheric analyses for an extended “Spin-up” (Case, 2016; Case & Zavodsky, 2018). SPoRT LIS is available in near-real-time, as atmospheric analyses become
available as input forcings. SPoRT LIS contains soil moisture data at a variety of levels (relative and volumetric), vegetation greenness fraction, surface thermal data, precipitation estimates among other data.

3.1.2 Global Forecast System 4 Analysis Data

The GFS4 data utilized in this study has a gridded spatial resolution of 0.5° x 0.5°. The spatial resolution is coarser than that of the North American Mesoscale model (NAM) (12km x 12km (~0.1° x 0.1°)). The use of a global model is not strictly necessary for this research, nor is the spatiotemporal resolution ideal. A higher resolution dataset would be preferred (such as NAM), however GFS4 was utilized for its simplicity of integration into the workflow. Too fine of resolution could confound the results as these data are used exclusively for general orientation of the underlying SPoRT LIS surface soil moisture (SSM) data to the background PBL flow. Nonclassical mesoscale circulation flow could be in direct opposition to the representative background flow offered by GFS4 due to the scale discrepancy.

The standard 10 m wind components from the GFS4 data were used to determine the low level PBL flow for the orientation of the underlying soil moisture data. These orientation data were used to transform the underlying data relative to the wind vectors such that “upwind” is always in the same location in the datasets for further analyses. Figure 3.2 illustrates the pose shift of these underlying data.

The 850hPa surface from GFS4 was used to determine the presence (lack) of the GPLLJ. This was done manually following the process laid out in previous research (Bonner, 1968; Frye & Mote, 2010). The surface was examined for flow originating from the Gulf of Mexico reaching speeds of 25kt (12m/s). If the 850hPa flow met these criteria and was over the convective initiation point the GPLLJ was presumed to exist for that datetime & location.
3.1.3 WSR-88D NEXRAD Level II Data

NEXRAD Level II Data were leveraged for the acute temporal coverage provided by the collection schema. NEXRAD Level II data utilized in this study are of a finer spatial scale than required, availability constraints dictated their use. These data have no broadly collected contemporary.

3.1.4 Haberlie and Ashley MCS Database

This study determined an initial set of dates with the potential to contain the phenomena under study as needed. Haberlie and Ashley (2019) identified and characterized MCS from National Operational Weather Radar data and Stage IV hourly precipitation analysis within several study areas across the United States. The north plains study area as defined by Haberlie and Ashley (2019) is partially contained by the study area employed by this research. Dates in

Figure 3.2 – Alignment surface soil moisture (SPoRT LIS RSM 0-10cm) to the direction of the surface wind background flow (GFS4 10m components).
which MCS activity within the north plains study are were obtained from this research. This initial dataset consisted of 370 datetimes encompassing 2007 through 2017. The datetimes comprising these data had a temporal resolution of 15min, identifying the time at which the criterion of MCS were met (Haberlie & Ashley, 2019).

3.1.5 Selection of Test Dates

Leveraging the list of datetimes for the study area provide by Haberlie and Ashley (2019). A search for any back-building systems (as defined by Bluestein & Jain (1985)) was performed via a three tiered filtering process.

Initial filtering: The times of events ad provided by Haberlie and Ashley were imported and NEXRAD mosaic loops covering a 12 hour period from -4hr to +8hr around the timestamp with a 30min interval were developed (archived data: http://www2.mmm.ucar.edu/imagearchive1/RadarComposites/). Each loop was examined for the presence of training thunderstorm behavior. The filtering process at this step was performed to err towards commission, as the temporal and spatial resolution at this stage of analysis were coarse. The decision to include an event at this stage was made under the refinement steps shown in Figure 3.3. The results from this initial thinning transformed the dataset from 370 candidate datetimes to 69 datetimes containing one or more instances of suspected thunderstorm training behavior.

Secondary filtering: After successful thinning of the initial dataset from Haberlie and Ashley, a more rigorous thinning was performed facilitation expansion to specific cases of the training phenomena was performed. This mirrored the selection methodology performed prior and implemented the following: 1) A more restrictive refinement step describing back-building
behavior (Figure 3.3, secondary refinement), 2) higher temporal resolution data, 3) exclusion of events within areas known for orthographic uplift (Black Hills, SD). NEXRAD Level III mosaics archived from the Iowa State University archive were examined to establish finer temporal (5 min) and spatial resolution attributes for each of the remaining systems via the online viewing application (https://mesonet.agron.iastate.edu/current/mcview.phtml). Systems determined to be non-back-building at the finer resolutions were removed. As before, if the nature of the developing system was found to indeterminate type but displaying training characteristics it was included. The level of spatiotemporal resolution at this stage was able to resolve several cases that exhibited multiple back-building systems occurring simultaneously. These were only included as separate cases in the resulting dataset if there was clear delineation between the two systems as defined by the decision tree used for this level of filtering. Several systems were suspect to orthographic forcing in the Black Hills region of SD, however, were not occurring directly over the Black Hills. These were denoted as suspect to orthographic forcing and were carried through to the next filtering step. At this stage, the dataset still contained 69 cases, this included several dates a multiplicity of cases > 1.
Figure 3.3 – Refinement and identification of convective initiation locations
Tertiary Filtering: Reflectivity characteristics up unto this point in filtering were not considered. There was some intrinsic selection bias in the previous two steps which erred towards events with reflectivity $<$35dBZ. An initial reflectivity threshold of $\geq$35dBZ was utilized hereafter to reduce noise and limit the convective elements (Frye & Mote, 2010). After applying the reflectivity threshold and excluding events occurring within 1.0° latitude/longitude (~111.0 km) radius of the Black Hills region (as tagged earlier). The final dataset contained 40 unique cases. The spatiotemporal distribution of these data is outlined in Figure 3.4. A single initiation point fell outside the bounds of the study area, however the main body of the event occurred within the study area (eastern Wyoming). Due to the limited number of test cases this event was included.

3.1.6 Data Availability

NEXRAD Level III data are available from 2010-present through several web servers at the time of writing. NEXRAD Level II data are available through the same channels, archived from the initialization of the station to the present. For this research NEXRAD Level II data were utilized over Level III based on availability. Due to the large number of files, these data were streamed ad hoc from the Amazon Web Services archive (https://s3.amazonaws.com/noaa-nexrad-level2/index.html). This presented a bottleneck in processing time, especially when performing sensitivity analyses over NEXRAD data (see section 3.4.3 for information on sensitivity analyses), however this is primarily due to a lack of optimization within downloading files and caching.

