January 2016

A Hydrometeorological And Geospatial Analysis Of Precipitation Within The Glacial Ridge Wildlife Refuge Using The R2ain-Gis Tool

William John Mokry

.

A HYDROMETEOROLOGICAL AND GEOSPATIAL ANALYSIS OF PRECIPITATION WITHIN THE GLACIAL RIDGE WILDLIFE REFUGE USING THE R^{2}AIN-GIS TOOL

by

William John Mokry Jr.
Bachelor of Science, Northland College, 2012

A Thesis

Submitted to the Graduate Faculty

of the

University of North Dakota

in partial fulfillment of the requirements

for the degree of

Master of Science

Grand Forks, North Dakota
December
2016

This thesis, submitted by William John Mokry Jr. in partial fulfillment of the requirements of the Degree of Master of Science from the University of North Dakota, has been read by the Faculty Advisory Committee under whom the work has been done and is hereby approved.

Dr. Matthew S. Gilmore, Chairperson

Dr. Mark A. Askelson, Committee Member

Dr. Cedric (Tony) A. Grainger, Committee Member

This thesis is being submitted by the appointed advisory committee as having met all of the requirements of the School of Graduate Studies at the University of North Dakota and is hereby approved.

Dr. Grant McGimpsey

Dean of the Graduate School

December 12, 2016

Date

PERMISSION

Title: A Hydrometeorological and Geospatial Analysis of Precipitation within the Glacial Ridge Wildlife Refuge Using the R^2AIn-GIS Tool

Department: Atmospheric Sciences

Degree: Master of Science

In presenting this thesis in partial fulfillment of the requirements for a graduate degree from the University of North Dakota, I agree that the library of this University shall make it freely available for inspection. I further agree that permission for extensive copying for scholarly purposes may be granted by the professor who supervised my thesis work, or in his absence, by the chairperson of the department or the dean of the Graduate School. It is understood that any copying or publication or other use of this thesis or part thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of North Dakota in any scholarly use which may be made of any material in my thesis.

William John Mokry Jr.

July 6, 2016

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

Acronym	Definition
ANOVA	Analysis of Variance
ASCII	American Standard Code for Information Interchange
CAP	Cooperative Agency Profiler
CFS	Cubic Feet per Second
Dual-pol	Dual-Polarization
ESRI	Environmental Systems Research Institute
GIS	Geographic Information System
GRNWR	Glacial Ridge National Wildlife Refuge
HUC	Hydrologic Unit Code
IDW	Inverse Distance Weighted
K_{DP}	Specific Differential Phase
KMVX	Mayville, North Dakota Radar
MLR	Multiple Linear Regression
NWS	National Weather Service

UND……..University of North Dakota

USGS……United States Geologic Survey

WSR-88D…….……Weather Surveillance Radar-1998 Doppler

Z.....……Reflectivity

Z_{DR}…..…….Differential Reflectivity

ρhv……Correlation Coefficient

ACKNOWLEDGMENTS

I would like to acknowledge the contributions and guidance of my committee members, most notably Dr. Matthew S. Gilmore, Dr. Mark A. Askelson, and Dr. Cedric (Tony) A. Grainger, for the assistance they have provided while completing this study. Their wisdom and advice is greatly appreciated. I would also like to thank my family, Melissa and Ephraim Mokry, Percy, Radar, Adam Schwantes, Chris Jondle, Kendell LaRoche, Matthew Eckhoff, Ron Stenz, and the other graduate students in the Department of Atmospheric Sciences for their support and assistance during my graduate studies.

“The best thing one can do when it’s raining is to let it rain”-Henry Wadsworth Longfellow.

To Melissa, Ephraim, Percy, Radar, Mom, Dad, and Keith.

ABSTRACT

Weather radar (radio detection and ranging) is a specialized meteorological tool used to sample and track meteorological objects. This tool is critical for meteorologists and public decision-makers to inform and provide for their constituents in a timely manner, often with the protection of lives and property on the line. With the application of using meteorological and geospatial data in the realm of geographic information systems (G.I.S.), the task of blending the two sciences to inhibit further research and dissemination of information occurs.

This study focuses on the creation and implementation of a new geospatial tool, the Radar and Rainfall Analyzed in GIS (R^2AIn-GIS) tool. The R^2AIn-GIS tool was built upon the initial concepts from Zhang and Srinivasan's (2010) NEXRAD validation and calibration (NEXRAD-VC) tool for G.I.S. R^2AIn-GIS is updated to support the latest software features present in the geospatial world as well as analyze dual-polarization radar products.

To test the R^2AIn-GIS tool, a warm seasonal precipitation study along with statistical analysis was performed over the Glacial Ridge National Wildlife Refuge in Minnesota, the largest prairie and wetland restoration site. Utilizing rain gauges operated by the United States Geological Survey, warm season precipitation events from 24 May 2012 to 31 August 2013 were analyzed using the R^2AIn-GIS tool.

The R^2AIn-GIS tool calculates the values from various dual-polarization radar products in conjunction with the recorded precipitation gauges to provide a detailed depiction of the weather event. Statistical tests including several iterations of multiple-linear regression of various combinations of dual-polarization radar variables allowed determination of rainfall rate prediction equations over the study area. This contributes to the body of radar literature regarding the best prediction equations for other locations. Unlike treatments in prior literature, most of the various assumptions in multiple linear regression are considered herein.

Based off the findings of the various statistical tests that adhere to the linear regression assumptions, regression models utilizing both reflectivity and correlation coefficient were the best models found during this study. These two variables had statistical significant p-values and their Durbin-Watson scores were among the highest even compared with the other radar variables of differential reflectivity and specific differential phase. Models including the radar variables reflectivity and correlation coefficient were found to be heteroscedastic along with the highest R Squared values. While the overall rainfall amounts were too small in terms of effective precipitation sampling, the results still positively contribute to the literature and provides the opportunity for future work.

CHAPTER I

INTRODUCTION

1.1 Problem Statement

Weather radar (radio detection and ranging) is a specialized meteorological tool that is used to sample and track meteorological objects. Using radar, meteorologists and other weather enthusiasts alike make decisions to serve the public based on storm intensity and radar-based products. Examples of use of radar-based products for public information include, but are not limited to, using reflectivity as a measurement of precipitation, using correlation coefficient (ρ_{hv}) to determine the presence of hail, and using radial velocity to determine rotation and possible presence of tornadoes. In recent years, the National Weather Service (NWS) has been utilizing the -upgraded dual-polarization (hereafter, "dual-pol") Weather Surveillance Radar-1988 Doppler (WSR-88D). Using both vertically and horizontally-polarized pulses of energy, dual-pol radars are used to sample the atmosphere. By sampling in two dimensions, dual-pol radars provide meteorologists with products that depict sizes and characteristics of precipitation as well as even debris lifted aloft by tornadoes.

A goal in this study is enhancement of understanding of the spatial and temporal complexities of data from ground-based rain gauges and dual-pol radar. Specifically, using a modified NEXRAD-validation and calibration (NEXRAD-VC) (Zhang and Srinivasan, 2010) for warm seasonal precipitation (May through August), this study

focuses on the Glacial Ridge Prairie Restoration site in Minnesota. Improving upon the concepts of NEXRAD-VC, the Radar and Rainfall Analyzed In GIS (R^2AIn-GIS) tool adds unto the ESRI® ArcGIS® Library for the spatial analysis of rainfall.

The study domain for this analysis is displayed in Fig. 1. The study area of Glacial Ridge is discussed in greater detail in a later section. The radar utilized in this study is KMVX, with its respective 250 km radar ring shown. The United States Geological Survey (USGS) maintains and operates a network of rain gauges in the Glacial Ridge Prairie site, of which seven gauges are used in this study. Data are recorded with these gauges at both fifteen-minute and one-hour intervals.

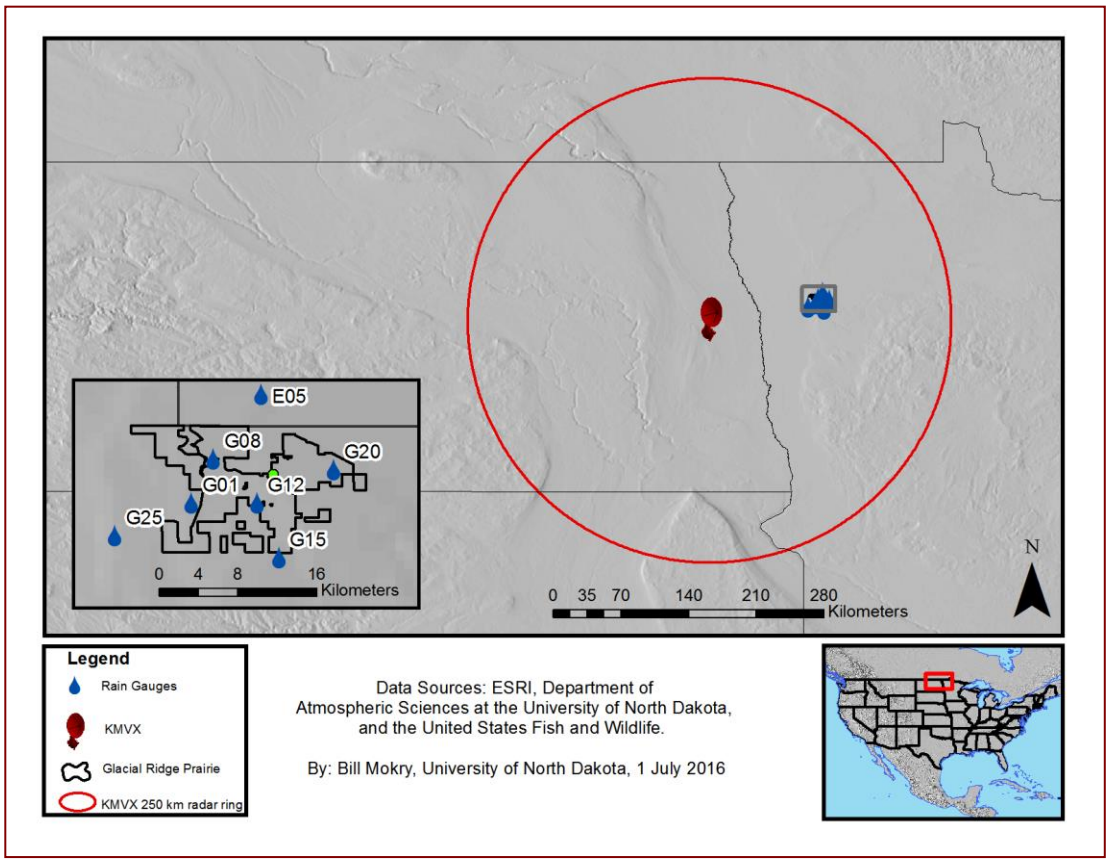


Figure 1. Study domain area showing the Glacial Ridge Prairie site.

The Glacial Ridge Prairie has important impacts on the surrounding communities. The study site major watershed is the Souris-Red-Rainy region, which encompasses the northwest corner of Minnesota as well as the northeast corner of North Dakota along the Red River. Breaking down the Souris-Red-Rainy watershed provides a look at smaller, more local watersheds that encompass Glacial Ridge. Glacial Ridge is located on the convergence of three hydrologic unit code (HUC) watersheds. The watersheds are the Red Lake, Clearwater, and Sandhill-Wilson. Figure 2 shows the major watersheds encompassing Glacial Ridge Prairie study site.

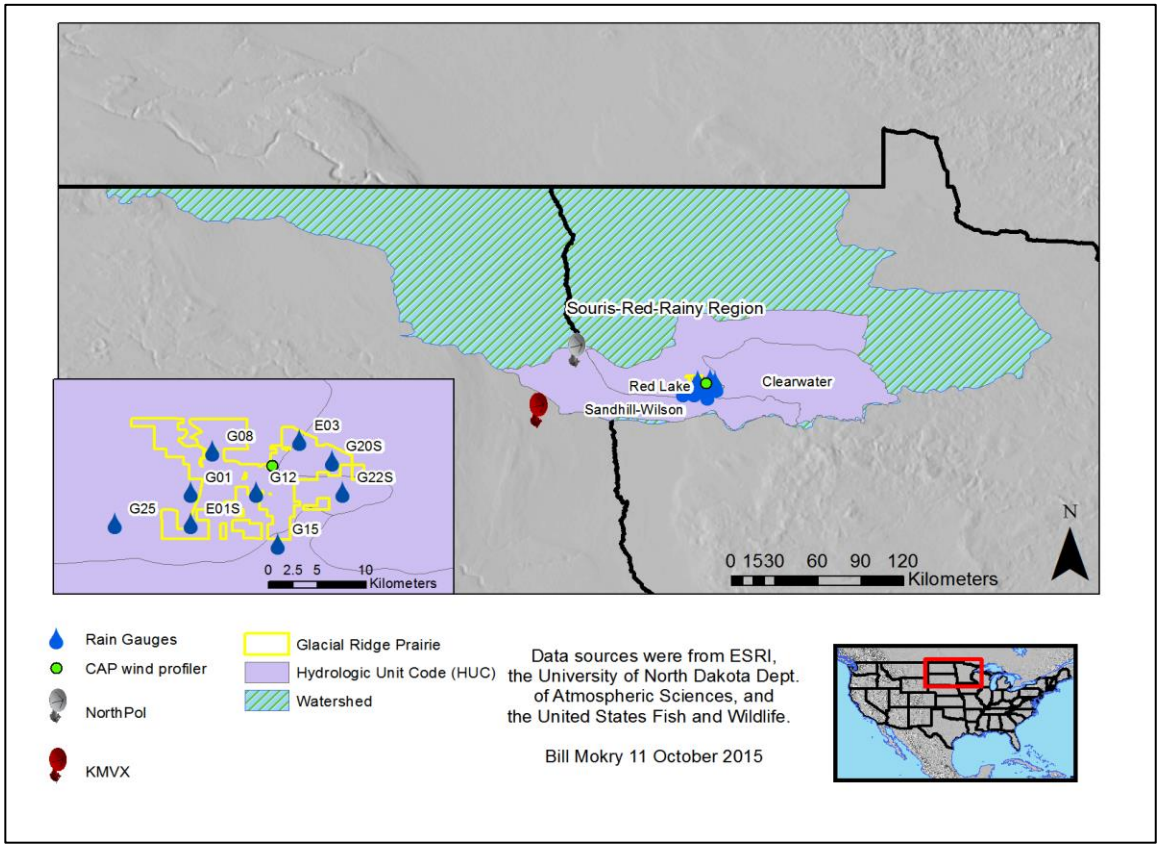


Figure 2. Map depicting the major and hydrologic unit code (HUC) watersheds for the Glacial Ridge Prairie.

The inset map in Fig. 2 provides an in-depth depiction of the influence of the region on the nearby watershed areas. The proximity of the seven rain gauges to each

other and the watersheds enables a unique study of spatial and temporal complexities of precipitation in a small area. From Zhang and Srinivasan (2010), traditional rain gauge measurements are too sparse to accurately depict the spatial variability. This study utilizes seven close proximity gauges to accurately depict the spatial variability of the rainfall. The spatial distance from the furthest east to west station is less than 33 km. The average spacing between all seven gauges is 9.63 km.