Spatial coverage of the NEXRAD network is not complete over the study area, especially below 10,000ft above ground level, see Figure 3.5. Volumetric scans are performed as a series of
Figure 3.4 – Spatiotemporal distributions of identified test cases
Figure 3.5 - WSR-88D NEXRAD coverage of study area at a minimum of 10,000 ft above ground level
radial sweeps from each sensor. Of these sweeps the lowest in elevation (~0.5°) was taken as the representative sample from the appropriate site. If these data were not available (i.e. sensor was offline, or the data were otherwise not available within the Amazon Web Services archive), the next closest sensor was utilized. This was determined by the great circle distance between the initiation point and all sensors providing coverage within the study area. This process was repeated until the subset of sensors providing coverage to convective initiation point was exhausted, where after the event would be discarded. NEXRAD Level II data do not have complete spatial coverage of the study area with sweeps close to the surface as a result of Earth’s curvature and topography, the coverage to 10,000 ft above ground level can be seen in Figure 3.5.

SPoRT LIS data were obtained for the list of dates. These data have a temporal resolution of 1 hour from 2015 to present, 12 hour resolution from 2011 through 2014, and 24 hour resolution from 2007 through 2010. The difference in the temporal resolution posed some issues when relating datetimes between datasets. SPoRT LIS data utilized always errs towards pre-storm conditions.

GFS4 data (0.5° x 0.5°) are available from 2007 to the present, earlier dates are available as GFS3 data at the coarser resolution of (1° x 1°) from 2004-03-02 to the present. This limitation of GFS4 availability is the primary constraint on the study period. The range of data available to serve the alignment function could be extended with NAM, which has data available from 2004-03-03 to the present. GFS4 analysis data are available in six-hour increments (0000z, 0600z, 1200z, 1800z). Temporal mismatches between GFS4 and other datasets (primarily NEXRAD data) when performing analysis were resolved by obtaining GFS4 data at the nearest neighbor datetime.
PRISM data were only utilized for the annual climatological means (1981-2010) (PRISM, 2019). These data are available at ~4km resolution and ~800m resolution. The 4km variant was utilized in this study. As these data were a baseline, there was no temporal mismatch that needed consideration. PRISM data are available through https://prism.oregonstate.edu.

3.1.7 Simplification Assumptions for Input Data

All input data sources were unprojected gridded data. Due to the relative ease of working with these data in this format all calculations were conducted without reprojecting these data. Maximum distances involved in any of the calculations were two decimal degrees (between 40°N and 49°N). One degree longitude at 40°N is ~85km whereas it is ~73 km at 49°N. The net result of this discrepancy in the context of this work in areas will be overestimated the further North the event takes place.

3.2 Data Processing and Region of Interest Extraction

NEXRAD and SPoRT LIS data are examined relative to the convective initiation point. The area local to the convective initiation point was extracted for each to facilitate inter-case comparisons and averages. Algorithms were developed for the extraction and derivation of metrics describing the LIS SSM surface. The resultant NEXRAD reflectivity is localized to each case. This process was performed on gridded, unprojected, data. The effects of geographic distortion were mitigated within the derived data but were not eliminated.
3.2.1 Analysis of WSR-88D NEXRAD Data

All code written for the analysis of NEXRAD Level II reflectivity data for the subsequent subsection was written in Python 3.7.1 and is available online at:
https://github.com/Skyehawk/LIS_Extract. Flowcharts dictating the high level structure of these algorithms are also available in Appendix A.

The latitude and longitude of each case were then determined based on the Refinement flowchart (Figure 3.2). Each case had the location of convection defined as the western-most edge of reflectivity ≥35dBZ and was the first timestep to contain returns ≥35dBZ. This location was then further refined by downloading these NEXRAD Level II data for the closest NEXRAD site and adjusting to the same criterion by hand for t=0min. For each case, all scans bound by t=0min to t=120min (inclusive) were obtained. The number of scans varied for each case, as NEXRAD scans are not synchronous.

In addition to the latitude & longitude of the convection, the propagation vector of the thunderstorm training was determined. This process was accomplished manually for each system, by defining a downrange polygon (region of interest (ROI)), fit to encompass the entirety of the lateral (crosswind) development of the back-building system up to (and including) 0.875° downwind (see Figure 3.6 for example ROI encapsulation polygon). Practical limits on the size of the ROI include, 1) the area encompassed by the training storms (upwind, crosswind, & downwind, 2) the chance of encroaching onto neighboring convective elements, 3) any non-linear training of the storms. The alignment of the ROI was performed by hand and based on the best fit between t=0min and t+120mins. The alignment vector applied in this step was not the same to that used in the alignment of the SPoRT LIS surface soil moisture characteristics to the background flow described in the next subsection. The alignment was done by hand due to two
factors: 1) lack of consistent data describing the upper-level flow (not all cases aligned with a GFS4 analysis timestep), 2) the occasional inclusion of encroaching cells that developed independently from the back-building structure, but were near enough to pass into the ROI. With sufficient data the back-building propagation vector alignment could be automated, however this would be beyond the scope of this work.

The ROI was chosen to approximate the scale at which NCMCs operate (Segal & Arritt, 1992). A trapezoid was chosen with the height axis parallel to the direction of storm movement. With the convective initiation point at (0,0) the coordinates of the ROI were: (0.8750, 0.25), (0.8750, 0.25), (-0.125, -0.125), (-0.125, 0.125), (0.8750, 0.25). In practice, the ROI could be any

Figure 3.6 - Region of Interest polygon masking underlying NEXRAD Level II Data
a trapezoid was chosen due to the increasing cross section as the storm(s) moved downwind from the convective initiation point and the limited number of vertexes. ROI proportions were estimated based on the observed storm spatial growth rate and storm track straight line tendency.

Each localized case provided a time series of reflectivity and coverage of returns ≥35dBZ. Each case had a 15min smoothing window applied to reduce noise, a temporal window was chosen over a specified number of samples due to the differences in scanning rates of WSR-88D NEXRAD sensors. Fifteen minutes was chosen as under most circumstances it will contain at least two scans but is still within the lifespan of training/back-building behavior. The rate of change (build-up rate) of mean reflectivity (Ref_Slope) and coverage ≥35dBZ (Area_Slope) were determined for each case across the observation period over the smoothed time series. The t+120min observation window was longer than the lifespan of most individual thunderstorm cells in these systems. These cells would often slowly move outside of the ROI due to upper level flow and interaction with other mesoscale features, conversely, surrounding systems would occasionally move into the ROI. To determine the build-up rates algorithmically, local minimums and maximums were calculated for each case on the smoothed data. The first and second local extrema from after t=15min were taken to compute a linear growth (reduction) rate, i.e. Figure 3.7). The maximum value of coverage ≥35dBZ (Max_Area) and reflectivity ≥35dBZ (Max_Ref) was taken for each time series. Maximum observed reflectivity was capped at 55dBZ (at an elevation of 0.5°) , as the observation of reflectivity of hail were a concern at this level, this was a rough approximation where hail could influence reflectivity, as the actual detection algorithms used by the NWS incorporate several parameters. However, for the test cases used in this study, Max_Ref never exceeded 50dBZ.