Before studying the spatial and temporal complexities of precipitation, several upgrades and changes were applied based upon Zhang and Srinivasan's existing tool for R^2AIn-GIS. The first change was made of necessity while the others were based upon convenience for the author. The tool was

1) modernized to incorporate dual-pol radar products, their derived products, and their associated geospatial statistics;
2) converted from Visual Basic® to Python®, and
3) converted from outdated ArcGIS® version 9.3 to ArcGIS® version 10.3.

CHAPTER II

STUDY AREA

The study area is the Glacial Ridge National Wildlife Refuge (GRNWR), situated about 82 km ENE of KMVX (Fig. 2). GRNWR is the United States' largest prairie and wetland restoration project undertaken by the United States Fish and Wildlife Service (USFWS) (The Nature Conservancy, 2015; hereafter NC, 2015). The scope of the restoration includes converting 24,000 acres of farmland back to their native state. After the retreat of Lake Agassiz, fluvial and glacial processes led to the formation of native wetlands (Coulter, 1910; Todhunter, 2001; Schwert, 2011; Todhunter, 2011; Rogers et al., 2013). A transition landscape between the Red River basin to the west and the forest landscape of north-central Minnesota, these wetlands provide a fire barrier between these two distinct ecosystems. The geomorphic, hydrologic, and hydroclimatic features of the GRNWR, located within the Red Lake Basin (Fig. 3) that drains into the Red River Valley, has an impact on the annual frequency flood risk from spring snowmelt and heavy precipitation (Todhunter, 2011).

The mission of the GRNWR is to preserve and restore prairie grass and wetlands for migratory birds, native vegetation and other wildlife (NC, 2015). The benefits of the revitalized area are not just realized by wildlife. In the process of restoring the wetlands, over 165 km of ditches were filled, restoring over 200 wetland areas that improved water quality and reduced the impacts of flooding along the Red River Valley (NC, 2015).

Since the mid to late 2000s, the United States Geological Survey (USGS), in collaboration with the USFWS, installed a small network of rain gauges to monitor rainfall within the GRNWR. A more detailed description of the rain gauge network is provided in Chapter 3.

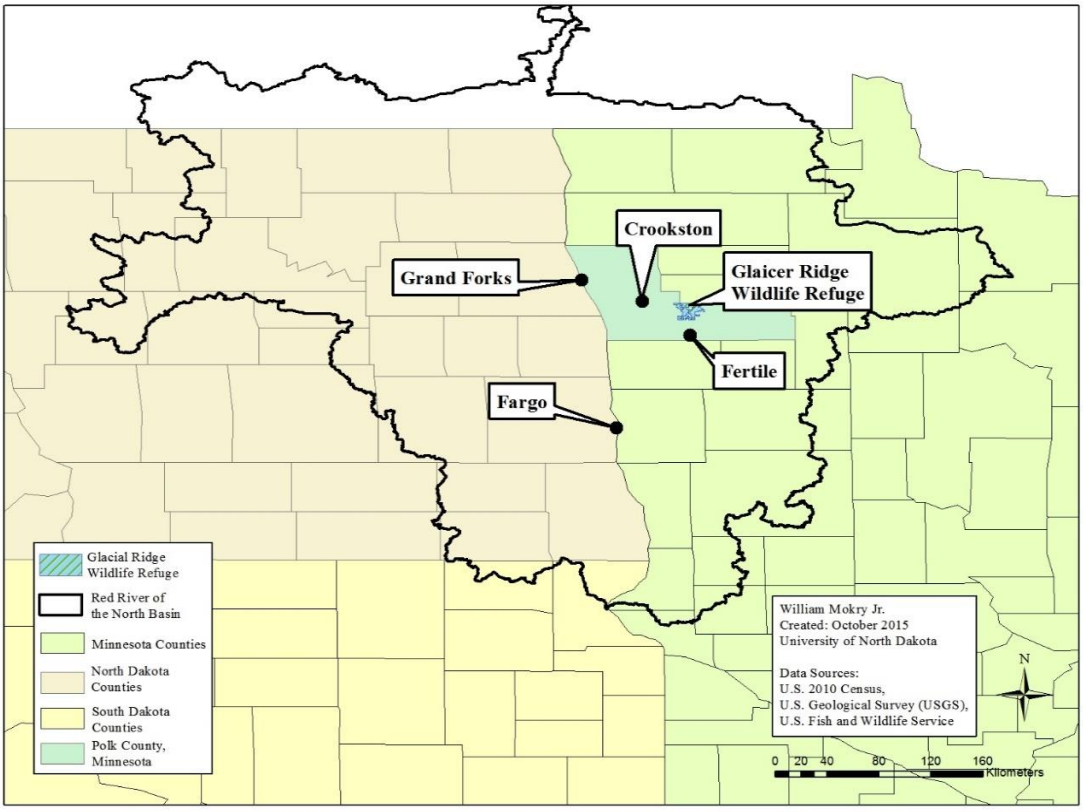


Figure 3. Map of the Red River of the North Basin.

The benefit of utilizing a small area rain gauge network is it enables analysis of small-scale variations, both meteorological and instrument-error-related that can exist across over just a few kilometers. For instance, variations in the radar reflectivity of isolated convective storms, and consequently rainfall, can occur over distances less than 1.6 km (Schilling, 1991). The average spacing of adjacent gauges in the GRNWR is about 4.8 km. Using a small-scale and high-density gauge network allows for improved

sampling of precipitation over a portion of a watershed. This improved spatial sampling can then be used for improved comparisons with radar data, similar to Gebremichael and Krajewski (2004).

Monitoring and investigating precipitation patterns on a watershed provides insight into the influences and the degree to which they impact nearby communities. In terms of hydrology, a watershed's accumulated precipitation is critical for the sustainability of the ecosystem as well as maintaining the water balance in the watershed. Hydrologic models rely upon precipitation totals to determine potential runoff and even overall health of the region (Schilling, 1991).

Restoration of this wetland area back to its native self has critically important impacts for the growing population center in Crookston, Minnesota (United States Census Bureau, 2015). The flat topography and northerly flow of the Red River of the North causes substantial flooding in the surrounding communities.

Crookston, like the communities 25 miles to the NW in East Grand Forks, MN, and Grand Forks, ND, has experienced extensive damage from flooding in the last sixty-five years. A major flood event in 1950, caused by snowmelt and a prolonged rain event, inundated portions of the residential area in lower elevations. Peak discharges in the streamflow measured 27,400 cubic feet per second (cfs), cresting on 7 May 1950 (Red Lake Watershed District, 2006; hereafter RLWD, 2006). Following the flood of 1950, Crookston enacted emergency procedures and mitigation responses to reduce flooding impacts upon the city. Flood prevention methods included construction of levees, sandbag operations, and blasting ice jams. Even though April of 1965 flooding from heavy rainfall on frozen soil surpassed the previous record flood stage, these preventative

measures lessened the damage in the city (RLWD, 2006). Major flooding occurred four years later when an inch of rain fell and along with 5 inches of meltwater equivalent snowpack, damaging the surrounding areas downstream (RLWD, 2006). These major flood events changed the landscape through erosion, which led to costly repairs and subsequent levee improvements aimed at flood mitigation (RLWD, 2006). Restoring previous land-use acreage in the GRNWR allows for the native wetlands to capture and slow the spread of water, thus decreasing the intensity of floodwaters.

Several organizations and research groups partner with the GRNWR to promote education and research outreach from various fields. The University of North Dakota (UND) Atmospheric Science Department collaborates with the GRNWR with an atmospheric research site. This site includes standard weather observation equipment and a vertical scanning wind profiler that is part of the Collaborative Agency Profiler (CAP) network. The CAP network has only one operating wind profiler in the Northern Great Plains that is jointly operated by the Atmospheric Sciences Department at UND. A wind profiler is a vertically scanning radar that can be used to measure the reflectivity (precipitation intensity) and reflectivity-weighted terminal fallspeed of precipitation.

By studying precipitation over GRNWR, further research and documentation of precipitation frequency can occur. These could then be used in future hydrologic modeling to monitor and capture streamflow intensity and duration during snowmelt runoff as well as heavy, long-duration rainfall.

The climate of GRNWR is driven predominately by continental airmasses that produce cold winters and warm summers in the region. For the sake of this study, only warm season precipitation is considered due to inherent errors in rain gauge catchment of

snowfall (e.g., Giangrande and Ryzhkov, 2003). During the warm season, GRNWR is affected by the Maritime Tropical airmass that originates over the Gulf of Mexico. The moist atmosphere coupled with intense daytime heating contributes to instability and development of convective precipitation over much of the region. Extensive wet periods with either widespread or localized precipitation can impact the local hydrology by flooding roads, farmlands, and other low-lying areas. With an increase in available wetlands and water retention, the impacts from extensive rainfall should lessen, but no one has recently tried to determine the impacts based on the new topographic land use.

CHAPTER III

LITERATURE REVIEW

3.1 Rain Gauges

A rain gauge network provides meteorologists and weather enthusiasts alike the ability to monitor rainfall in a given location over an extended period of time. This spatial and temporal system has set the standard in precipitation measurements and has enabled creation of climatologies for a given location's annual or seasonal precipitation. To quantify precipitation at a point, a standard rain gauge is used to measure precipitation fairly accurately with a 5 to 10% error (Neff, 1977). Rain gauge monitoring is critical for decision-making by stakeholders that include farmers, ranchers, fisherman, emergency managers, and natural resource personnel. Over shorter timescales, extreme precipitation events captured using rain gauges can provide information needed for emergency managers to prepare and inform the at-risk public. Adequate time between rainfall and subsequent surface runoff enables stakeholders to prepare and mitigate the effects of flooding (James and Korom, 2001).

The fine-scale spatial and large temporal variability inherent in a convective rain event drives the need for a fine-scale network of gauges capable of sampling the precipitation over short time periods. To aid in hydrologic and ecological modeling, accurate accumulations of rainfall are critical (Arnold et al., 1998; Fodor and Kovacs, 2005). In many instances, rain gauge networks are spaced sparsely in an effort to provide equal spacing over a region. These spread-out rain gauges sample a convective

precipitation event inadequately (Zhang and Srinivasan, 2010). Numerous authors have developed and compared methods to comprehend the spatial distribution of precipitation among a sparse network of gauges (Goovaerts, 2000; Jeffrey et al., 2001; Jolly et al., 2005; Hancock and Hutchinson, 2006; Bannayan and Hoogenboom, 2008; Zhang and Srinivasan, 2008, 2010). Not often does a study focus on a dense group of rain gauges on the scale of 32 km to calculate and determine the spatial and temporal variability from precipitation. Shepard et al. (2004) states that using dense networks of rain gauges provide detailed rainfall measurements to evaluate convective-mesoscale precipitation variability on a spatial and temporal scale.

Meteorologists monitor precipitation inside clouds via radar through its path till it hits the rain gauge at the surface and joins the groundwater. "Ground-truth" measurements, or what is recorded using a rain gauge, provide the best depiction of the amount of water that fell at that location. However, gauges are not without error. The effects of wind through turbulence and increased speeds near and around the gauge can result in undercatchment (Stellman et al., 2001). Even hard rain can lead to undercatchment when the gauge is tipping the bucket and recording precipitation yet rain is still falling into the gauge not being collected and recorded. Another source of error occurs when precipitation is interpolated between a sparse collection of gauges.

This study focuses on a close-gridded network of rain gauges in Glacial Ridge Prairie in Minnesota over the warm-precipitation season from April through October in 2012 and 2013. This study utilizes gauges having periods of record of less than fifteen years, making a long-term climatology impossible until the period of record increases considerably. Incorporated with the gauges are dual-pol radar products from KMVX.

Because KMVX was only upgraded to dual-pol capabilities on 24 May 2012, this study focuses on rain events after KMVX completed its transition to dual-pol. Because of the sparse density in the national rain gauge system, NWS forecasters use Weather Service Radars to estimate rainfall between gauges, which aids in watch and warning operations. This has reduced the negative impact of spread-out gauge networks (Stellman et al., 2001). The polarimetric capability of such radars has existed since 2012.

3.2 Radar and Dual-Polarization Products

Since the late 1940s, weather radars have been utilized to remotely sense precipitation and other various weather phenomena. The benefit of utilizing radar is to scan over a large area in a reasonable time. It is beneficial to scan over a large area in a reasonable time.

The NWS analyzes radar products from the NEXRAD radars that serve each weather forecast office (WFO). For overviews of weather radar theory and derivation of the radar equation and how it applies to rainfall measurements, see Doviak and Zrnić (1993) and Battan (1973).

Probably one of the most important products from radar is radar reflectivity factor (Z). When radar pulses of electro-magnetic energy hits a hydrometeor target, the energy is absorbed and/or scattered by the hydrometeor based on the incident radiation wavelength and the phase and size of the hydrometeor (Clement, 1995). Hydrometeors will scatter the incident radiation following the laws of Rayleigh scattering when the radiation wavelength is at least 16 times smaller than the diameter of the largest

hydrometeors. For this study, KMVX is an S-band wavelength radar allowing for valid Raleigh approximation hydrometeors smaller than 6.3 - 6.9 mm diameter, which nearly all raindrops on earth are. With a Raleigh approximation, the hydrometeor's direction and scattering pattern is known. The backscattering cross section for a liquid hydrometeor (σ) is

$$\boldsymbol{\sigma = \frac{\pi^5}{\lambda^4} |K_w|^2 D^6}, \tag{1}$$

where λ is the radar wavelength, $|K_w|$ is the complex index of refraction of water, and *D* is the equivalent volume spherical raindrop diameter. One can see that radar reflectivity factor is related to the sixth power of the diameter of the hydrometeor. The backscatter measurements are a collection of all targets and hydrometeors within the volume of space sampled with the radar. The volume covered by a radar pulse of energy is based on the distance from the radar, radar antenna beam width, and radar pulse length (e.g., Clement, 1995; Rinehart, 2010). The radar reflectivity calculated by the radar's post-processor is a function of this radar beam, the antenna characteristics, as well as the power of the backscattered energy received by the radar using the so-called radar equation (not shown). When contours of radar reflectivity are color-filled and animated on a map, meteorologists can monitor the movement and rainfall intensity of storms at range and in locations without rain gauges.

However, using (1), linear radar reflectivity ($mm^6\ m^{-3}$) can also be calculated using actual raindrop size measurements with either a discrete (not shown) or the integral/continuous equation,

$$z = \frac{\lambda^4}{\pi^5 |K_w|^2} \int_0^{D_m} D^6 N(D) dD, \quad (2)$$

where D_m is maximum drop size diameter and the drop size distribution is $N(D)$.

Marshall and Palmer (1948) found that raindrop size distributions, plotted on a D versus ln(D) scale, can be functionally fit by a line that intercepts the y-axis at N_0, with negative slope, $-\lambda$ using

$$N(D) = N_0 e^{-\lambda D}, \quad (3)$$

and this can be plugged into the continuous form of Z in (2) and is easily integrated to obtain an analytic solution (not shown).

Just as one may use measured raindrop sizes with the discrete form of (2) to obtain radar reflectivity, rainfall rate may additionally be calculated with raindrop terminal fallspeed $v_t(D)$, over different raindrop diameters, D. As rain drops fall, they can collide and coalescence to form a larger drops or break-up into smaller droplets due to hydrodynamic effects and raindrop collisions. Rain rate, R (usually in mm h $^{-1}$), can be expressed as

$$R = \frac{\pi}{6} \int_0^{D_m} D^3 N(D) v(D) dD, \quad (4)$$

and there are a variety of functions fitted to raindrop $v_t(D)$ that may be used. However, this is not the most common way to measure rainrate. A simpler way is to use rain gauge data.