43
A composite time series was created and then smoothed following the prior method. The product of the normalized area and normalized coefficient of variation of reflectivity was computed across the time series. The linear growth rate of this metric was then computed, similar to that of the area growth. Descriptions of these variables are available in Appendix C.

3.2.2 Analysis of NASA SPoRT LIS Relative Soil Moisture Data

All code written for the analysis of LIS data for the subsequent subsection was written in Python 3.7.1 and is available online at: https://github.com/Skyehawk/LIS_Extract. Flowcharts dictating the high level structure of these algorithms are also available in Appendix B.

Analysis of the NEXRAD Level II data provided the precise latitude, longitude, and temporal information on the convective components. These data were used in conjunction with the closest pre-event set of SPoRT LIS data. For those years where SPoRT LIS was available on
an hourly basis (2015-present) the analysis was also conducted on the hour preceding rainfall over the convective initiation point. The selection of these LIS datetimes does not exclude instances where precipitation may have fallen within the previous 24hrs not associated with the convective initiation. The years 2011-2014 have SPoRT LIS data available for 0000z and 1200z, for these dates the nearest 12hr run preceding rainfall associated with the convective initiation was utilized. 2007-2010 only have data available for 0000z, the nearest available data preceding rainfall associated with convective initiation was utilized.

Alignment of the gridded SPoRT LIS data based on the surface wind vectors was performed in two steps. The high level process was as follows: 1) extraction of the underlying information was performed via a set of points defining a polygon (a square of side length 2.0° was utilized in this study, although any polygon could theoretically be used). This polygon was sized following Taylor 2015. The polygon was then rotated and translated to be centered on the latitude/longitude of the point of convection. A masked array was then constructed to exclude all points not falling within the defined polygon, but still within the orthogonal bounding rectangle of the region of interest. 2) These data were then translated and rotated to place them on the origin.

Transformation of these data was processed by two algorithms, two dimensional transformation matrices (transformations of the input polygon for masking), and the Scikit Image package. The Scikit approach of image rotation was chosen over the direct application of rotation via a transformation matrix for the resampling algorithms built into the Scikit package. The nearest neighbor resampling scheme was utilized to preserve the distribution of the underlying data.
SPoRT LIS soil moisture data was used to compute several variables. All computations follow an identical methodology with varying inputs, these are outlined in Appendix C. For several variables, a skewed two-dimensional weighted average was applied, (size: 2x2, $\sigma_x = 1.0$, $\sigma_y = 2.0$) (see Figure 3.8 for a visualization of the distribution used for the weighted variants of derived variables). Each case had the raw relative soil moisture descriptive statistics computed, including the weighted mean, and relative location of the minimum and maximum values. The same statistics were computed for the following derived data: mean departures, the first derivative of surface, ratios between upwind to downwind means, as well as differences between upwind and downwind means. These data were output in the ASCII format for later analysis in ArcMap 10.7.1 (see section 3.3). Smoothing of these data was not performed, as the soil moisture surface never exhibited a high degree of noise. This was determined via visual inspection and scaling of these data. If smoothing were deemed to be necessary, a bilateral blur

![Bivariate gaussian distribution applied in weighted calculations](image_url)

*Figure 3.8 – Bivariate gaussian distribution applied in weighted calculations.*
would have been leveraged to preserve the significant transitions of soil moisture by space and value.

The raw SPoRT LIS SSM and derived data were plotted to facilitate visual identification of patterns and as an error mitigation check to ensure the data were extracted correctly. The averages along the primary axes were also computed (row wise and column-wise averages), these correspond to the upwind/downwind averages and the crosswind averages. Additionally, the magnitude of the soil moisture gradient was also derived at the native scale of the data without smoothing (Figure 3.9).

*Figure 3.9 – Upwind versus crosswind means of a LIS SSM surface.*
3.3 Data Relationships and Distributions

The multidimensional aspects of these data necessitate several analysis techniques. These include the quantification of relationships between variables, within single variables, the influence of some variables upon others, and the spatiotemporal distributions. The purpose of the following techniques is based on the identification of relationships between variables, within variables, and describe how these data are spatially distributed. These techniques are employed to provide a manner means to answer:

1) What are the general characteristics of surface soil moisture local to points of convective initiation exhibiting back-building/training behavior within the selected test cases?

2) Is the spatial distribution of these events the result of random clustering? Do hotspots of convective initiation exist?

Correlations between independent and dependent variables (Pearson’s R and point Biserial) were used in conjunction with the analysis of covariance (ANCOVA) to answer the first research question. Whereas it is nearly impossible for a single independent variable to control for a dependent variable some weak correlations may exist. To evaluate the spatial parameters of the data quadrat analysis was performed, as well as frequency analysis based on climatological annual precipitation means. Finally, the Getis-Ord Gi* statistic was used to check for broad clustering a frequency analysis would have not been able to detect. All spatial statistics were computed across grids of varying scales.

3.3.1 Intra-Variable Relationships

Several variables derived from these NEXRAD Level II data are temporal in nature. The area and reflectivity ≥35dBZ and derived variation are subject to autocorrelation within the
signals. However, within the fluctuating time periods over which these variables are calculated in conjunction with the varying polling rate of the NEXRAD sweeps between stations, poses an issue: these data may not have sufficient resolution to reliably detect these signals. This has potential ramifications in the calculation of storm build-up rates for the respective variables. Furthermore, since the ROIs for each storm are spatially fixed, errors of omission and commission may obscure any auto correlation signal.

Collected interval and ratio level data were found to be following a normal distribution in most cases. This was validated through the use of q-q plots with a reference line between the first and third quantiles (Cleveland, 1993). A Pearson’s r coefficient matrix was computed across all independent/dependent variable pairs. The GPLLJ was a dichotomous categorical variable, all other variables were continuous. This made the correlations between the GPLLJ and the dependent variables point bi-serial. Correlations between independent and dependent variable were used to answer the first research question:

3.3.2 Analysis of Covariance

The presence (lack) of the GPLLJ has been observed as a significant factor in the development of convective systems in the northern plains (Frye & Mote, 2010). To test for the significance of this source and the local contribution of moisture in the development of cells ANCOVA was conducted across all independent continuous variable and GPLLJ pairs versus the dependent variables. This analysis was done with Python 3.7.1 utilizing the Scipy.Stats and the Statsmodels packages.

3.3.3 Spatial Distribution of Cases

The mean surface soil moisture surface was calculated in ESRI ArcMap 10.7.1. The soil moisture data as aligned to the background flow, as part of the initial analysis of the SPoRT LIS
data, were exported for the region of interest. These data were assembled as a raster stack, optionally queried by date or presence of GPLLLJ, then averaged across all remaining layers. Focal statistics (mean, circular kernel (radius 3)) were then computed across the surface.

Evaluation of the spatial relationship was conducted in ESRI ArcMap 10.7.1. The average nearest neighbor analysis would not return meaningful results as the number of samples was limited to 40, to combat this, quadrat analysis was performed. The binning of the data was conducted at multiple scales when developing the spatial climatology from these data (see section 3.4.3 for details concerning grid scales). For simplicity, the convective initiation points were used as the center point of convection, however it should be noted the effects of back-building systems can be observed far downwind of the convective initiation point, potentially entering several downwind bins. The Getis-Ord Gi* statistic was then computed for each of the bins at each scale (Euclidian distances, Inverse Distance of 0.0).

The frequency distribution of convective initiation points in a regular grid was calculated at multiple scales to manage the modifiable areal unit problem (MAUP) (see section 3.4.3 for details concerning grid scales). The total extent of the grid and the study area were identical, with each cell of the respective grid being equivalent in area. A random distribution of points was generated for the study area and binned with the same grid. This random distribution run over 100 iterations and then the mean frequency distribution was taken. The observed distribution and the generated distribution were then evaluated with the Chi-Squared test.

A similar frequency distribution was conducted using polygons generated from the PRISM annual precipitation data (climatological period 1981-2010). The PRISM data were masked to the study area, reclassed via 1) Jenks natural breaks, standard deviation (1 standard deviation bins, 3) quantiles. The class breaks can be seen in Table 3.1. These raster data were
then converted to multipart polygons, the expected frequency distribution was generated (mean of frequencies of 100 iterations of random point generation), and the observed distribution was computed.

Table 3.1: Class breaks for PRISM Mean Annual Precipitation (liquid equivalent), climatological period: 1981-2010

<table>
<thead>
<tr>
<th>Classification</th>
<th>Annual Mean Precipitation (mm)</th>
<th>Area (km²)</th>
<th>Convective Initiation Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jenks</td>
<td>336 – 420</td>
<td>91074.8</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>421 – 481</td>
<td>166577.5</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>482 – 549</td>
<td>126634.8</td>
<td>6</td>
</tr>
<tr>
<td>Natural Breaks</td>
<td>550 – 617</td>
<td>105411.2</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>618 – 665</td>
<td>105222.5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>666 – 767</td>
<td>63256.6</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>768 – 916</td>
<td>36293.8</td>
<td>0</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>336 – 367</td>
<td>12256.2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>367 – 484</td>
<td>255257.5</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>485 – 602</td>
<td>207599.3</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>603 – 719</td>
<td>152064.4</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>720 – 836</td>
<td>60152.3</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>837 – 916</td>
<td>7323.0</td>
<td>0</td>
</tr>
<tr>
<td>Quantiles</td>
<td>336 – 429</td>
<td>113737.5</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>430 – 465</td>
<td>115150.0</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>466 – 524</td>
<td>114667.5</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>525 – 595</td>
<td>115550.9</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>596 – 667</td>
<td>115968.4</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>668 – 916</td>
<td>119212.2</td>
<td>6</td>
</tr>
</tbody>
</table>

3.4 Error Sources, Error Propagation, Error Management

Error throughout these analyses was of constant concern. Steps were taken to identify probable sources of error, how these may influence downstream calculations, and how to report results within the context of the accumulated error. Due to the small sample size, this posed issues, particularly when selecting statistical tests.
3.4.1 Error Sources

This study inherits all biases and errors from the source datasets. There was little that can be done in terms of correction in these cases, as the original source and magnitude are often unknown. In some instances (such as incomplete data) corrective action could have only been performed there was a suitable auxiliary dataset to pull the required data from without inducing additional error.

Error and biases produced internally within this study’s methodology and interpretation of results are perhaps easier to correct but induces additional complications. To the best of the researcher’s knowledge and abilities this error has been mitigated such that the results are robust to small variation.

3.4.2 Error Propagation

The methodology employed in this research contained several serial processes. By this very nature, any induced error could quickly compound and drown any underlying signal. Many redundant/unnecessary processes were removed for the final implementation of the various algorithms, however, to maintain code readability, performance, and implementation time some of these processes needed to retain these inefficiencies.

3.4.3 Error Management

This study leveraged batch processing across all datasets to eliminate as much human-induced error as possible. However, the selection of test dates, creation of scripts, interpretation of results, etc. all involved human interaction and thus could contain errors. In many of these instances’ safeguards were put in place to minimize this risk. This included utilizing quality controlled data from reputable sources, multiple independent passes over the initial data for the
selection of test dates when data were analyzed by hand, decision trees where appropriate, checksums across gridded data, and a variety of data storage and file management schemas.

Data used in this analysis are of varying spatial scales. To limit the impact of MAUP on the analysis, GFS4 data was queried such that any derived GFS4 variables (i.e. mean 10m flow) were treated as a constant. GFS4 gridded data were never overlayed on finer temporal resolution data. When querying the GFS4 data, the variance of the wind vectors was examined to ensure the robustness of the generalized background flow. Other datasets are encapsulated such that direct spatial comparisons between datasets do not occur. In other instances where there are multiple layers of data, only those data under the same spatial scales were overlayed/compared.

To further address the effects of MAUP sensitivity analysis were performed for both the SPoRT LIS data sets and the NEXRAD datasets. This varied the positioning of the convective initiation (base point for each case) for the SPoRT LIS data, as well as the size of the bounding polygon utilized for NEXRAD Level II data. In the case of the SPoRT LIS analysis, error tended not to occur until greater than 0.25° of latitude/longitude from the convective initiation point. For the NEXRAD Level II data, the spatial scale of the ROI was commonly not inducing issues until ≥120% of the base polygon. In depth results of the sensitivity analyses can be found in Chapter IV – Results.