Combining (4) and (2) provides a relationship between the z and R of the form

$$z = AR^b. \quad (5)$$

with z ($mm^6\ m^{-3}$) and R ($mm\ hr^{-1}$). Such a relationship can also result from a best fit to radar-observed z and gauge-observed R as will be discussed later. Since many different drop size distributions can exist at different locations within a single storm and between nearby storms, the use of just one z-R relationship can prove ineffective. Battan (1973) and Joss et al. (1970) stressed the importance for varying z-R relationships based on storm type, season, and location. A new z-R relation is created herein using the reflectivity and rainfall values over the gauges to better calculate rain rate over the region. If one is working with radar reflectivity data already provided in Z (dBZ), one may convert to z ($mm^6\ m^{-3}$) using

$$Z_H = 10 \log_{10}(z_H), \quad (6)$$

although the horizontal polarization subscript is usually omitted.

One of radar products available from polarimetric radars (such as KMVX), being utilized in this study is differential reflectivity (Z_{DR}), which utilizes the transmission of horizontally- and vertically-polarized pulses of radiation (Seliga and Bringi, 1976). In this process, the radar emits a horizontally-polarized pulse and receives the returned power and then emits a vertically-polarized pulse and receives the associated returned power. The vertically- and horizontally-polarized pulses are used to compute radar reflectivity factors for each polarization (Rinehart, 2010). Mathematically it is given by

$$Z_{DR} = 10 \log_{10}\left(\frac{z_H}{z_V}\right), \quad (7)$$

where Z_{DR} is in decibels (dB), and z_H and z_V are the reflectivity values ($mm^6\ m^{-3}$) at horizontal and vertical polarization, respectively. Using logarithmic units for the radar reflectivity factors, Equation 6 can also be expressed as $Z_{DR}=Z_H-Z_V$ (e.g., Rinehart, 2010).

As with differential reflectivity, another dual-pol product utilized in this study is specific differential phase (K_{DP}). Similar to how reflectivity incorporates backward scattering, K_{DP} is a measure of forward scattering that is used to determine rain rate. However, it is important to note that the physical causes are different. K_{DP} (degrees phase shift km^{-1}) can be described by Clement (1995) as the range derivative of differential phase and is defined as

$$\boldsymbol{K_{DP} = \frac{180}{\pi} \lambda Re \int [f_H(D) - f_v(D)]\, N(D) dD,} \tag{8}$$

where $f_H - f_v$ is the phase shift between the horizontal and vertical forward-scattered polarized waves, integrated over all drop sizes.

Both Z_{DR} and K_{DP} are helpful to identifying regions with large raindrops and distinguishing them from hail. The science behind this is in the shape of large raindrops and hail as they fall. Both small spherical raindrops and hail that is, or appears spherical to the radar due to tumbling, result in the same phase shift and same reflectivity measured between the horizontal and vertical components. This gives near-zero K_{DP} and Z_{DR}. In contrast, large falling raindrops are oblate, meaning their horizontal axes are larger than their vertical axes and they, thus, produce positive K_{DP} values and positive Z_{DR} values. One benefit of using K_{DP} compared to Z_{DR}, however, is for cases where there is a mixture of both hail and large raindrops in the same volume. In that case, hail would dominate the backscattering of the H and V pulses, resulting in Z_{DR} closer to zero. However,

spherical hail would not affect the phase shift of those pulses that were caused by large oblate raindrops, thereby still allowing positive K_{DP}.

Correlation coefficient, ρ_{hv}, in radar meteorology is the correlation between the radar signals at the horizontal and vertical scans for a given point at the same time. Following Brandes (2000), the equation to define ρ_{hv} can be given as

$$\boldsymbol{\rho_{HV}} = \frac{\langle \boldsymbol{s_{VV} s^*_{HH}} \rangle}{\langle |\boldsymbol{s_{HH}}|^2 \rangle^{1/2} \langle |\boldsymbol{s_{VV}}|^2 \rangle^{1/2}}, \tag{9}$$

where s and s* are scattering matrices and H and V represent the horizontal and vertical signals of the received and transmitted polarizations from the radar. In application, meteorologists use ρ_{hv} to identify areas of mixed-phase, shapes, or sizes of hydrometeors and other items aloft. Perfect spheres will have a ρ_{hv} value of 1. Raindrops however have values ranging from 0.97 to 0.99 (Rinehart, 2010). The benefit to this study in utilizing ρ_{hv} is to be able to identify and distinguish areas of rain from non-rain. Low values of ρ_{hv} symbolize areas of non-homogeneous particles and are used to detect melting regions, tornado debris, and even non-spherical hail. High values of ρ_{hv} conversely symbolize areas of homogeneous regions where hydrometeors are similar in phase, shape, and/or size.

3.3 Geographic Information Systems (GIS)

Geographic Information Systems (GIS) are software used in many disciplines to input, store, analyze, manipulate, and display spatially-defined data. Users can map and represent data with points, lines, polygons or rasters within a consistent coordinate system. Points are features that can be represented geographically using an x,y

coordinate. Lines are connections between points. A polygon is a closed, connected group of lines defining an area. Raster datasets are cellular data comprised of rows and columns, with cell groups representing features and each cell value being the value of the feature. Lindhult et al. (1998) defines the composition of GIS to be of a data input, analysis and manipulation, and extracting and displaying data. A variety of software platforms exist for GIS users to interpret and analyze geospatial data. In some instances, some GIS software are open source and free to use, while other GIS software require paid subscriptions to operate. For the study herein, Environmental Systems Research Institute (ESRI, 2015) ArcGIS® 10.3 software is used.

Data input is a critical step to take real-world information and convert it into an applicable digital format for display. One input method includes geo-referencing an image. Another method involves reading point data in tabular form with appropriate headers (that include latitude and longitude) and spatially joining them to the geo-referenced information already being displayed. Once information has been properly digitized, the information present in an attribute table is linked and hence associated to the spatial information and vice-versa. The attribute table is a collection of the variable data for a shapefile such as latitude-longitude coordinates, feature values, and description of feature. In this study, the network of rain gauges were features in a shapefile. Figure 4 provides a glimpse into the variables associated with rain gauges used in this study and

the attributes associated with each gauge.

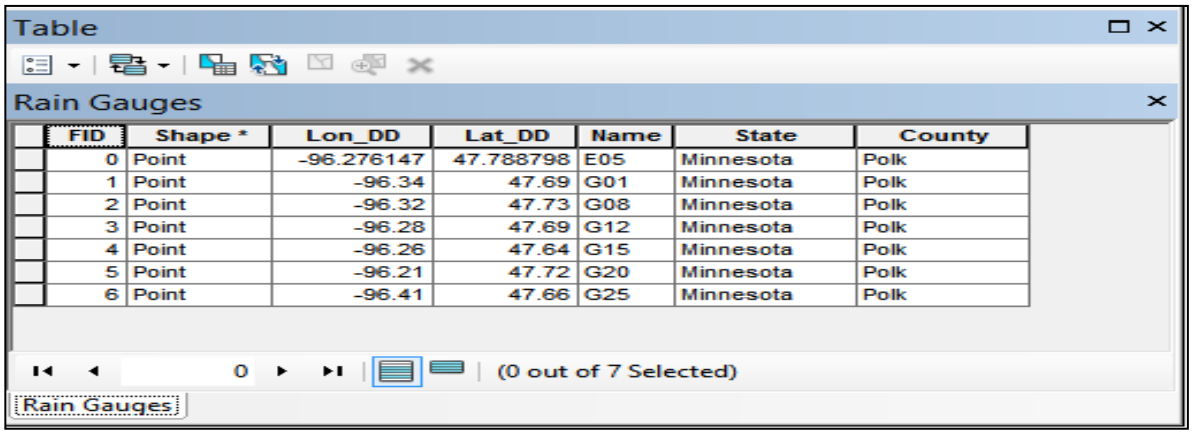


Figure 4. Depiction of attribute table for the seven rain gauges in ArcGIS®.

A feature utilized in the ESRI ArcGIS® product known as ArcMap has a built in section for displaying descriptive information. The metadata are provided in a user-input section where details about the file are input. Items such as theme keywords, abstract, purpose, bounding coordinates, number of records, time period for when data are relevant and publication information are described by the user for use by others.

In the analysis and manipulation component of GIS, spatial and/or tabular data undergo user-selected procedures to determine values of particular attributes in a given region. In this study, radar data are overlaid upon rain gauges. Typical attributes of radar data include reflectivity values, coordinates of the radar grid cell, and time of scan. Typical attributes of rain gauge data include coordinates of the rain gauge, total rainfall, time period of rainfall, state, and name. Utilizing both sets of spatial data enables determining the relationship between the two sets of data. Much of the analysis performed in this study utilizes spatial analysis tools provided by ESRI in the ArcGIS®

software. Spatial analysis tools within the ArcGIS® suite of products are used to extrapolate and interpolate between irregularly-spaced point data (like the gauge data).

The final composition as defined by Lindhult et al. (1998) for GIS is the extracting and displaying of data. With the GIS realm of software, data can searched and portrayed in a multitude of ways. A simple method is a choropleth map using colors to shade and pattern trends to visually display the data. One can focus on either a particular region or an entire domain. Many of these GIS-generated maps are drafted and disseminated with a client or group in mind. When exporting the data using GIS, many file output formats are available. Many raster-based images can be transformed into the following file types: ASCII, binary, Google Earth®, Excel®, and image-based (GIF, JPG, etc).

Zhang and Srinivasan (2010) state that the GIS applications available to process WSR-88D data are limited and lack processing power. Having the means to process radar data along with other data sources, in this case rain gauges, is a critical component that is addressed and improved upon in this study. The fact that the NEXRAD data system is not readily available in formats used in different scientific divisions suggests a "lack of geo-processing and geo-referencing" that could be useful, especially in agriculture, meteorology, and natural resource management (3 et al., 2008). Currently, the National Oceanographic and Atmospheric Administration (NOAA) maintains the Weather and Climate Toolkit (WCT), a software application that can be used to display and export radar data into other formats.

The WCT can be used to display radar data, but currently cannot be used to produce additional products and information about the spatial and temporal

characteristics of the data. The benefit of the WCT, however, is that it enables exporting of radar data to a variety of different formats. For this research, all NEXRAD data were exported from the WCT into an American Standard Code for Information Interchange (ASCII) format for visualization and processing in the ArcGIS®-software ArcMap. A header file is created when the data are exported.

ArcGIS® software provides the opportunity to geospatially reference and process data. The toolset in ArcGIS® allows for analyzing and displaying spatial data through either user-defined or pre-defined calculations. It is noted that this study utilized the applications and tools available in the ArcGIS® Spatial Analyst license. The licensing provides the user authority to utilize products from ESRI. These products add additional spatial modeling and analysis tools to ArcGIS® that are otherwise not available.

CHAPTER IV

DATA AND METHODS

4.1 Data Sources and Case Selection

The data for this study were collected from two different sources and input into ArcGIS® 10.3. As shown in Table 1, the datasets include: 15-min rain gauge accumulation data from USGS and 0.5 degree NEXRAD Level 3 products for Z, Z_{DR}, K_{DP} and ρ_{hv} at regular scan intervals. First, the 15-min rain totals were entered into an Excel® spreadsheet for each gauge in the network for each storm event. An initial list of storm events was produced for the period from 24 May 2012, beginning when the upgrade of NEXRAD KMVX to dual-pol was completed, to 29 August 2013. Storms utilized for this study produced, at one of the gauges, greater than 0.1 inches of rain, and must have produced precipitation at a majority of the seven gauges.

Table 1. Sources and datasets included in the study.

Data Source	**Type of Data**	**Years Included in Analysis**
United States Geological Survey	15-min Rain Gauge Precipitation (in)	2012-2013
National Oceanic and Atmospheric Administration	0.5 degree radar data: Z, Z_{DR}, K_{DP} and ρ_{hv} variables	2012-2013

Characteristics of the USGS rain gauges are provided in Table 2. All of the gauges within the GRNWR have been in operation since 28 August 2008. The earliest gauge in operation is G08 and dates back to 28 April 2003, giving it a period of record of less than thirteen years. Table 3 lists the storm events that met the initial criteria for analysis in this study, along with USGS rain gauge availability. Rain gauge data were provided by the USGS in fifteen-minute increments ending at 00, 15, 30, and 45 minutes past the hour, measured in inches. For this study only seven gauges were available out of the 10 total operated by the USGS. The other three were under provisional status and being reviewed by USGS staff to verify the measurements (T. Cowdery, personal communication, May 8, 2014). As such, the precipitation data for these gauges were not sent to the author for analysis.

For storm consideration, recorded precipitation for the day needed to exceed 0.10 inches along with the majority of gauges being in operational status. Only allowing the warm season limits uncertainties caused by mix-phased or solid precipitation near the surface along with freezing temperatures that could affect gauge reporting. These dates were cross-checked with storm data reports from the Storm Prediction Center for hail activity that could potentially interfere with gauge reports, resulting in the final list shown in Table 3. Figure 5 shows a detailed flow chart of the R^2AIn-GIS tool.

4.2 Methodology

Upon receipt of the rain gauge information from the USGS, the rainfall information were entered into a Microsoft Excel® spreadsheet. The radar data availability for each storm event was confirmed using the National Centers for

Environmental Information (NCEI 2015; formerly the National Climatic Data Center - NCDC). From NCEI, the radar data at base-scan elevation angle of 0.5° was ordered and obtained for variables Z, Z_{DR}, K_{DP} and ρ_{hv} for each initial storm event for the duration of precipitation being recorded with the gauges. Level 3 radar data were ordered through NCEI and obtained through an ftp download in the WCT. Additional time buffers of an hour on each end were ordered as well to ensure that precipitation events were captured in their entirety. Once the data were obtained from NCEI, the NOAA WCT was launched to retrieve and export the storm event radar data. Using the toolkit, the radar files for each scan were exported into an ASCII grid file (.asc), with adjoining projection file (.prj) for spatial reference. The spatial reference for each file is in a Geographic Coordinate System North American 1983 for a spheroid. The projection file is essential for drawing the grids appropriately in ArcGIS®. Within ArcGIS®, a modelbuilder script was incorporated to quickly and efficiently process the rain gauge and data for each timestep of each storm event. Modelbuilder is a tool within ArcGIS® for visually and computationally building scripts that automate data processing. The R^2AIn-GIS tool was created by the researcher within the confines of the modelbuilder script.

For this study, the goal of developing a multiple linear regression (MLR) equation to describe rainrate in terms of the reflectivity, differential reflectivity, specific differential phase, and correlation coefficient. The user provides SPSS with a table of values for each variable and a MLR is developed.

This study focuses on the application of two overarching methodologies—geospatial and statistical analysis. Specifically, these approaches provide critical information to further improve the field of hydrometeorology and applications of dual-

polar radar data in the realm of GIS over the Glacial Ridge Prairie. The results were analyzed in ArcGIS® 10.3, Microsoft Excel®, as well as the statistical software, SPSS.