The sensitivity of the SPoRT LIS data to alignment was conducted by generating a regular 5x5 grid of points centered on the identified convective initiation point aligned North/South. This resulted in a maximum offset of ±0.25° in latitude and in longitude. Since there was no guarantee the observed point of convective initiation was directly above the anomaly, the extent of this analysis was determined to be larger than that of a single large thunderstorm cell. No cells approached that threshold within the first 30mins in any of the test
cases. Extraction and alignment to the background flow was then conducted. The calculation of
the background flow was obtained independently from the GFS4 Analysis data for each case.
The algorithm was then run for each point on the 5x5 grid, the results for each location were then
aggregated and examined for outliers. There were two common variables that showed more than
five instances of outliers within the 5x5 grid: 1) background flow, 2) difference in mean
departures between upwind and downwind. It was common for the background flow direction
and magnitude to shift dramatically; this was observed for 15 of the 40 cases. Of those 15, all but
one occurred on the edge or corner of the 5 x 5 grid of test locations. The center (3x3 grid) only
had one case where the background flow and magnitude shifted. The ratio of upwind to
downwind of the departures when weighted experienced nine instances of outliers all occurred
on an edge or corner of the 5x5 regular grid. (see Figure 3.8 on weighting)

The range of the sensitivity analysis for the NEXRAD Level II data was examined from a
100% to 125% scaling of the (ROI) bounding polygon in 1% steps. This resulted in the range of
area within the polygon covering a nominal range of 4,620km$^2$ to 5,775km$^2$ at 45°N. The
polygons were only scaled up as the underlying statistics were not robust to the presence of zeros
in the resultant time series for each percentage of the nominal value. This approach guaranteed
the data present in the 100% case were a subset of the scaled up cases. The results were then
aggregated and examined for outliers across each respective variable. The sensitivity of the
upscaled values was found to marginally deviate as the polygon grew. Seventeen of the forty
cases saw outliers on the large end of the spectrum (scale ≥120%) for one or more variables. Of
those seventeen cases, three cases were found to be more sensitive to the size of the polygon,
triggering anomalous conditions after an increase in polygon size at 101%, 105% and 115%. The
occurrence these anomalies, particularly at the larger end of the spectrum, can be attributed to neighboring thunderstorm activity entering the polygon as it was enlarged.

Considerations for MAUP were also made for the spatial distribution of the storms across the study area. The analysis was conducted at five separate scales across the study area. These were conducted on a regular grid encompassing the entire study area. These grids ranged in scale from ~0.5° x 0.5°, ~0.6° x 0.6°, ~0.75° x 0.75°, ~1.0° x 1.0°, ~1.5° x 1.5°. The distribution at ~0.5° x 0.5°, 0.6° x 0.6° was found to be too fine for further analysis, as only 1.3% of cells contained multiple convective initiation points.
CHAPTER IV

RESULTS

The results of this study are organized into five sections: 1) correlations, 2) analysis of covariance 3) quantification of intra-variable homogeneity across cases, 4) spatial distributions, 5) mean surface conditions. These sections roughly follow the order of methods described within Chapter III – Methodology from section 3.3 onward.

4.1 Correlations

The correlations found between the independent variables (GFS4, SPoRT LIS, presence of GPLLJ) and the dependent variables (NEXRAD) are addressed in this section. Significant Pearson’s r correlation coefficients (α=0.05) are shown in Figure 4.1. Definitions of variables are available in Appendix C. There exists four significant, albeit weak, correlations between these variables. The standard deviation of the aligned SPoRT LIS soil moisture (relative soil moisture (RSM), 0-10cm – hereafter referred to as LIS SSM) surfaces correlated with the maximum coverage (≥35dBZ) value observed in the NEXRAD data. This had a coefficient of 0.381. The difference between the sum of the upwind LIS SSM and that of the downwind portion of the surface soil moisture correlated with the maximum coverage (≥35dBZ) observed in the NEXRAD data. This had a coefficient of -0.349. The presence (lack) of the GPLLJ was denoted as either present (1) or not (0), thus was a point-biserial correlation. A point-biserial correlation was observed between the presence (lack) of the GPLLJ and the maximum coverage (≥35dBZ) value observed in the NEXRAD data.
There exist several other correlations at the $\alpha=0.10$ significance level. These are depicted in Figure 4.2. Of note is the presence of a positive Pearson’s $r$ correlation coefficient (0.304) between the upwind/downwind difference in the weighted departures. And a negative Pearson’s $r$ correlation coefficient (-0.311) between the difference of upwind mean departures and downwind mean departures. The remainder of the correlation coefficients at this significance level are $< 0.300$ in absolute magnitude.

Figure 4.1 – Correlation matrix between derived SPoRT LIS Soil Moisture (0-10cm), and presence of GPLLJ versus the derived values from NEXRAD data ($\alpha=0.05$)
4.2 Analysis of Covariance

The analysis of covariance (ANCOVA) between the presence (lack) of the GPLLJ, each remaining independent variable, and the dependent variables was calculated. The results of this can be seen in Figure 4.3. There were three instances of significance at the $\alpha = 0.05$ level. These occurred between: 1) Standard deviation of the LIS SSM surface and the maximum NEXRAD coverage ($\geq 35\text{dBZ}$) 2) difference of upwind versus downwind raw LIS SSM values, 3) difference between the mean departure of the upwind values versus the mean departure of the downwind values.

In this case, for ANCOVA, our null and alternate hypotheses were:

Null: *There was no significant effect of GPLLJ on <dependent variable> controlling for the <independent variable>.*
Alternate: *There was a significant effect of GPLLJ on* <dependent variable> *controlling for the* <independent variable>.*

Comparing Figures 4.1 and 4.3 the GPLLJ had a significant effect on both Max_Area vs. StdDev and on Max_Area vs. Diff_Mean.

![Figure 4.3 – Statistical significance of ANCOVA between the GPLLJ and the dependent variable, controlling for the independent variables (α = 0.05)](image)

The ANCOVA matrix was also calculated for a significance level of α = 0.10. This yielded the results seen in Figure 4.4. As with the previous matrices, comparing the results of the correlation matrix (Figure 4.2) and the ANCOVA matrix (Figure 4.4) none of the correlations can be decoupled from the GPLLJ with any statistical certainty.
4.3 Spatial Distributions

Quadrat analysis was performed at several scales to test for the spatial dispersion patterns of the convective initiation points utilizing the Chi-Squared statistic (see sections 3.3.3 and 3.4.3). This analysis did not find any statistically significant results indicating the spatial distribution of convective initiation points were anything other than random. The results of the quadrat analysis are available in Table 4.1.

Table 4.1 - Multi-scale quadrat analysis of convective initiation points

<table>
<thead>
<tr>
<th>Grid</th>
<th>Grid Cell Dimensions</th>
<th>Chi Sq. - p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 x 18</td>
<td>0.50° x 0.50°</td>
<td>0.7942</td>
</tr>
<tr>
<td>15 x 15</td>
<td>0.60° x 0.60°</td>
<td>0.9098</td>
</tr>
<tr>
<td>12 x 12</td>
<td>0.75° x 0.75°</td>
<td>1.0000</td>
</tr>
<tr>
<td>9 x 9</td>
<td>1.00° x 1.00°</td>
<td>0.8259</td>
</tr>
<tr>
<td>6 x 6</td>
<td>1.50° x 1.50°</td>
<td>0.1698</td>
</tr>
</tbody>
</table>

Figure 4.4 – Statistical significance of ANCOVA between the GPLJJ and the dependent variable, controlling for the independent variables (α = 0.10)
The spatial distribution of the convective initiation points can be compared against other known physical distributions. Precipitation has commonly been used as a proxy for potential available soil moisture. The study area was split based on PRISM precipitation climatological norms. Of the several classifications performed, only one, Jenks natural breaks was found to be statistically significant at the $\alpha = 0.10$ level. The results are available in Table 4.2 and in Figure 4.5.