4.3 ASCII to Raster Conversion

As part of the R^{2}AIn-GIS, the radar variables downloaded from WCT are converted from an ASCII file to an ArcGIS® float raster for viewing and calculations. When the user downloads radar data from WCT, the user has the option to export any or all of the radar data to a variety of different file extensions. For this study, the exported data were set to be that of an ArcGIS® float raster. The reason for the specific file format is for accessibility and processing of the raster within the software of ArcGIS®. Once the ASCII files are exported, the conversion from ASCII to raster must occur to properly display the data and to utilize other raster tools. A script in the R^{2}AIn-GIS tool loops through the user-defined directory where the WCT exported radar files are located and converts them to a raster format in an output directory. To retain the details and character length of the radar data, the rasters were output to a file geodatabase, which allows for all characters in the file name to be retained. Raster-based files outside of a geodatabase are limited to a length of only thirteen characters.

Table 2. United States Geological Survey gauges and period of record included in the study.

Station Name	**Latitude & Longitude (Decimal Degrees)**	**Period of Record Begin**	**Period of Record End**
E05-R 150N44W27ABBAA L058 (Denoted: E05)	47.7889 -96.276125	27 August 2008	Operational
G01-R 149N44W30CAAD 0000620661 (Denoted: G01)	47.692972 -96.341736	30 April 2003	Operational
G08-R 149N44W17ABAD 0000620668 (Denoted:G08)	47.729439 -96.315225	28 April 2003	Operational
12-R 149N44W27CDBB 0000620672 (Denoted: G12)	47.690381 -96.281289	29 April 2003	Operational
G15-R 148N44W10CCCC 0000620675 (Denoted: G15)	47.6447 -96.258608	29 April 2003	Operational
G20S-R 149N43W18DDBACA01 0000620680 (Denoted: G20)	47.719017 -96.205625	6 May 2004	Operational
G25-R 148N45W05DDDD 0000620685 (Denoted: G25)	47.659256 -96.410411	6 May 2004	Operational

Table 3. Storm events that met criteria for analysis.

Storm Date	Number of Gauges Reporting	Name of Gauges Unavailable
27 May 2012	6	E05
28 May 2012	6	E05
10 June 2012	6	E05
14 June 2012	6	E05
16 June 2012	6	E05
20 June 2012	6	E05
23 June 2012	6	E05
3 August 2012	5	E05, G25
25 August 2012	6	E05
19 May 2013	7	None
20 May 2013	7	None
21 May 2013	7	None
30 May 2013	7	None
31 May 2013	7	None
22 July 2013	5	G1, G15

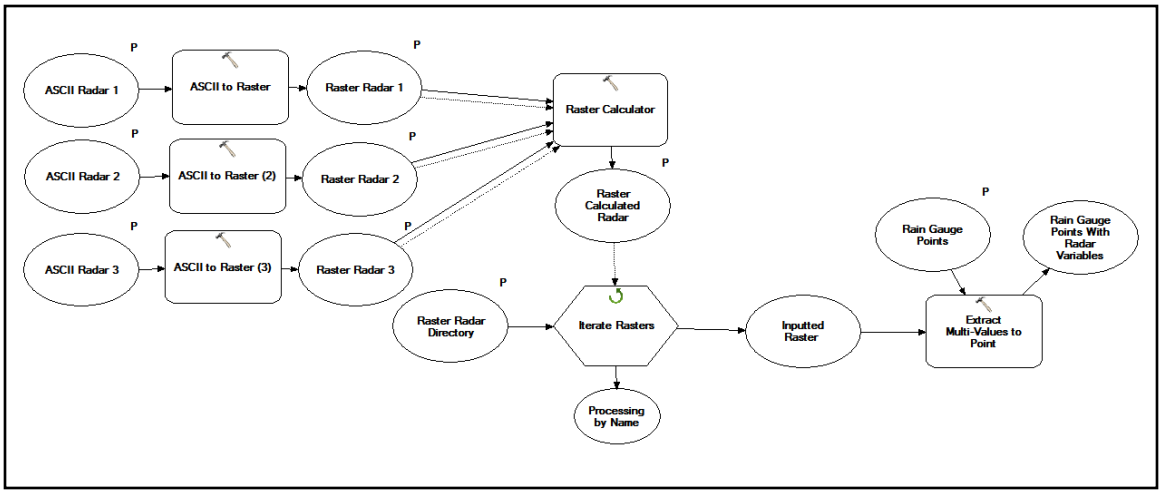


Figure 5. R^{2}AIn-GIS tool flow chart.

4.4 Raster Calculator

Once all the WCT radar ASCII files are exported to an ArcGIS® float-raster format, the process of calculating radar time-averaged data begins. The purpose of averaging the three sequential radar scans is to match them up with the temporal spacing of rain gauge data. The next step in R^{2}AIn-GIS is the utilization of the raster calculator tool to combine three consecutive radar files and compute average values. The user defines the three sequential radar files that are used for each fifteen-minute period and sets an output for the newly created, calculated raster. For the scanning strategies used by the NWS in this study, averaging base scan data for three consecutive volume scans resulted in periods of about 13 min to 18 min. While not exactly an averaging period of 15 min, it was deemed better than using only a single instantaneous radar scan. The reason for user-definition in the calculation of the rasters is that the radar scans rarely begin at precisely 00, 15, 30, or 45 minute after every hour like the gauge data. Oftentimes the radar scan times are offset by several minutes relative to the recording times of the gauges. These cases, take the three consecutive radar volume scan periods

with the last grid being closest to the gauge recording time. The radar times were included in the naming scheme of each radar file. Some raster calculations did end on either of those four intervals, providing a precise timing between gauge and radar. For example, the period ending exactly at 15 minutes past the hour will include the average of the sum of the 5 minute, 10 minute and 15 minute rasters. If there was a missing raster during the sequence of the 15 minute averaging, the average of the two rasters closest to the time of rain gauge reporting were used. If the radar variable had no values present within each raster included for the 15 minute average, the value was set to missing upon entry into the Excel® file. The output raster calculations for this study were exported to another file geodatabase where the entire file name would be retained. An example is depicted in Fig. 6. It should be noted that for the averaging of the rainrate and for this study that each gauge was treated independently.

4.5 Point-Based Data Extraction and Table Conversion

Using the latitude and longitude as the basis to display the rain gauges, the points are exported to a shapefile format for use in ArcGIS®. When the points are exported to a shapefile, an attribute table is created based on the input text or Excel® file of origin. For each storm event, the gauges have columns for each fifteen-minute precipitation total. Using the file geodatabase that contains the raster calculations, radar variable values are exported to each gauge location using the "extract values to point" tool within R^2AIn-GIS. This tool utilizes a loop iterator to extract each radar variable value (in rasters). The user defined point shapefile, in this case the rain gauge locations, will then have the radar variable values added on as additional columns in the attribute table. Once the

values were appended to the attribute table, the table is exported to an Excel® file where the rows and columns were organized for further analysis in the statistical software SPSS.

4.6 Statistical Analysis of Data

Within Excel®, the tables with rain gauge rainfall and averaged radar variable values for each gauge are formatted and saved to a comma-delimited ASCII file format for input into SPSS, a statistical software. The total number of observations for all case dates for R, Z, Z_{DR}, K_{DP} and ρ_{hv}, are respectively 738, 1327, 1417, 971, and 1422. The number of observations for each radar variable differs owing to missing data for certain variables. Also, there are more radar observations than gauge observations shown here because times with zero or missing rain were included but considered null in the dataset.

As part of the formatting within Excel®, an initial scatterplot of the variables were plotted to verify the need to perform transformations of the data (not shown). First, a units conversion for rainfall (inches per 15 minutes to mm per hour) was done. Then, in keeping with the literature, and because the raw data plots were curved (not shown), R was changed into a logarithmic variable by calculating $\log_{10}(R)$. Z and Z_{DR} are both already logarithmic variables, however, because $Z=10\log_{10}(z)$, then $Z/10$ (dBZ) was calculated and prepared for input. Z_{DR} (dB) was input directly (though prior authors seem to have calculated and used $Z_{DR}/10$). No transformation was applied to ρ_{hv} (although prior authors seem to have applied $\log_{10}$). Finally, following Giangrande and Ryzhkov (2003), the log of the absolute value of specific differential phase was calculated and then multiplied by its sign to preserve the integrity and avoid mathematical error. Thus, the general form of the regression equation that was tested was as follows,

$$\log_{10}(R) = b_0 + b_1(Z/10) + b_2(Z_{DR}) + b_3\rho_{HV} + b_4 \log_{10}|K_{DP}|sign(K_{DP}), \quad (10)$$

which can also be written, after substitution of Equations (6) and (7),

$$\log_{10}(R) = b_0 + b_1 \log_{10}(z) + b_2\, 10\log_{10}(z_{dr}) + b_3\rho_{HV} + b_4 \log_{10}|K_{DP}|sign(K_{DP}) \quad (11)$$

Where $z_{dr} = \left(\frac{z_H}{z_V}\right)$ is referred to by Giangrande and Ryzhkov (2003) as the unitless "linear zdr term". After the SPSS is run, one may write the resulting equation in power law form by raising each side of the equation as the power to 10 giving

$$R = 10^{b_0}\, z^{b_1}\, {z_{dr}}^{10b_2}\, 10^{(b_3\rho_{HV})} K_{DP}^{b_4} sign(K_{DP}) \quad (12)$$

Where z ($mm^6\ m^{-3}$), z_{dr} and ρ_{HV} (both unitless), and K_{DP} (deg km^{-1}). The sign(K_{DP}) is logic used to keep rain rate from becoming negative. Within SPSS, each comma-delimited ASCII file is imported. Following the import steps, the user, when applicable, sets any missing values that might exist due to missing radar variables found during periods. The user can specifically specify the correct value used to denote a missing value such as in the case of having missing values for correlation coefficient being -999 but reflectivity after transformation to reflectivity/10 becomes -99.9. The user also sets the decimal placement for the variables to ensure the values are properly read into the program. Then, statistics are computed —these include Analysis of Variance (ANOVA), Pearson Correlation, linear regression, and multiple linear regression (MLR). Linear regression is performed to obtain a predictive equation for rainrate based upon each radar variable. The so-called z-R relationship is one example and results from performing

linear regression only between z and R observations, as first done by Marshall and Palmer (1948). In that case, Equation (12) only has one independent variable $R = 10^{b_0} z^{b_1}$

The correlation (positive [>0], no correlation [0], or negative [<0]) between the independent (radar products) and the dependent variable (rainfall) is used to evaluate the strength of the linear relationship between the two variables. Similarly, multiple linear regression is used to evaluate the relationship between the rainfall reported at each gauge (dependent variable) and two to four radar variables mentioned before (independent variables). Multiple correlation evaluates the combined strength of the linear relationship between the independent variables and dependent variable. Such an equation serves as a predictor for Glacial Ridge Prairie rainfall amounts based on the corresponding radar variable values. By having an equation to apply to predict rainfall, people with an interest or stake in the Glacial Ridge Prairie can monitor and react appropriately to rainfall or lack thereof.

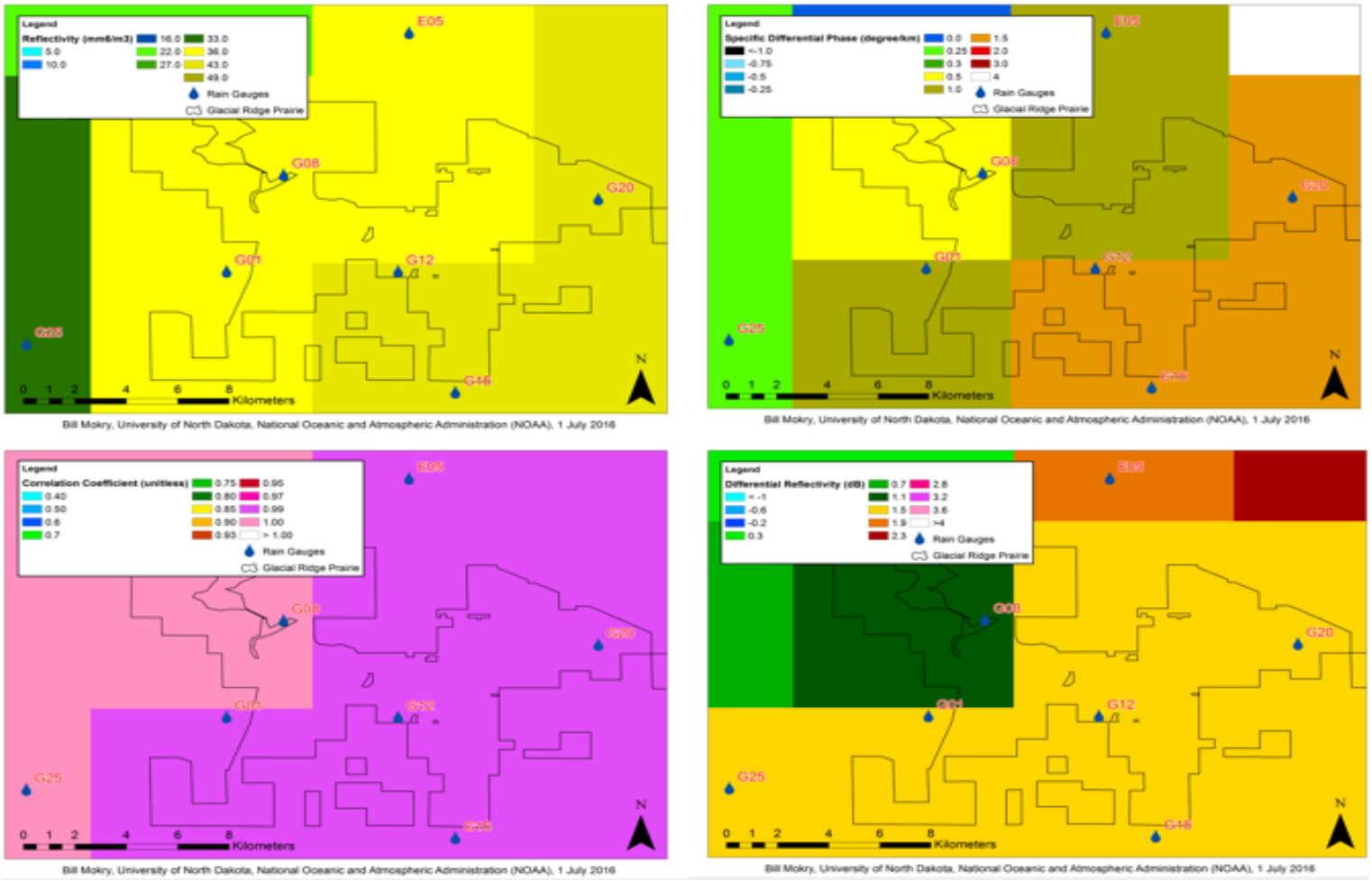


Figure 6. Four panel depiction of the radar variables: reflectivity (upper left), specific differential phase (upper right), correlation coefficient (lower left), and differential reflectivity (lower right)

Following Giangrande and Ryzhkov (2003), multiple linear regression was performed with different combinations of the four radar variables to determine which combination is the best predictor of rainfall. Each of the models were evaluated independently using Analysis of Variance (ANOVA). The multiple correlation coefficient is used to evaluate the overall best model. Each variable within each model is also evaluated separately (with t-statistic and p-values) to determine how much of the variance in rainrate is explained by each variable. Specifically, the p-value provides an indication of the probability that the independent variable is explaining the variability using a non-zero slope by random chance when the slope is actually zero (no effect). Statisticians refer to the p-value as "the probability of falsely rejecting the null hypothesis". In this case, the null hypothesis is that the radar variable has no effect on rain rate (zero slope). If the p-value is sufficiently small, we reject the null in favor of the alternative (that the slope is actually "real" and not an accident). Most commonly, it is said that a p-value of 0.050 or less is statistically significant, where it is "safe" to reject the zero slope notion, but variables with smaller p-values are considered even better because that means that those variables are less likely to have zero slope (more likely to have a nonzero slope value). If the p-value is greater than 0.050 then it is said to not be statistically significant as there is a larger probability that the variable actually has zero slope (zero predictive capability), even though ANOVA may calculate a non-zero slope. This allows one to determine whether an unhelpful variable can be excluded from a particular model.