*Table 4.2 – Statistical significance of PRISIM classifications*

<table>
<thead>
<tr>
<th>PRISM Classification</th>
<th>Number of classes</th>
<th>Mann – Whitney U</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jenks Natural Breaks</td>
<td>7</td>
<td>0.0794*</td>
</tr>
<tr>
<td>Standard Deviation (6)</td>
<td>6</td>
<td>0.2862</td>
</tr>
<tr>
<td>Quantiles</td>
<td>6</td>
<td>0.4680</td>
</tr>
</tbody>
</table>

* Statistically significant at $\alpha = 0.10$ level

The Getis-Ord Gi* statistic was then computed for each of the bins at each scale of the regular grids (see section 3.3.3) to identify hotspots of back-building convection within the study area. These can be observed in Figures 4.6 – 4.8. Hotspots were found at the $\alpha = 0.01$ confidence level; however, it should be noted that due to MAUP, the geographical locations and significance slightly vary.

Mean LIS SSM distributions relative to the point of convection for all events are shown in Figure 4.9. A subset of events occurring in the driest two PRISM Jenks natural brakes classifications (mean annual precipitation water equivalent: 365 – 483mm (~13.25 - 19.00 in) are shown in Figure 4.10. Another Subset of those events occurring over the peak months of June/July/August are shown in Figure 4.11 – Figure 4.13, respectively. Differences between the
surface of all events and those falling in the driest portions of the study area were ~7.8% drier, as would be expected.
Figure 4.5 – Precipitation polygon bins used for frequency distribution (left) and observed convective initiations by bin (right)
Figure 4.6 – Hotspot analysis (reg. rect. grid of 0.75° by 0.75°)

Figure 4.7 – Hotspot analysis (reg. rect. grid of 1.0° by 1.0°)
Figure 4.8 – Hotspot analysis (reg. rect. grid of 1.5° by 1.5°)
Figure 4.9 – Mean LIS SSM relative to the convective initiation point, all cases

Figure 4.10 – Mean LIS SSM relative to the convective initiation point, 13.25-19.00in precipitation, based on PRISM annual means
Figure 4.11 – Mean LIS SSM relative to the convective initiation point, June

Figure 4.12 – Mean LIS SSM relative to the convective initiation point, July
Figure 4.13 – Mean LIS SSM relative to the convective initiation point, August
CHAPTER V

DISCUSSION

This chapter is organized into three sections: 1) evaluation of methods employed in this study, 2) synopsis of results, 3) conclusions, including strategies that could be employed in future work to make the results of a study similar to this on more robust.

5.1 Methods Employed

This study utilized some methods derived from Taylor (2015). However, most methods, or at least the application thereof, have not been broadly applied to research utilizing similar data. Those data utilized in this study, processing algorithms, and methods of analysis are commonly leveraged in other applications (i.e. computer vision). The fusion of these sources and techniques is unique to this study.

5.1.1 Data Sources

Data utilized in this study were from actively managed and quality controlled sources, this does not guarantee an absence of error, however reasonable precautions in data sourcing, interpretation, manipulation, and storage were taken to limit the chances of error. All data are sourced are from US government agencies (National Oceanic and Atmospheric Administration (NOAA), NASA) or from accredited research institutions (Oregon State University). There are some resources not utilized in this study that could have been leveraged such as infrared imagery to identify the extent of cold cloud shields or state mesonet data.

5.1.2 Data Processing

Data processing techniques employed to localize the surface conditions and the storms were based on the author’s prior experience with computer vision algorithms. These techniques
were not initially developed for use with geographic data. This potentially induces some
distortion to the data for each dataset. LIS SSM data were the most susceptible for two reasons:
1) extracted data covered a grided area of ~2.0° x 2.0°. 2) data were subject to unprojected
geometric transformations (translation and rotation). This was done to optimize the development
time of the software within the time constraints of the study. Geometric transformations were not
applied to NEXRAD data, and calculations were carried out on a per-case basis; there was no
inter-comparisons of geographic NEXRAD data between cases. NEXRAD Level II data was
subject to the polling rate of the NEXRAD collection schema: resulting in temporal resolution
available for the development of the timeseries of the reflectivity.

5.1.3 Statistical Measures

Test cases in this study were limited. This presented an issue with the application of and
with the statistical power of any inferential statistical tests. Whereas care was taken to ensure the
integrity of the results, the low number of samples was taken as a constant overhead issue. Thus,
assumptions on distribution were kept to a minimum (especially when these data were split into
subsets). Underlying statistical distributions were analyzed. However, in many cases, the true
distribution of these data was unknown.

5.2 Results

Statistical tests were performed at multiple confidence intervals and different spatial
scales. Care was taken to limit the effects of MAUP on the results. Still, this cannot be truly
overcome, especially with the sparse dataset. For this reason, conclusions were drawn carefully,
as the limits of these input data do not provide vast amounts of statistical power.
5.2.1 Correlation Within the Context of the GPLLJ

Working under the assumption of normally distributed data (validated with q-q plots) a Parsons’s r correlation coefficient matrix was computed. Several statistically significant correlations (α = 0.05) were found between the LIS SSM derived variables and the dependent variables derived from NEXRAD data (Figure 4.1). These correlations were strongest between:

1) standard deviation of the moisture surface & the maximum observed NEXRAD coverage ≥ 35dBZ (Pearson’s r correlation coefficient of 0.381),
2) the difference between the upwind and downwind soil moisture values and the maximum observed NEXRAD coverage ≥ 35dBZ (Pearson’s r correlation coefficient of -0.349). When these data were evaluated significance level (α = 0.10) more significant correlations emerged (Figure 4.2).

Surface heterogeneities (high variance over the surface and/or differences between the upwind moisture versus downwind moisture) may be responsible for the generation of NCMCs, which in turn provide a pathway for moisture transport aloft. The greater the standard deviation of the surface corresponds with an uptick in maximum observed reflectivity over the lifetime of the training/back-building behavior. Similarly, the difference between upwind and downwind moisture negatively correlates with the maximum observed reflectivity values. This implies drier soils upwind of moister soils correspond to a greater observed maximum reflectivity for the cases evaluated in this study.