The significance of the results (smallness of the p-value) in any statistical test, as well as the representativeness and accuracy of a regression equation, is highly dependent

upon the number of independent observation samples. This means that if the sample size of rainfall to the radar variables is too small, obtaining a predictive rainfall rate equation using linear regression based on the radar variables will be neither accurate nor representative of the region. Following Giangrande and Ryzhkov (2003), fifteen rain events with 114 hours of observations were analyzed, similar to their 15 rain events with 52 hours of observations. It should be noted that Giangrande and Ryzhkov (2003) did not make mention of whether they checked that their linear regression was appropriate. This study strives to ensure the linear regressions are appropriate with various statistical testing performed. Again, the storm events for this study follow the criteria previously mentioned in Chapter 4 Section 1.

There are also several assumptions that are made when performing a multiple linear regression that must be checked and satisfied before trusting the models resulting from a multiple linear regression. Some of these would ideally be checked prior to conducting the statistical ANOVA and multiple linear regression while others can be checked by examining the results of the regression. Usually statistical tests are involved that describe the degree of violation of the assumption. If the assumptions are violated in some way, sometimes data transformations can be applied such that the assumptions are satisfied.

The first assumption is that the data exhibits a linear relationship between the dependent and each independent variable. A scatterplot of each radar variable (independent variables) versus rain rate (dependent variable) can be used to check a linear relationship prior to performing the regression (figures were checked but are not shown in results). If the scatter appears nonlinear, a transformation can be applied (such

as modeling the log of the variable rather than the variable itself) prior to performing the regression. In addition, after performing regression (and particularly useful for multiple regression with multiple independent variables), a plot of observed versus predicted rain rate can be used to check that the overall pattern was indeed linear (shown later in the results section). If it is not, then that would indicate one or more independent variable has a non-linear relationship with rain rate.

The second assumption is that there is little multicollinearity and singularity among the radar products. Multicollinearity occurs when the independent variables are highly correlated—for instance an R Squared value[1] is ≥ 0.9 and when the tolerance (1 minus R Squared) is less than 0.1 would be indicative of multicollinearity. Alternatively, multicollinearity is revealed when the variance inflation factor (VIF; the inverse of the tolerance) js greater than 10. This will be shown in the results section.

The third assumption is that each of the variables going into the regression is normally distributed. This can be evaluated by checking that the plotted distribution of any variable involved in the regression is normal. Non-normality can also be checked using skewness or kurtosis computations. Should data be found to be non-normal, it may be transformed into a normal distribution using the Box-Cox transformation (Box and Cox, 1964). This assumption was not checked for the present study.

The fourth assumption is that the model is homoscedastic (errors being equally distributed about the regression line). Heteroscedasticity can be revealed by visually examining the predicted versus observed rain rate plot (shown later in results) or by using

[1] The statistical computation of R Squared and its square root, R, should not be confused with the rain rate (abbreviated as *R*).

the test developed by Goldfeld and Quandt (1965). SPSS software does not support the Goldfeld-Quandt Test so further testing of homoscedasticity would need to be addressed in later research beyond the existing plots and brief analysis mentioned in this study.

The fifth and final assumption is that the residuals are not auto-correlated. This can arise if the data being sampled are highly correlated in time. This can occur in time series data. Since both rain gauge data and radar data from a time series, this is a potential issue. The Durbin-Watson test can be used to check for significant auto-correlation (Tillman 1975). If significant auto-correlation exists, the auto-correlation can be used to decrease the N (effectively reducing the number of "repeated" samples), which results in increased p-values.

CHAPTER V

RESULTS AND DISCUSSION

5.1 Statistical Results

With 15 storm events (regression models) and 114 hours of rainfall observations, statistical analysis provides insight into how well the radar variables estimate the rainfall amounts over the Glacial Ridge Prairie. Utilizing the outputs from the single and multiple linear regressions, the resultant data provide an assessment of the warm season rainfall for the first two seasons of dual-pol availability. Furthermore, it is critical to provide further discussion on the individual regression models and the subsequent results to better understand its applicability and also utility for future research projects.

In the process of running the statistical tests, each of the radar variables were separately plotted (not shown) against the rainrate to ensure there existed a linear trend between the independent and dependent variables. Without transforming the variables appropriately, the plots (not shown) based off the data would not follow a linear trend at all. It should be noted that for all statistical tests performed throughout this study with the exception of the initial scatterplots (not shown), variables that were originally transformed remained that way for all of the statistical tests. It was imperative to transform the data accordingly to accurately and appropriately run the statistical tests, beginning with Pearson Correlation and followed by ANOVA.

Table 8 describes the Pearson Correlation of the various regression models to determine the correlation between the various radar variables and the rainfall. It is imperative to determine the correlation between the variables in an effort to verify which of the four radar variables are crucial in estimating rainfall. It is crucial to assess that the independent variables (reflectivity, correlation coefficient, specific differential phase, and differential reflectivity) have some relationship with the dependent variable (rainfall) included in the study. Ideally, a numerical representation of this essential relationship should be ±0.30 (Cohen, 1988). A Pearson Correlation value of 0.30 indicates a weak but positive linear relationship, whereas a Pearson Correlation value of -0.30 indicated a weak negative linear relationship. The positive or negative symbol only shows the relationship between the variable and with no influence on the magnitude. Furthermore, if the Pearson Correlation value is 0, this indicates that there is no direct linear relationship between independent variables and the dependent variable. Pearson Correlation values that range from 0.50-0.99 have a moderate to strong positive linear relationship, while Pearson Correlation values that range from -0.50- -0.99 would have a moderate to strong negative linear relationship. It is also important to note that Pearson Correlation values of 1 represent an ideal positive linear relationship and conversely, Pearson Correlation values of -1 represent an ideal negative linear relationship. As seen in Table 8, all of the independent variables did not exceed the 0.30 threshold for the Pearson Correlation results. However, the Pearson Correlation value for the Z and R was 0.288, which was the closest correlation to the 0.30 threshold. This indicated that there was a weak but positive linear relationship between Z and R. Additionally, the Pearson Correlation values for ρ_{hv}, K_{DP}, and Z_{DR} indicated little to no linear relationship.

Specifically, ρ_{hv} had a Pearson Correlation of -0.097, K_{DP} had a Pearson Correlation value of -0.073, and Z_{DR} had a Pearson Correlation of -0.048. It is critical to identify and evaluate the relationship between the independent variables against each other. The Pearson Correlation values of the other radar variables to ρ_{hv} include a value of 0.164 for K_{DP}, a value of 0.722 for Z, and a value of 0.23 for Z_{DR}. For Z_{DR} and K_{DP}, there was a weak but positive linear relationship to ρ_{hv}, while Z had a strong, positive linear relationship. When investigating the linear relationship of the radar variables to K_{DP}, there was a Pearson Correlation value of -0.012 with Z and a value of 0.092 for Z_{DR}. For Z and K_{DP}, there was little to no linear relationship given its low negative value close to zero. In addition, for Z_{DR} there was also little to no linear relationship since the value was positive and close to zero. Furthermore, it is also important to assess the Pearson Correlation for Z and Z_{DR}, which is 0.112. Again, this indicated a weak and little to no linear relationship given its close proximity to zero. As a result, the data indicated that a majority of the independent variables when compared to their dependent counterpart, showed very little linear relationship with the exception of Z. While these results are not ideal, this is not to say that the data were invalid or not helpful. Rather, it provides the opportunity for further application in future research studies and serves as a framework.

The Pearson Correlation assessment provided an initial review of the four radar variables for estimating rainfall. In order to further understand how the different radar variables influence rainfall and to better understand the five assumptions of linear regression, further statistical analyses such as ANOVA were conducted and assessed for this research study

As seen in Table 4, insight is provided into the overall goodness of fit[2] for each of the regression models by analyzing the Coefficient of Determination (R Squared). Specifically, the R Squared value describes how much of the variance in the dependent variable can be explained by all of the independent variables that together make up the model. It provides key insight into how well a given model may fit the data presented. It is commonly known that R Squared values range from 0-1, with 0 indicating that a given model does not explain any of the variance of the data centered on the mean, whereas, values near or equal to 1 indicate that a given model accounts for most or all of the variance, respectively, and the better the predictor that model will be. Thus, if R squared is larger, the prediction (regression line) will better explain the rain rate data and those points will be distributed along it.

For this study, the R Squared value provides a measure of how well each model explains total variance in observed rainfall. As seen in Table 4, none of the R Squared values provide a clear and strong indication that any of the models fits the regression line exactly since they all explain less than 30% of the variability in rainfall. However, it is still important to mention the specific results and compare between the different models. Four out of 15 models provide the highest R Squared values. For the $R(Z,K_{DP},Z_{DR},\rho_{hv})$ model, the R Squared value is highest at 0.295 thus explaining 29.5% of the variability in rainfall recorded. Three other cases also produced a similar R Squared value including $R(Z,Z_{DR},\rho_{hv})$, $R(Z,K_{DP},\rho_{hv})$, and $R(Z,\rho_{hv})$ with respective R Squared values of 0.292, 0.280, and 0.278. As seen in Table 4, the remaining models have much lower R Squared.

[2] Adjusted R Squared, the Coefficient of Multiple Correlation, R, and Standard Error of the Estimate are also included in this table from the SPSS output for completeness.

The lowest-performing models with the lowest R Squared values, not even reaching 0.01, include R(Z_{DR}), R(K_{DP}), and R($K_{DP,}$ Z_{DR}). The seven remaining models in Table 4 are middle-of-the-road in explaining only between 1% and 10% of the variability in rainfall.

Upon inspection, it is evident that the model utilization of both reflectivity and correlation coefficient were associated with the highest R Squared. Though none of the models produced R Squared values over 0.3, low R Square values do not necessarily mean bad results. Unfortunately, other prior authors did not report their R Squared values in their studies, so while 0.3 might seem low, it is unknown what to expect.

How important was each predictor variable within each model? ANOVA provides information in how likely each regression model explains the variance in rainfall simply due to chance as well as how significant each independent variable is to the overall model. In ANOVA, each independent variable achieves a p-value and those less than 0.050 are typically considered statistically significant, with smaller p-values indicating more statistical significance and less chance of a type-1 error. All output from SPSS for ANOVA testing is included for completeness herein, but only the p-value and statistically significant columns will be further discussed below.

Table 4 already provided information on poorly-performing models. The ANOVA results in Table 5 show that the same poor models with small R squared (from Table 4) have non-significant p-values. For instance, $R(K_{DP})$, $R(K_{DP},Z_{DR})$, and R(Z_{DR}) all have p-values ranging from 0.07 to 0.22 and thus those models have a greater probability that they are explaining the rain variance simply by chance rather than having true predictive value. This means that the values pertaining to specific differential phase and differential reflectivity were, alone and together, no good in predicting rainfall. In

contrast, eight of the regression models have very good p-values of <0.001, indicating a very high statistical significance and these were also the ones that had the highest R squared values in Table 4: $R(Z,K_{DP},Z_{DR},\rho_{hv})$, $R(Z,K_{DP},Z_{DR})$, $R(Z,K_{DP},\rho_{hv})$, $R(Z,Z_{DR},\rho_{hv})$, $R(Z,K_{DP})$, $R(Z,\rho_{hv})$, $R(Z,Z_{DR})$, and R(Z). Four additional models, that had mediocre R Squared values in Table 4, were still found significant in Table 5 with p-values < 0.04, including $R(K_{DP},Z_{DR},\rho_{hv})$ with a significance of 0.034, $R(Z_{DR},\rho_{hv})$ with a value of 0.022, $R(K_{DP},\rho_{hv})$'s significance value of 0.021, and $R(\rho_{hv})$ with its significance value of 0.012.

From the ANOVA testing (Table 5), it is clear that the best regression models had to include reflectivity. Furthermore, models including one or more of correlation coefficient, specific differential phase, and differential reflectivity were helpful when included with reflectivity but not helpful in any combination in models that were missing reflectivity. Furthermore, from Table 4, one can surmise that the most important variable after reflectivity is correlation coefficient as those models have the largest R Squared values. However, Table 5 also shows that models missing correlation coefficient but having differential reflectivity and/or specific differential phase joining reflectivity were also highly statistically significant. Next, to determine the relative importance of each independent variable to the different regression models, an examination of the regression models' statistical coefficients must be assessed.

Table 6 provides the statistical coefficients of the regression models with computed correlations. The entire output from SPSS for the statistical coefficients for the regression models are included but only the unstandardized coefficient *B* and the p-value column ("Sig") will be further discussed for the purpose of this study. In this context, the p-value in Table 6 shows how well each of the radar variables individually explain rain

rate for the combination of independent variables in that given regression model. As before a p-value of less than 0.05 is considered statistically significant (only 5% chance of a type-1 error) and $p<0.001$ is considered very significant (highly unlikely for a type-1 error). The unstandardized coefficient *B* provided the values for the resultant equations to determine the rainfall are provided in Table 9. It should be noted that the *B* values were applied after transforming the equations to a power law form from an initial logarithmic form. The slopes of each independent variable show up in the associated exponent, with the power law version of the equation. For the Z_{DR} exponents that are extremely small, based on taking the *B* values directly from Table 6, the value for Z_{DR} gets mathematically solved to be 1. Obviously, if the variable were to actually have no effect, the zero slope is plugged into the exponent and any variable to zero power is like just multiplying by 1 (no effect on rain rate). This is an indication of that even though the p-value can be of statistical significance, the *B* value constrains the variable to be of no beneficial use in the regression model equations to determine rainfall.

Delving deeper into Table 6, let's first consider the models that include reflectivity. The first regression model $R(Z,K_{DP},Z_{DR},\rho_{hv})$ has all variables very significant with a p-value < 0.001 with the exception of specific differential phase ($p=0.107$). Conversely, the regression model of $R(Z,K_{DP},Z_{DR})$ only has the constant value and reflectivity $p<0.001$ with specific differential phase and differential reflectivity values of $p=0.108$ and $p=0.058$ respectively. For the $R(Z,Z_{DR},\rho_{hv})$ regression model, all variables are very significant with differential reflectivity $p=0.001$ and others $p<0.001$. In both the $R(Z,K_{DP},\rho_{hv})$ and $R(Z,K_{DP})$ regression models, specific differential phase is the only variable that is not statistically significant ($p= 0.236$ and 0.074, respectively) while

the constant and other variables were very significant ($p<0.001$). The $R(Z,\rho_{hv})$, $R(Z)$, and $R(Z,Z_{DR})$ regression models had all variables very significant ($p <0.001$) except for differential reflectivity that was only significant ($p= 0.035$). One will note by comparing the model results, that the individual performance of certain variables differed depending upon what other variables were included in the equation.