The GPLLJ is a significant source of moisture for the study area. The point bi-serial correlations were generated between the presence (lack) of the GPLLJ and the dependent variables. This resulted in statistically significant correlations between: 1) the GPLLJ and the maximum observed NEXRAD coverage ≥ 35dBZ, 2) the GPLLJ and a composite metric (rate of change in the time series of NEXRAD coverage ≥ 35dBZ scaled by the variance of the
reflectivity). These correlations are consistent with previous literature relating the GPLLJ to the generation and strength of convection during the warm season months, as well as a shift in GPLLJ related precipitation from more in May towards less in August (Bonner, 1968; Higgins et al., 1997). This is the result of the influence of the GPLLJ shifting south during the second half of the warm season, despite an uptick in moisture availability from the GPLLJ during warm seasons months. The second half of the warm season also sees the largest differences between observations and reanalysis data are found. The observed variance has been hypothesized to be the result of PBL interactions with the stability characteristics of higher atmospheric layers throughout the diurnal cycle (Higgins et al., 1997; Hitchcock & Schumacher, 2020). Such interactions between local convective circulation and the GPLLJ and how the stability parameters of the atmosphere prime the environment for the development of training/back-building precipitation remain an area of continued research.

The presence (lack) of the GPLLJ moisture source was not able to be decoupled from the correlations through analysis of covariance (ANCOVA) over the test cases. This study was unable to reject any covariate from being coupled to the presence (lack) of the GPLLJ:

1) There was a significant effect of the GPLLJ on the maximum NEXRAD coverage \( \geq 35\text{dBZ} \), when controlling for the standard deviation of the LIS SSM surface.

2) There was a significant effect of the GPLLJ on the maximum NEXRAD coverage \( \geq 35\text{dBZ} \), when controlling for the difference of the LIS SSM surface between upwind and downwind.

This leads to the conclusion for these data: the presence (lack) of the GPLLJ is the strongest single indicator for the potential of the convective elements. Under certain circumstances the surface conditions may lend themselves to the generation of strong convection
regardless of the GPLLJ, however the sample size employed in this research was not large enough to definitively identify any such influence generated from a single metric derived from the soil moisture surface. This is similar to the findings of Frye and Mote (2010).

ANCOVA was conducted for the GPLLJ and those correlations significant at the $\alpha = 0.10$ level. As at $\alpha = 0.050$, the GPLLJ was found to exert a significant effect on the respective dependent variables when controlling for the respective independent variables.

This does not rule out the interaction of surface conditions, initial boundary layer conditions, or upper level conditions on the development of NCMCs and convective elements. It is possible the conditions necessary for the development of back building convective elements can be forced by the surfaces conditions in the study area but were not observed in large enough quantities to yield statistically significant results.

5.2.2 Spatial Distributions

Quadrat analysis over the study area was performed at five discrete scales (Sections 3.3.4, 3.4.3, and 4.4). The limited number of datapoints set the lower bound of the spatial scale to ~0.75° x 0.75°, as scales with finer resolutions exhibited difficulty resolving multiple data points in cells. The result was quadrat analysis frequency distributions performed on grids of ~0.75° x 0.75°, ~1.0° x 1.0°, and ~1.5° x 1.5°. To discern if the observed frequency at a given scale was due to a random distribution a chi squared test was performed at each scale for the observed values and the expected frequency, none of these proved to be significantly different than a random distribution. The observed frequency was thus ascertained to be random across all tested scales. The observed random distribution is consistent with that found by Frye and Mote (2010).
Computation of the Getis-Ord Gi* statistic across the regular grids utilized in the frequency analysis resulted in clusters in two locations: southern NE and southwest ND. This was done primarily for data visualization, as the number of test cases was limited. Southern NE hotspots were not tightly clustered. The presence of increased center pivot irrigation throughout the region (Figure 1.3) may have influenced the frequency of back-building storms within the region, this would warrant further study. Southwest ND hotspots were clustered at multiple scales, this potentially indicated a hotspot between the soil surface and the presence of back-building behavior as a function of location (Figures 4.8 through 4.10). If the clustering of convective initiation points is a true positive signal, this is in opposition to the lack of clustering observed around Oklahoma City in similar work (Frye & Mote, 2010). However, the observed frequency distribution within the study area was found to be random, and thus could have been a false positive signal.

The spatial distribution of observed events was compared against the expected frequency for polygons derived from the climatological precipitation norms. The observed distribution of events was found a bias towards the events occurring in areas of lesser rainfall: ~365 – 483mm (~13.25 - 19.00 in). When grouped by precipitation means (trending from lesser in the northwest corner of the study area to greater in the southeast corner), the spatial distribution of back-building events is biased towards the western (drier) portion of the study area.

The mean surface of soil moisture as aligned to the background flow of the PBL at 10m was calculated. This was done for all events, events within the drier portion of the study area, and for each: events occurring in June, July, and August.

The difference in the mean surface between all events and that fell into the drier portion of the study area appeared to be the magnitude of the driest portion of the study area. As would
be expected the drier areas exhibited a greater lack of moisture at the driest points juxtaposed to the rest of the study area. This difference was found to be up to 7.8% less than the saturation value of the soil and occurred near the convective initiation point.

Differences between the mean LIS SSM surface for all events and those occurring in June resulted in less overall available moisture at the surface, ranging from 1.4% to 7.2%. The differences were less pronounced near the convective initiation point, and more extreme downwind. This pattern was inverted for August, with the mean surface having more moisture available overall (-2.0% - -11.7%). The convective initiation point was found to be drier overall in comparison to the mean of all cases. Downwind in August was observed to have the lowest available moisture (relative to mean moisture for the respective subset) (-11.7% RSM) when compared to June (7.2% RSM). The shift in preferential precipitation over surfaces with greater than mean moisture versus those with less than the mean surface may be linked to two factors: 1) shift in the surface state from an energy-limited regime to that of a moisture one, as this could influence the development characteristics of NCMC based on surface soil moisture and vegetation, 2) moisture supplied by the GPLLJ on June versus that in August.

5.3 Conclusions

This study explores some of the complex, multi-parameter interactions of soil moisture distribution and metrics of back-building behavior exhibited by storms in ND, SD, and NE. This study was conducted at a scale finer than that of previous forms of this analysis and over an area that has had limited analysis of this type. Additionally, this study examines the underlying surface conditions before and during storm development to examine the spatial distribution associated with changes in the surface energy budget as the land surface switches between an energy-limited regime to a moisture-limited one. Precipitation has been found to be favored over
both dry and wet surfaces in previous work. This study brings into question the conditions local to an event as well as the temporal element.

This study was severely limited by the number of test cases available. Future work may focus on developing a larger and more robust dataset through an exhaustive search through NEXRAD Level II data for an extended study period. Further modifications to the size and shape of regions of interest used to extract these NEXRAD Level II data, as well as attaching them relative to any storm movement (as opposed to fixed in space) may provide data that is more robust to the presence of other storms in the area.