Regarding the models that performed poorly in Tables 4 and 5 because they do not include reflectivity, but still included correlation coefficient, for the $R(K_{DP},Z_{DR},\rho_{hv})$ regression model, the only variables being statistically significant are the constant and correlation coefficient. The variables specific differential phase and differential reflectivity, similar to the previous model are not of significance statistically with p-values of 0.182 and 0.322 respectively. For the regression model of $R(Z_{DR},\rho_{hv})$, the constant value was very significant ($p < 0.001$), the correlation coefficient was significant ($p= 0.013$), but differential reflectivity was not significant ($p=0.24$). Also, the regression model $R(K_{DP},\rho_{hv})$ had the constant value very significant ($p<0.001$), correlation coefficient significant ($p=0.034$), and specific differential phase not significant ($p=0.153$). For the $R(\rho_{hv})$ model, the constant variable was very significant ($p<0.001$) and correlation coefficient significant ($p<0.012$). These last four models show that while the overall equation predicted poorly without reflectivity, the constant and the correlation coefficient were still much more important to these models than were differential reflectivity or specific differential phase.

For the models without correlation coefficient and without reflectivity, what were the most important variables? For $R(K_{DP},Z_{DR})$, $R(K_{DP})$, and $R(Z_{DR})$, only the constant variable was statistically significant ($p<0.05$).

What about some of the assumptions regarding multiple linear regression? The above results may be corrupt to a degree if one does not check that the multiple linear regression assumptions were all satisfied. Were those assumptions satisfied? One important assumption is to check that there is little to no multicollinearity.

Table 7 showcases the collinearity statistics of the multiple linear regression models. These statistics are important to determine if multicollinearity exists. The tolerance value indicates how much of the variability of each of the independent variables cannot be explained from the other independent variables. Recall that tolerance is mathematically defined as one minus the R Squared value and VIF is just the inverse of the tolerance. Tolerance values less than 0.10 (or Variance Inflation Factor, VIF, greater than 10) indicate that multicollinearity might exist. As seen in Table 7, none of the multiple regression models have multicollinearity. Specifically, three of the regression models $R(Z,K_{DP},Z_{DR},\rho_{hv})$, $R(Z,\rho_{hv})$, and $R(Z,K_{DP},\rho_{hv})$ had tolerance values greater than 0.4 for each independent variable while all other models had tolerance values greater than 0.9. (Table 7). Thus, it is evident by the results in Table 7 that none of the values indicate multicollinearity. The results presented in Table 7 fulfill one of the multiple linear regression assumptions of little to no multicollinearity. To further fulfill the assumptions of multiple linear regression, it is also important to assess the correlation between the variables for each regression model.

The Durbin-Watson test was necessary to perform in order to test the multiple linear regression assumption of autocorrelation, since the data gathered were part of a time series. The range of outputs commonly found from running a Durbin-Watson test range from 0 to 4. Values that are under 2 tend to have a positive autocorrelation while

values above 2 tend to have a negative autocorrelation. Ideally, it is best to have the Durbin-Watson results centered around the value of 2. According to Field (2013), generally any Durbin-Watson values <1 or >3 could be problematic, implying the assumption of autocorrelation is invalid related to the multiple linear regression criteria. As seen in Table 10, none of the regression models have an ideal Durbin-Watson value, and more specifically, two of the regression models had values of concern ($R(Z,Z_{DR})$'s value of 0.982 and R(Z)'s value of 0.983). Even though these two regression models are <1, they were still close overall to the Durbin-Watson value of 1, and for this study the autocorrelation assumption would not be considered invalid for these regression models. However, two of the regression models, $R(Z,K_{DP},Z_{DR},\rho_{hv})$ and $R(Z,Z_{DR},\rho_{hv})$ had the highest Durbin-Watson values with 1.287 for both. The regression model for $R(Z,K_{DP},\rho_{hv})$ had the third highest Durbin-Watson result with a value of 1.279 followed by $R(Z,\rho_{hv})$ with 1.198. The regression cases of $R(Z,K_{DP},Z_{DR})$, $R(Z,K_{DP})$, and $R(K_{DP})$ values around 1.1, specifically 1.106, 1.107, and 1.114 respectively. The regression models of both $R(K_{DP},\rho_{hv})$ and $R(K_{DP},Z_{DR})$ had Durbin-Watson values of 1.075. $R(Z_{DR},\rho_{hv})$ and $R(\rho_{hv})$ had similar Durbin-Watson values of 1.030 and 1.031 respectively. Furthermore, the case of $R(Z_{DR})$ had a value of 1.013 and $R(K_{DP},Z_{DR},\rho_{hv})$'s value of 1.073. For all of the regression models, their Durbin-Watson values were below the threshold value of 2, implying positive autocorrelation based off the data but not less than 1.0 (Field's "problematic" trigger) (Field, 2013). The Durbin-Watson test suggests that the data were indeed autocorrelated. Again, if one were to account for the autocorrelation by reducing N (repeated data), the favorable p-values reported earlier would not have looked as favorable. To accomplish this (though not performed in this study

specifically), one would apply an equation that adjusts these values accordingly with the end result of reducing the impact of autocorrelation.

To further understand one other assumption related to linear regression, linearity, Figures 7-21 show the regression scatterplots for each of the regression models. The x-axis is the standard predicted value in this case the rainfall amount based on the given model of radar variables. The y-axis is the standard residual value of the model and can be alternatively described as the observed value minus the predicted value. A perfect prediction would have all data points distributed along the y=0 line, implying the difference between the observed and predicted rainfall be zero. When assessing scatterplots it is important to discuss the direction, form, and strength of the given plots. The direction can be either a positive or negative association (gradient). The form of a scatterplot follows a perfect linear association (all points are plotted on the trend line), linear association (some of the points but not all are on the trend line, all points though follow the trend whether it be positive or negative), no linear association (points follow a trend like a bell curve but are not linear), or no association at all (points are scattered but with no overall trend) depending on the data. Also, the strength of a scatterplot is based upon the correlation also known as the clustering of the data points. By utilizing the Coefficient of Multiple Correlation (R value) as seen in Table 4, the correlation strength can identified as either near zero, weak, moderate, or strong. R values roughly under 0.20 were identified as near zero and this applied to seven models. R values between 0.20-0.49 were identified as weak and this applied to four models. R values ranging from 0.50-0.69 were classified as moderate and four models fell in that category. There were

no R values above 0.69 which would have been categorized as having a strong correlation strength.

As seen in Figure 7, there was a negative association between the predicted rainfall and the difference between the model and the observed. When evaluating the form of Figure 7, it is evident that there was a linear association. In addition, when examining the strength of the data, there was a moderate correlation (R value of 0.543, as seen in Table 4). Figure 8 has an overall negative association between the predicted rainfall and the difference between the model and the observed. However, when evaluating the form of the scatterplot, the points did not have a distinct linear association. Furthermore, the strength of the correlation was weak, with a R value of 0.288 (Table 4). Additionally, Figure 9 also has a negative association with no apparent linear association either. The scatterplot also indicated a weak correlation with a R value of 0.299 (Table 4). As seen in Figure 10, there was a negative gradient of the predicted rainfall versus the differences in the observed and predicted values. This scatterplot indicated no association in terms of its linear form. The strength of the scatterplot was near zero with a R value of 0.084 (Table 4). In contrast, the scatterplot in Figure 11 had a slight negative gradient and no distinct linear association. This scatterplot also had a near zero strength with a R value of 0.073 (Table 4). As seen in Figure 12, there was a negative association between the predicted rainfall and the difference in the observed and predicted. The form had no linear association and the strength of the correlation was weak with an R value of 0.306 (Table 4). Figure 13 also had a negative association gradient and no linear form. It also had correlation strength near zero (R value of 0.120 as seen in Table 4). When assessing the direction of the scatterplot in Figure 14, it was

apparent that there was a negative gradient and a slight linear association. Furthermore, the correlation strength was moderate with a R value of 0.529 (Table 4). As seen in Figure 15, there was a negative association between the predicted rainfall and the difference in the observed and predicted and the form had a slight linear association. The strength of the correlation was moderate with a R value of 0.540 (Table 4). Additionally, Figure 16 had a negative direction and no associated form, while the correlation strength was near zero (R value of 0.113 as seen in Table 4). As seen in Figure 17, once again there is a negative direction and the potential for a slight linear association. The strength of the correlation is weak with an R value of 0.297 as seen in Table 4. In contrast, Figure 18 provides a unique display of the data with a slight negative direction and no linear association. Also, in Figure 18 it was apparent that the strength of the correlation was near zero with an R value of 0.107 (Table 4). Figure 19 also presented a negative direction of the data and a slight linear association. The correlation strength was moderate (R value of 0.527 as seen in Table 4). The scatterplot in Figure 20 was similar to that of Figure 18. Here, the data had a slight negative direction with no linear association. The strength of the correlation was also near zero with an R value of 0.097 (Table 4). Figure 21 presented the worst results and basically shows no predictive value as the dots are not varying along the x-axis. Remember that the three distinct characteristics of direction, form, and strength are needed when interpreting such results. Here, the data presents no distinct direction given the majority of the points align on the x=0 value. . The form was not apparent and the correlation strength was reported near zero with a R value of 0.048 (Table 4). As seen in the discussion of the scatterplots, there is quite a lot of variability in terms of strength and form given the various independent

variables utilized. All of the scatterplots had either a slight or distinct negative direction implying an inverse relationship between the difference of the predicted and observed and the predicted based on the regression model Finally, to test the last assumption of linear regression analysis, it is important to assess for homoscedasticity (as seen in subsequent figures mentioned below).

Figures 22-36 represent the scatterplots of each of the regression models for assessing homoscedasticity. In analyzing for homoscedasticity it is imperative to identify any deviations from the overall trend and clustering of points. Generally, when evaluating the scatterplots for homoscedasticity the points are tightly clustered in a linear manner along the entire x-axis, whereas scatterplots that present points in a fan-like manner are classified as heteroscadastic because the errors get larger in one direction along the x-axis. Recall that the assumption in linear regression is that the errors are homoscedastic. In order to fully understand and assess for homoscedasticity, a Goldfeld-Quandt test would need to be conducted. However, unfortunately, SPSS does not support the statistical test. Thus, for the purpose of this study, scatterplots were created to provide some degree of analysis related to homoscedasticity. While this is not ideal, it is better to provide a brief overview than nothing at all in order to assess the five assumptions of linear regression.

As seen in Figure 22, the scatterplot points appeared to follow a general fan-like distribution, initially the points clustered around the smaller x-axis values and expanded in point distribution the higher on the x-axis. This would suggest heteroscedasticity for the regression model $R(Z,K_{DP},Z_{DR},\rho_{hv})$. Figure 23 for the regression model of $R(K_{DP},Z_{DR},\rho_{hv})$, conversely displayed a scatterplot with points tightly bound and

relatively close to each other, suggesting homoscedasticity. The regression model of $R(Z,Z_{DR},\rho_{hv})$ in Figure 24 showed a similar heteroscedastic trend of point distribution as seen in Figure 22. Additionally, Figure 25 representing the regression model of $R(Z,K_{DP},Z_{DR})$ was more like Figure 23 in displaying a homoscedastic trend given the points are tightly clustered from the lower bounds of the x-axis to the upper bounds. The $R(Z,K_{DP},\rho_{hv})$ regression model shown in Figure 26 has quite the fan-like distribution among the points. This implied that the data points were in a heteroscadastic pattern. Figure 27 of the regression model $R(Z_{DR},\rho_{hv})$ displayed a homoscedastic trend with the points tightly clustered in a linear manner. The $R(Z,\rho_{hv})$ regression model in Figure 28 had a heteroscedastic pattern to it with the points again fanning out from the lower to upper bounds of the x-axis. Conversely, the regression model of $R(Z,K_{DP})$ in Figure 29 had tightly bound array of points implying homoscedasticity. Figure 30 displayed the regression model $R(Z,Z_{DR})$ and had the points generally following a heteroscedastic form. Even with the outliers shown in Figure 31 of the regression model $R(K_{DP},Z_{DR})$, the scatterplot overall indicated a homoscedasticity of the points. In addition, $R(K_{DP},\rho_{hv})$ displayed in Figure 32 also followed a homoscedastic trend with the points tightly clustered together in a linear manner across the graph. Similarly, the regression model of $R(\rho_{hv})$ in Figure 33 also showed a homoscedastic pattern of the points. Conversely, as shown in Figure 34, the pattern of points followed a heteroscedastic pattern for the regression model of R(Z). The regression model of $R(Z_{DR})$ displayed in Figure 35 had tight clustering of the points implying homoscedasticity. Finally, as seen in Figure 36, $R(K_{DP})$ regression model followed a homoscedastic trend with a close clustering of points following in a near-linear manner. In summary the heteroskedastic patterns were seen in

R(Z), R(Z,Z_{DR}), R(Z,ρ_{hv}), R(Z,K_{DP},ρ_{hv}), R(Z,Z_{DR},ρ_{hv}), and R(Z,K_{DP},Z_{DR},ρ_{hv}) which happen to also have the highest R Squared values. In contrast, the homoscedastic patterns were seen in all other models with the lower R Squared values: R(K_{DP}), R(Z_{DR}), R(ρ_{hv}), R(K_{DP},ρ_{hv}), R(K_{DP},Z_{DR}), R(Z,K_{DP}), R(Z_{DR},ρ_{hv}), R(Z,K_{DP},Z_{DR}), and R(K_{DP},Z_{DR},ρ_{hv}). There is not a clear indication that a specific radar variable is contributing to heteroscedasticity or homoscedasticity. Rather, it is seen that those models that do the best in predicting rainrate are also the models that are heteroskedastic. Thus, there may unfortunately be worse errors due to heteroscedasticity precisely in the models where we would wish they would not occur –affecting the significance (p-values) of the regression equations as well as individual variables within the equations.

Finally, as part of the output from SPSS, P-P plots were created. These plots (not shown) are a way to address the non-normality assumption. It was evident upon review of the P-P plots that no significant deviations from the normality existed. With this conclusion, the remainder of the statistical tests were performed with other plots analyzed in regards to the linear regression assumptions.

Overall, the numerous figures and tables provided thoroughly discussed the potential for computing an equation to determine rainfall based on the combination of different regression models of the radar variables. Since, there was no clear indication of an equation that will satisfy all the linear regression criteria required, it is important to discuss the limitations of this study and overall summary of this study.

In an effort to establish a relationship for the GRNWR between the radar variables and the rainfall rate based off the gauges, it is useful to compare these to the relationships the NWS used operationally: convective, tropical, Marshall-Palmer, east-

cool stratiform, and west-cool stratiform respectively. These five main z-R relationships (R in mm h^{-1} and z in mm^6 mm^{-3}) are as follows: deep convection ($z=300\ R^{1.4}$), tropical convective systems ($z=250\ R^{1.2}$), general stratiform ($z=200\ R^{1.6}$), winter rain east of the Continental Divide ($z=130\ R^{2.0}$), and winter rain west of the Continental Divide ($z=75\ R^{2.0}$). While this form of the equation is shown, the equation is inverted to solve for R (mm h^{-1}). Thus, for deep convection $R = (z/300)^{1/1.4}$ or, after inverting the equation, $R = 0.017\ z^{0.714}$. One may convert Z (dBZ) to z ($mm^6\ m^{-3}$) for use in this equation using $z = [10^{Z/10}]$. These relationships are chosen by forecasters at each weather forecast office to effectively sample and quantify precipitation. In light of the recent upgrades to dual-polarization, quantitative precipitation estimations also utilize dual-pol products.

Since the upgrade to polarimetric, the NWS also uses a quantitative precipitation estimation algorithm consisting of four location-dependent equations for computing rain rate. The first is the traditional convective z-R relationship (already inverted) with $R(z)=(0.017)z^{0.714}$ with z ($mm^6\ m^{-3}$) and R (mm h^{-1}) (Prentice, 2016). Converting the leading constant for consistency with linear regression results herein gives $R(z)=10^{-1.77}z^{0.714}$ (Prentice, 2016). The second equation combines differential reflectivity and reflectivity to handle cool season and warm season deep convection. That equation is $R(z,z_{DR})=(0.0142)z^{0.77}\ z_{dr}^{-1.67}$ with z ($mm^6\ m^{-3}$) and z_{dr} (nondim) (Prentice, 2016). Converting the leading constant for consistency with results herein, $R(z,z_{dr})=(10^{-1.84})\ z^{0.77}\ z_{dr}^{-1.67}$ (Prentice, 2016). A third equation is used for warm seasons and tropical events dominated by warm rain processes: $R(z,z_{DR})=(0.0067)z^{0.927}z_{dr}^{-3.43}$ with z ($mm^6\ m^{-3}$) and z_{dr} (nondim) (Prentice, 2016). Adjusting the leading constant, $R(z,z_{dr})=(10^{-2.174})\ z^{0.927}z_{dr}^{-3.43}$ (Prentice, 2016). Lastly, the fourth rain estimation equation is based upon specific

differential phase when hail has been identified and known as $R(K_{DP})=44.0|K_{DP}|^{0.822}\text{sign}(K_{DP})$ with K_{DP} (deg km^{-1}) (Prentice, 2016). Adjusting the leading constant for comparison with results herein, $R(K_{DP})=10^{1.64}|K_{DP}|^{0.822}\text{sign}(K_{DP})$. The quantitative precipitation algorithm has built in logic to handle when this last equation is negative (Prentice, 2016). In deriving new regression models based off of this research study, initial testing was performed to validate and determine if the rainrate from these equations and the equations from Giangrande and Ryzkhov (2003) were in agreement to the observed rainfall and the derived equations from the regression models of this study (not shown). It should be noted that an initial assessment was undertaken to see if the equations compared well to the observed rainfall based on the radar variable values from this study. The derived equations from the regression models herein did not compare well to the other equations used by the NWS and Giangrande and Ryzkhov (2003). Further research should be addressed to obtain higher rainfall amounts to more appropriately sample and refine the regression models to do a more accurate comparison.

Table 4. Regression summary of the Coefficient of Multiple Correlation (R), Coefficient of Determination (R Squared), Adjusted R Squared, and Standard Error of the Estimate (rounded for display purposes).

Regression Model	R	R Sq.	Adj. R Sq.	Std. Error of the Est.
$R(Z,K_{DP},Z_{DR},\rho_{hv})$	.543	.295	.290	.355
$R(Z,K_{DP},Z_{DR})$	.306	.093	.089	.403
$R(K_{DP},Z_{DR},\rho_{hv})$	.120	.014	.010	.420
$R(Z,K_{DP},\rho_{hv})$	.529	.280	.276	.359
$R(Z,Z_{DR},\rho_{hv})$	.540	.292	.288	.356
$R(Z,K_{DP})$	.297	.088	.085	.404
$R(Z_{DR},\rho_{hv})$	.107	.012	.009	.420
$R(Z,\rho_{hv})$	.527	.278	.276	.359
$R(K_{DP},\rho_{hv})$	.113	.013	.010	.420
$R(K_{DP},Z_{DR})$	.084	.007	.004	.421
$R(Z,Z_{DR})$	.299	.089	.087	.403
$R(Z)$	.288	.083	.082	.404
$R(K_{DP})$	.073	.005	.004	.421
$R(\rho_{hv})$	.097	.009	.008	.420
$R(Z_{DR})$	.048	.002	.001	.422

Table 5. Analysis of variance (ANOVA) summary where the rain gauge is the dependent variable and the various radar variables are the independent variables. The "Significant Value" is also known as the p-value and is the probability that the model would explain the rain variability by chance (probability of type-1 error).

Regression Model	Sum of Squares	Degrees of Freedom (df)	Mean Square	F	Significant-Value	Significant?
$R(Z,K_{DP},Z_{DR},\rho_{hv})$				62.10	<.001	Yes, Very
Regression	31.380	4	7.845			
Residual	75.034	594	.126			
$R(Z,K_{DP},Z_{DR})$				20.43	<.001	Yes, Very
Regression	9.940	3	3.313			
Residual	96.474	595	.162			
$R(K_{DP},Z_{DR},\rho_{hv})$				2.92	0.034	Yes
Regression	1.541	3	.514			
Residual	104.872	595	.176			
$R(Z,K_{DP},\rho_{hv})$				77.00	<.001	Yes, Very
Regression	29.759	3	9.920			
Residual	76.654	595	.129			
$R(Z,Z_{DR},\rho_{hv})$				85.97	<.001	Yes, Very
Regression	32.659	3	10.886			
Residual	79.270	626	.127			
$R(Z,K_{DP})$				28.73	<.001	Yes, Very
Regression	9.356	2	4.678			
Residual	97.058	596	.163			
$R(Z_{DR},\rho_{hv})$				3.86	.022	Yes
Regression	1.361	2	.681			
Residual	116.797	662	.176			
$R(Z,\rho_{hv})$				120.68	<.001	Yes, Very
Regression	31.111	2	15.556			
Residual	80.819	627	.129			

Table 5 (Cont.). Analysis of variance (ANOVA) summary where the rain gauge is the dependent variable and the various radar variables are the independent variables. The "Significant Value" is also known as the p-value and is the probability that the model would explain the rain variability by chance (probability of type-1 error).

Regression Model	Sum of Squares	Degrees of Freedom (df)	Mean Square	F	Significant-Value	Significant?
$R(K_{DP},\rho_{hv})$				3.88	.021	Yes
Regression	1.368	2	.684			
Residual	105.046	596	.176			
$R(Z,Z_{DR})$				30.803	<.001	Yes, Very
Regression	10.014	2	.5.007			
Residual	101.916	627	.163			
R(Z)				56.851	<.001	Yes, Very
Regression	9.292	1	9.292			
Residual	102.638	628	.163			
$R(K_{DP})$				3.238	.072	No
Regression	.574	1	.574			
Residual	105.840	597	.177			
$R(\rho_{hv})$				6.338	.012	Yes
Regression	1.119	1	1.119			
Residual	117.217	664	.177			
$R(Z_{DR})$				1.510	.220	No
Regression	.269	1	.269			
Residual	117.890	663	.178			

Table 6. Coefficients of regression models where the rain gauge is the dependent variable and the various radar variables are the independent variables. Singular regression does not provide intercorrelations among independent variables.

Regression Model	Unstandardized Coefficients		t	Sig.	95.0% Confidence Interval for B		Correlations Part
	B	Std. Error			Lower Bound	Upper Bound	
$R(Z,K_{DP},Z_{DR},\rho_{hv})$							
Constant	.577	.037	15.571	<.001	.504	.650	
Z	.285	.019	15.369	<.001	.249	.322	.530
K_{DP}	.016	.010	1.616	.107	-.003	.035	.056
Z_{DR}	<.001	<.001	-3.582	<.001	<.001	<.001	-.123
ρ_{hv}	-.771	.059	-13.03	<.001	-.887	-.655	-.449
$R(Z,K_{DP},Z_{DR})$							
Constant	.249	.031	8.079	<.001	.188	.309	
Z	.107	.014	7.527	<.001	.079	.135	.294
K_{DP}	-.017	.011	-1.608	.108	-.039	.004	-.063
Z_{DR}	>-.001	<.001	-1.897	.058	<.001	<.001	-.074
$R(K_{DP},Z_{DR},\rho_{hv})$							
Constant	.514	.044	11.818	<.001	.429	.600	
K_{DP}	-.015	.011	-1.338	.182	-.038	.007	-.054
Z_{DR}	>-.001	<.001	-.992	.322	<.001	<.001	-.040
ρ_{hv}	-.100	.047	-2.114	.035	-.192	-.007	-.086

Table 6 (Cont.). Coefficients of regression models where the rain gauge is the dependent variable and the various radar variables are the independent variables. Singular regression does not provide intercorrelations among independent variables.

Regression Model	Unstandardized Coefficients		t	Sig.	95.0% Confidence Interval for B		Correlations Part
	B	Std. Error			Lower Bound	Upper Bound	
$R(Z,K_{DP},\rho_{hv})$							
Constant	.575	.037	15.376	<.001	.502	.649	
Z	.275	.059	14.845	<.001	.239	-.631	.517
K_{DP}	.012	.010	1.187	.236	-.008	.031	.041
ρ_{hv}	-.747	.019	-12.58	<.001	-.864	.311	-.438
$R(Z,Z_{DR},\rho_{hv})$							
Constant	.555	.034	16.521	<.001	.489	.621	
Z	.279	.018	15.739	<.001	.244	.314	.529
Z_{DR}	<.001	<.001	-3.497	.001	<.001	<.001	-.118
ρ_{hv}	-.746	.056	-13.37	<.001	-.856	-.636	-.450
$R(Z,K_{DP})$							
Constant	.254	.031	8.257	<.001	.193	.314	
Z	.104	.014	7.344	<.001	.076	.132	.287
K_{DP}	-.019	.011	-1.791	.074	-.040	.002	-.070
$R(Z_{DR},\rho_{hv})$							
Constant	.535	.039	13.882	<.001	.459	.611	
Z_{DR}	>-.001	<.001	-1.176	.240	<.001	<.001	-.045
ρ_{hv}	-.110	.044	-2.488	.013	-.197	-.023	-.096
$R(Z,\rho_{hv})$							
Constant	.559	.034	16.507	<.001	.492	.625	
Z	.271	.018	15.269	<.001	.236	.306	.518
ρ_{hv}	-.730	.056	-13.01	<.001	-.840	-.620	-.442

Table 6 (Cont.). Coefficients of regression models where the rain gauge is the dependent variable and the various radar variables are the independent variables. Singular regression does not provide intercorrelations among independent variables.

Regression Model	Unstandardized Coefficients		t	Sig.	95.0% Confidence Interval for B		Correlations Part
	B	Std. Error			Lower Bound	Upper Bound	
$R(K_{DP},\rho_{hv})$							
Constant	.514	.044	11.824	<.001	.429	.600	
K_{DP}	-.016	.011	-1.432	.153	-.039	.006	-.058
ρ_{hv}	-.100	.047	-2.122	.034	-.193	-.007	-.086
$R(K_{DP},Z_{DR})$							
Constant	.432	.020	21.977	<.001	.393	.471	
Z_{DR}	>-.001	<.001	-1.006	.315	<.001	<.001	-.041
K_{DP}	-.019	.011	-1.699	.090	-.041	.003	-.069
$R(Z,Z_{DR})$							
Constant	.262	.029	9.087	<.001	.206	.319	
Z	.107	.014	7.749	<.001	.080	.135	.295
Z_{DR}	>-.001	<.001	-2.108	.035	<.001	<.001	-.080
R(Z)							
Constant	.270	.029	9.377	<.001	.213	.326	
Z	.104	.014	7.540	<.001	.077	.131	.288
$R(K_{DP})$							
Constant	.432	.020	21.972	<.001	.393	.471	
K_{DP}	-.020	.011	-1.799	.072	-.042	.002	-.073
$R(\rho_{hv})$							
Constant	.537	.038	13.948	<.001	.461	.613	
ρ_{hv}	-.111	.044	-2.518	<.012	-.198	-.024	-.097

Table 6 (Cont.). Coefficients of regression models where the rain gauge is the dependent variable and the various radar variables are the independent variables. Singular regression does not provide intercorrelations among independent variables.

Regression Model	Unstandardized Coefficients		t	Sig.	95.0% Confidence Interval for B		Correlations Part
	B	Std. Error			Lower Bound	Upper Bound	
$R(Z_{DR})$							
Constant	.448	.016	27.381	<.001	.416	.480	
Z_{DR}	>-.001	<.001	-1.229	.220	<.001	<.001	-.048

Table 7. Tolerance and variance inflation factor (VIF) statistics for multiple regression models where the raingauge is the dependent variable and the various radar products are the independent variables.

Regression Model	Collinearity Statistics	
	Tolerance	VIF
$R(Z,K_{DP},Z_{DR},\rho_{hv})$		
Z	.449	2.225
K_{DP}	.924	1.082
Z_{DR}	.967	1.034
ρ_{hv}	.443	2.257
$R(Z,K_{DP},Z_{DR})$		
Z	.987	1.013
K_{DP}	.991	1.009
Z_{DR}	.979	1.022
$R(K_{DP},Z_{DR},\rho_{hv})$		
K_{DP}	.965	1.036
Z_{DR}	.991	1.009
ρ_{hv}	.973	1.028

Table 7 (Cont.). Tolerance and variance inflation factor (VIF) statistics for multiple regression models where the raingauge is the dependent variable and the various radar products are the independent variables.

Regression Model	Collinearity Statistics	
	Tolerance	VIF
$R(Z,K_{DP},\rho_{hv})$		
Z	.461	2.170
K_{DP}	.938	1.066
ρ_{hv}	.461	2.229
$R(Z,Z_{DR},\rho_{hv})$		
Z	.469	2.131
Z_{DR}	.981	1.020
ρ_{hv}	.475	2.106
$R(Z,K_{DP})$		
Z	1.000	1.000
K_{DP}	1.000	1.000
$R(Z_{DR},\rho_{hv})$		
Z_{DR}	.999	1.001
ρ_{hv}	.999	1.001
$R(Z,\rho_{hv})$		
Z	.478	2.091
ρ_{hv}	.478	2.091
$R(K_{DP}, \rho_{hv})$		
K_{DP}	.973	1.028
ρ_{hv}	.973	1.028
$R(K_{DP},Z_{DR})$		
K_{DP}	.991	1.009
Z_{DR}	.991	1.009
$R(Z,Z_{DR})$		
Z	.988	1.013
Z_{DR}	.988	1.013

Table 8. Pearson Correlation results for regression models including the following transformed variables: *$log_{10}(R)$, ρ_{hv}, $log_{10}|K_{DP}|sign(K_{DP})$, Z/10, and Z_{DR}.*

Variable	$log_{10}(R)$	ρ_{hv}	$log_{10}\|K_{DP}\|$ sign(K_{DP})	Z/10	Z_{DR}
$log_{10}(R)$	1.00	-.097	-.073	.288	-.048
ρ_{hv}		1.000	.164	.722	.023
$log_{10}\|K_{DP}\|sign(K_{DP})$			1.000	-.012	.092
Z/10				1.000	.112
Z_{DR}					1.000

Table 9. Power law form of regression equations where the rain rate (mm h^{-1}) is the dependent variable and the various radar variables are the independent variables: z (mm^6 m^{-3}), K_{DP} (deg km^{-1}), and z_{DR} (nondim).