Limitations aside, this study was able to answer the questions posed:

1) **What are the general characteristics of surface soil moisture local to points of convective initiation exhibiting back-building/training behavior within the selected test cases?**

This study found several weak correlations between the soil moisture surface and the behavior of the observed reflectivity. These correlations were strongest between the maximum area of observed reflectivity $\geq 35$dBZ and both: the standard deviation of the soil moisture surface and the difference in raw values between the surface soil moisture upwind versus downwind. However, this analysis was not able to decouple these correlations from the presence (lack) of outside influence of the GPLJJ.

The surface conditions at the point of convection relative to the background flow were extracted, subsets were extracted as necessary, averages were taken, and smoothing was performed. The resultant mean SSM surface evolved from June through August, a time period where the surface would likely transition from an energy-limited regime to a moisture-limited
one. Temporally, June saw the point of convection initiation occurring over a drier surface than that of August, although mean moisture within the ROI was higher in June versus August. Geographically, those events in the climatologically drier western portion of the study area were occurring over greater ranges of soil moisture, albeit overall moisture was lower. The precipitation preference toward drier or wetter soils in noted across literature, with some studies noting a preference of precipitation over drier soils (i.e. Taylor (2015)) and others showing preference or associating the enhanced evapotranspiration over wetter soils (Findell & Eltahir, 1997). The dichotomy of precipitation preference is well observed with some areas tending towards one or the other (Koster et al., 2004). Further still, Koster et al. (2009) established connections between this phenomenon and the evaporative regime of the local area around the convective initiation.

2) Is the spatial distribution of these convective initiations random within the study area? Are they more common in certain location as opposed to others?

Frequency analysis performed at multiple scales suggests two hotspots may exist. One in south central NE, this may be explained by irrigation practices in the area, however, would require additional study (Segal & Arritt, 1992). South central NE is also bounded by two river systems, which could potentially contribute. A stronger a signal in southwest ND / northwest SD. Irrigation is not a major agricultural practice within this region, however the hotspot rests squarely on the Missouri Plateau, suggesting a potential physiographic component. The geospatial patterns observed in this study corresponded with some analyses of soil moisture feedback (Dirmeyer et al., 2009). However, they did not correspond with other research (Meng & Quiring, 2010). This is to be expected considering the narrow scope of phenomena under study, the size of the study area, and limited data of this study.
APPENDIX A

NEXRAD Data Extraction
APPENDIX B

SPoRT LIS Data Extraction
# APPENDIX C

Variable Descriptions

## Dependent Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max_Area</td>
<td>Maximum observed coverage of reflectivity ≥35dBZ within ROI, over the timeseries</td>
</tr>
<tr>
<td>Max_Ref</td>
<td>Maximum observed value of reflectivity ≥35dBZ, within ROI, over timeseries, capped at a maximum of 65dBZ (hail threshold)</td>
</tr>
<tr>
<td>Slope_Area</td>
<td>Build-up rate (slope) of line defined by time series (within ROI) describing mean coverage ≥35dBZ from t≈0min to the first coverage local minimum/maximum after t≥15min. *</td>
</tr>
<tr>
<td>Slope_Ref</td>
<td>Build-up rate (slope) of line defined by time series (within ROI) describing mean reflectivity ≥35dBZ from t≈0min to the first reflectivity local minimum/maximum after t≥15min. *</td>
</tr>
</tbody>
</table>

*See section 3.2.1 for additional information on calculation of build-up rates

## Independent Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Short Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>StdDev_SSM</td>
<td>StdDev</td>
<td>Standard deviation of LIS Relative Surface Soil Moisture (0-10cm)</td>
</tr>
<tr>
<td>Mean_SSM</td>
<td>Mean</td>
<td>Mean of LIS Relative Surface Soil Moisture (0-10cm)</td>
</tr>
<tr>
<td>Mean_SSM(weighted)</td>
<td>Mean_W</td>
<td>Mean of LIS Relative Surface Soil Moisture (0-10cm), weighted*</td>
</tr>
<tr>
<td>Ratio_Mean_SSM up:down</td>
<td>Ratio_Mean</td>
<td>Ratio of LIS Relative Surface Soil Moisture (0-10cm), mean of entire surface, upwind to downwind</td>
</tr>
<tr>
<td>Ratio_Mean_SSM up:down(weighted)</td>
<td>Ratio_Mean_W</td>
<td>Ratio of LIS Relative Surface Soil Moisture (0-10cm), mean of entire surface, upwind to downwind, weighted*</td>
</tr>
<tr>
<td>Variable</td>
<td>Short Name</td>
<td>Description</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Difference_Mean_SSM up-down</td>
<td>Diff_Mean</td>
<td>Difference of LIS Relative Surface Soil Moisture (0-10cm), mean of entire surface, upwind less downwind</td>
</tr>
<tr>
<td>Difference_Mean_SSM up-down(weighted)</td>
<td>Diff_Mean_W</td>
<td>Difference of LIS Relative Surface Soil Moisture (0-10cm) mean, mean of entire surface, upwind less downwind, weighted*</td>
</tr>
<tr>
<td>Difference_Mean Departures_SSM up-down</td>
<td>Diff_Depart</td>
<td>Difference of LIS Relative Surface Soil Moisture (0-10cm) mean departures, departures calculated separately for upwind and downwind, upwind less downwind</td>
</tr>
<tr>
<td>Difference_Mean Departures_SSM up-down(weighted)</td>
<td>Diff_Depart_W</td>
<td>Difference of LIS Relative Surface Soil Moisture (0-10cm) mean departures, departures from means calculated separately for upwind and downwind, upwind less downwind, weighted*</td>
</tr>
<tr>
<td>GPLLLJ</td>
<td>GPLLLJ</td>
<td>Presence (lack) of Great Plains Low Level Jet, as identified from the 850hPa surface (Bonner, 1968; Frye &amp; Mote, 2010).</td>
</tr>
</tbody>
</table>

*see Section 3.2.2 for information on weighting parameters


Case, J. L. (2016). From drought to flooding in less than a week over South Carolina. *Results in*
Physics, 6, 1183–1184. https://doi.org/10.1016/j.rinp.2016.11.012


83


Joshi, D., Carrera, M., Belair, S., & Leroyer, S. (2017). Influence of Open Water Bodies on the


Parker, M. D., & Johnson, R. H. (2000). Organizational modes of midlatitude mesoscale


PRISM. (2019). *Descriptions of PRISM Spatial Climate Datasets for the Conterminous United States Long-Term Average (“Normals”) Datasets* (pp. 1–14).


https://doi.org/10.2134/agronj2015.0536


Yang, Z., Qian, Y., Liu, Y., Berg, L. K., Hu, H., Dominguez, F., Yang, B., Feng, Z., Gustafson,