Regression Model	Rainfall Equation following form shown in Eq. 12
$R(z,K_{DP},z_{DR},\rho_{hv})$	$R=10^{.577}\ z^{.285}\ z_{dr}^{0}\ 10^{-.771\rho hv}\ \lvert K_{DP}\rvert^{016}\ \mathrm{sign}(K_{DP})$
$R(z,K_{DP},z_{DR})$	$R=10^{.249}\ z^{.107}\ z_{dr}^{-6.869E-5}\ \lvert K_{DP}\rvert^{-.017}\ \mathrm{sign}(K_{DP})$
$R(K_{DP},z_{DR},\rho_{hv})$	$R=10^{.514}\ z_{dr}^{-3.720E-5}\ 10^{-.100\ \rho hv}\ \lvert K_{DP}\rvert^{-.015}\ \mathrm{sign}(K_{DP})$
$R(z,K_{DP},\rho_{hv})$	$R=10^{.575}\ z^{.275}\ 10^{-.747\ \rho hv}\ \lvert K_{DP}\rvert^{012}\ \mathrm{sign}(K_{DP})$
$R(z,z_{DR},\rho_{hv})$	$R=10^{.555}\ z^{.279}\ z_{dr}^{0}\ 10^{-.746\ \rho hv}$
$R(z,K_{DP})$	$R=10^{.254}\ z^{.104}\ \lvert K_{DP}\rvert^{-.019}\ \mathrm{sign}(K_{DP})$
$R(z_{DR},\rho_{hv})$	$R=10^{.535}\ z_{dr}^{-4.169E-5}\ 10^{-.110\ \rho hv}$
$R(z,\rho_{hv})$	$R=10^{.559}\ z^{.271}\ 10^{-.730\ \rho hv}$
$R(K_{DP},\rho_{hv})$	$R=10^{.514}\ 10^{-.100\ \rho hv}\ \lvert K_{DP}\rvert^{-.016}\ \mathrm{sign}(K_{DP})$
$R(K_{DP},z_{DR})$	$R=10^{.432}\ z_{dr}^{-3.758E-5}\ \lvert K_{DP}\rvert^{-.019}\ \mathrm{sign}(K_{DP})$
$R(z,z_{DR})$	$R=10^{.262}\ z^{.107}\ z_{dr}^{-7.41E-5}$
$R(z)$	$R=10^{.270}\ z^{.104}$
$R(K_{DP})$	$R=10^{.432}\ \lvert K_{DP}\rvert^{-.020}\ \mathrm{sign}(K_{DP})$
$R(\rho_{hv})$	$R=10^{.537}\ 10^{-.111\ \rho hv}$
$R(z_{dr})$	$R=10^{.448}\ z_{dr}^{-4.374E-5}$

Table 10. Durbin-Watson summary where the rain gauge is the dependent variable and the radar products denoted are the independent variables. Items from other tables are repeated, and the table has been sorted in descending order of R Squared.

Regression Model	Durbin-Watson	Hetero-skedastic	R Squared (from Table 4)	p-value (from Table 5)
$R(Z,K_{DP},Z_{DR},\rho_{hv})$	1.287	YES	.295	<.001
$R(Z,Z_{DR},\rho_{hv})$	1.287	YES	.292	<.001
$R(Z,K_{DP},\rho_{hv})$	1.279	YES	.280	<.001
$R(Z,\rho_{hv})$	1.198	YES	.278	<.001
$R(Z,K_{DP},Z_{DR})$	1.106		.093	<.001
$R(Z,Z_{DR})$	.982	YES	.089	<.001
$R(Z,K_{DP})$	1.107		.088	<.001
$R(Z)$	.983	YES	.083	<.001
$R(K_{DP},Z_{DR},\rho_{hv})$	1.073		.014	0.034
$R(K_{DP},\rho_{hv})$	1.075		.013	.021
$R(Z_{DR},\rho_{hv})$	1.030		.012	.022
$R(\rho_{hv})$	1.031		.009	.012
$R(K_{DP},Z_{DR})$	1.075		.007	<.001
$R(K_{DP})$	1.114		.005	.072
$R(Z_{DR})$	1.013		.002	.220

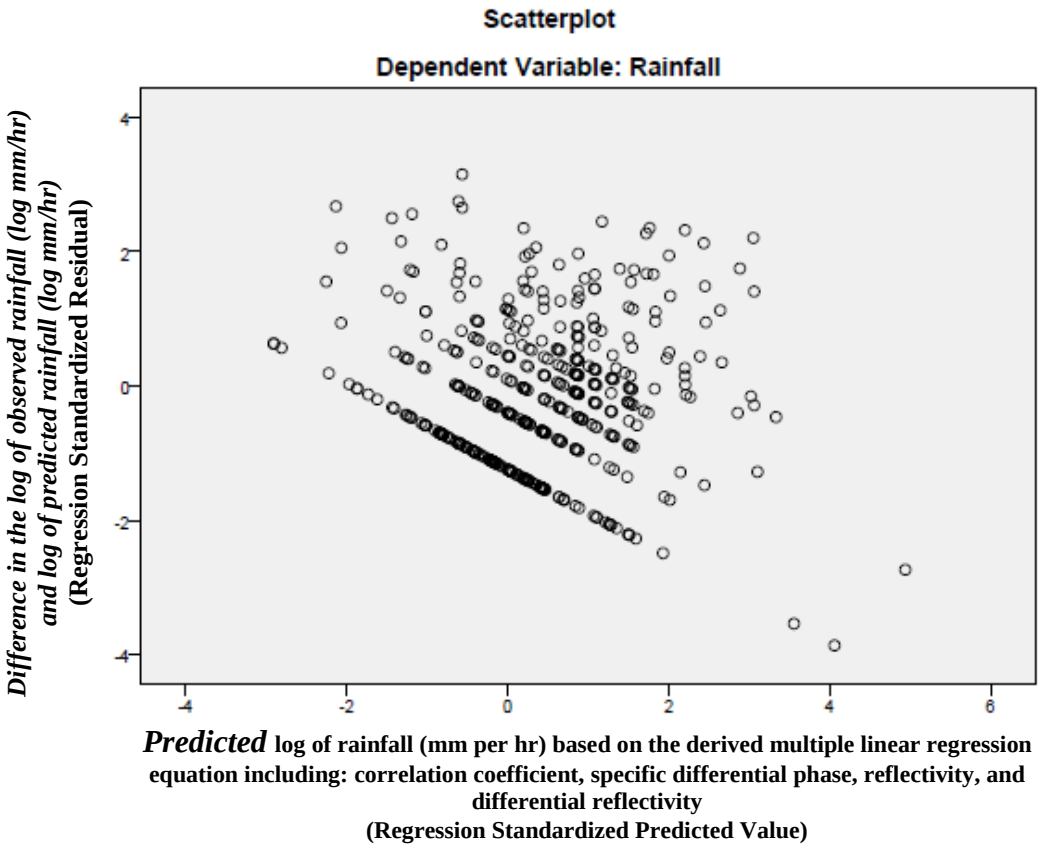


Figure 7. Scatterplot of the multiple regression case utilizing all four radar variables: correlation coefficient, specific differential phase, reflectivity, and differential reflectivity, for all storm events.

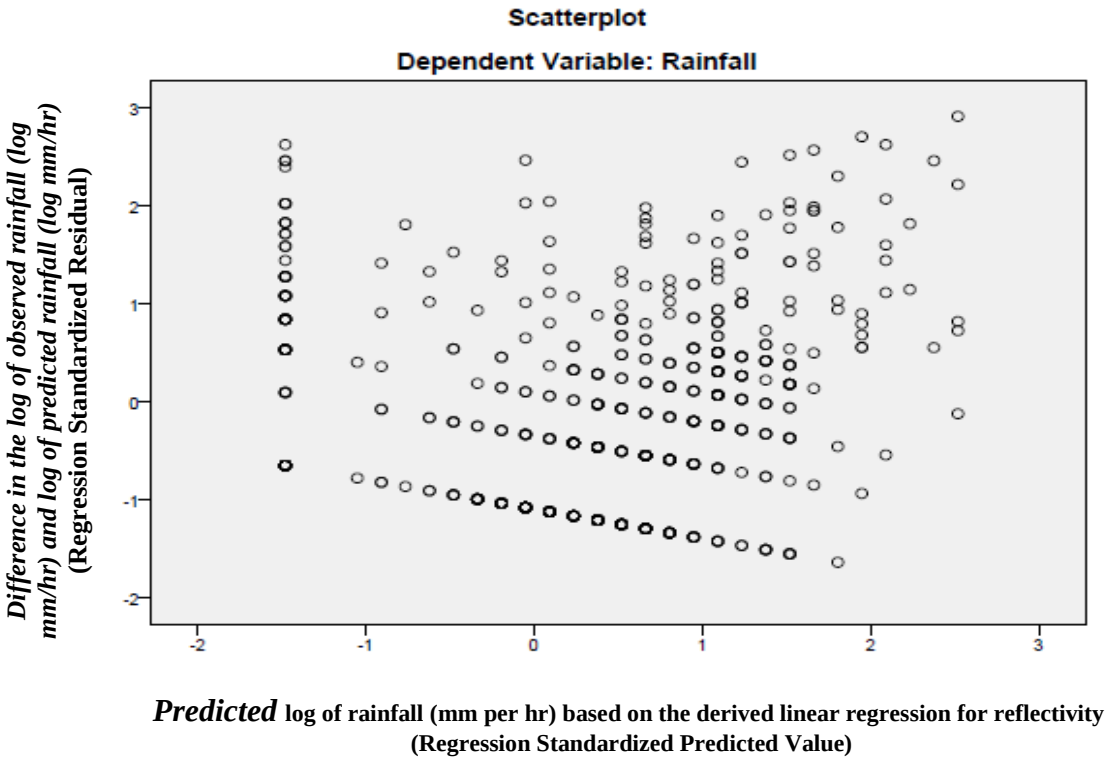


Figure 8. Scatterplot for the singular regression case of reflectivity for all storm cases.

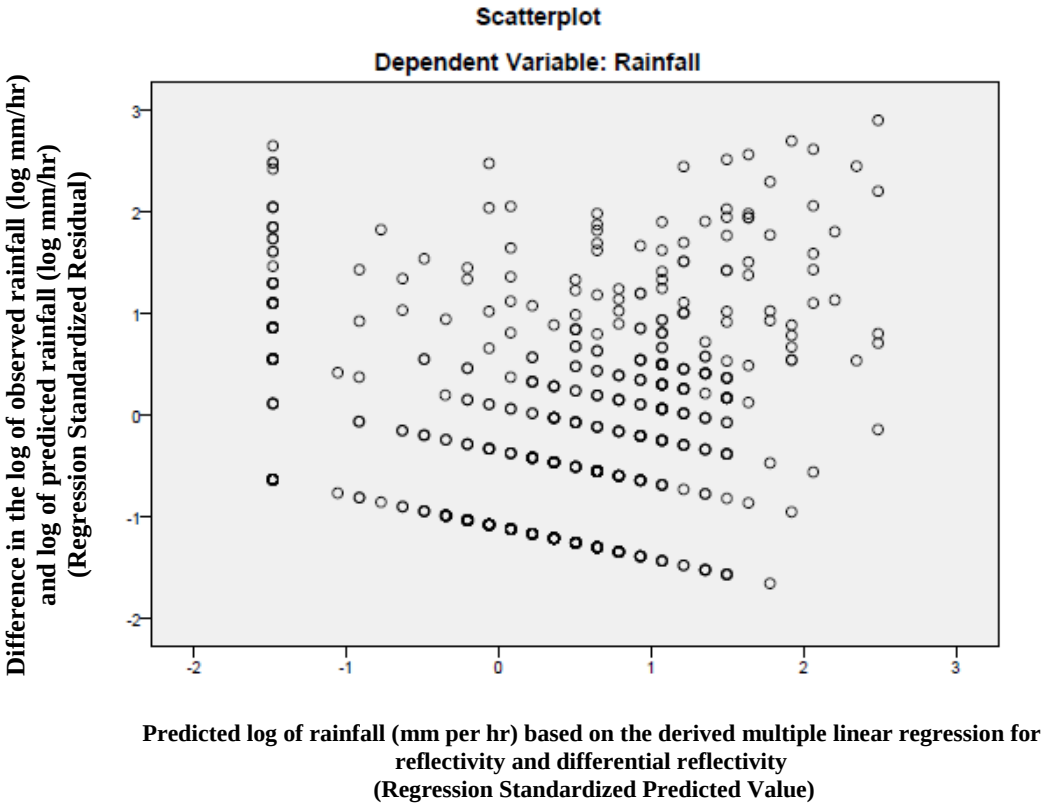


Figure 9. Scatterplot of the multiple regression case of the radar variables reflectivity and differential reflectivity for all storm cases.

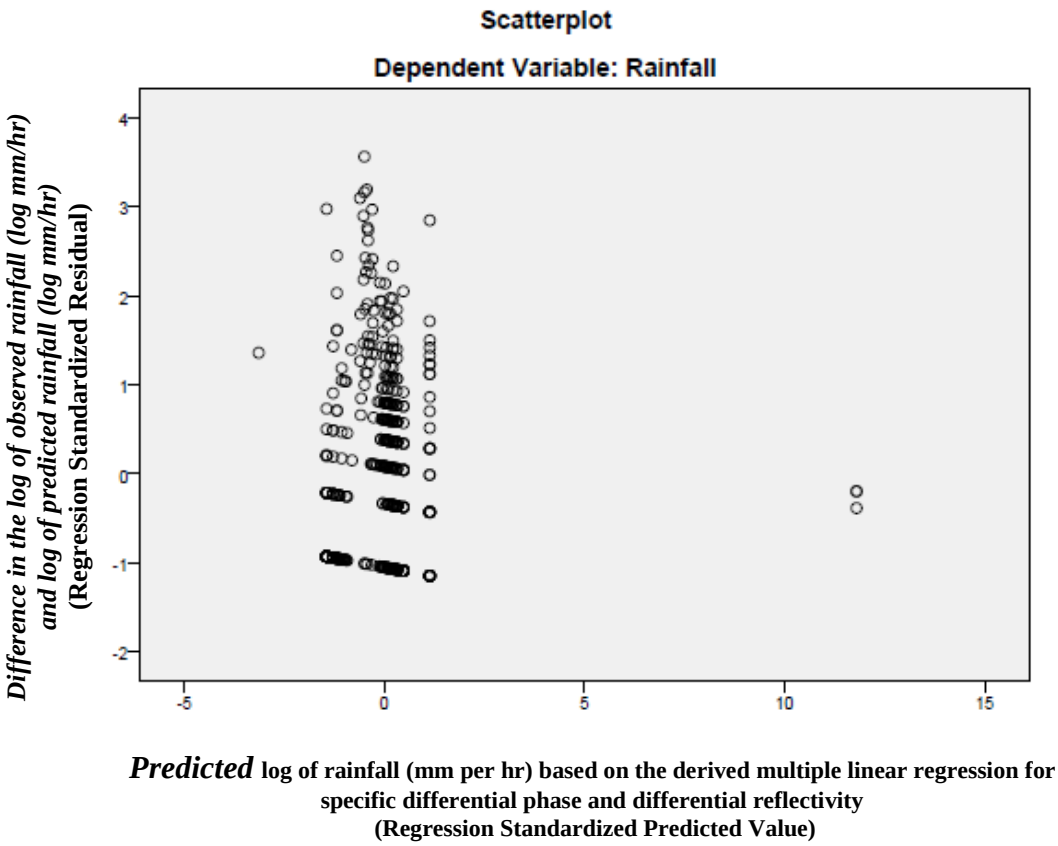


Figure 10. Scatterplot for the multiple regression case involving the radar variables specific differential phase and differential reflectivity for all storm cases.

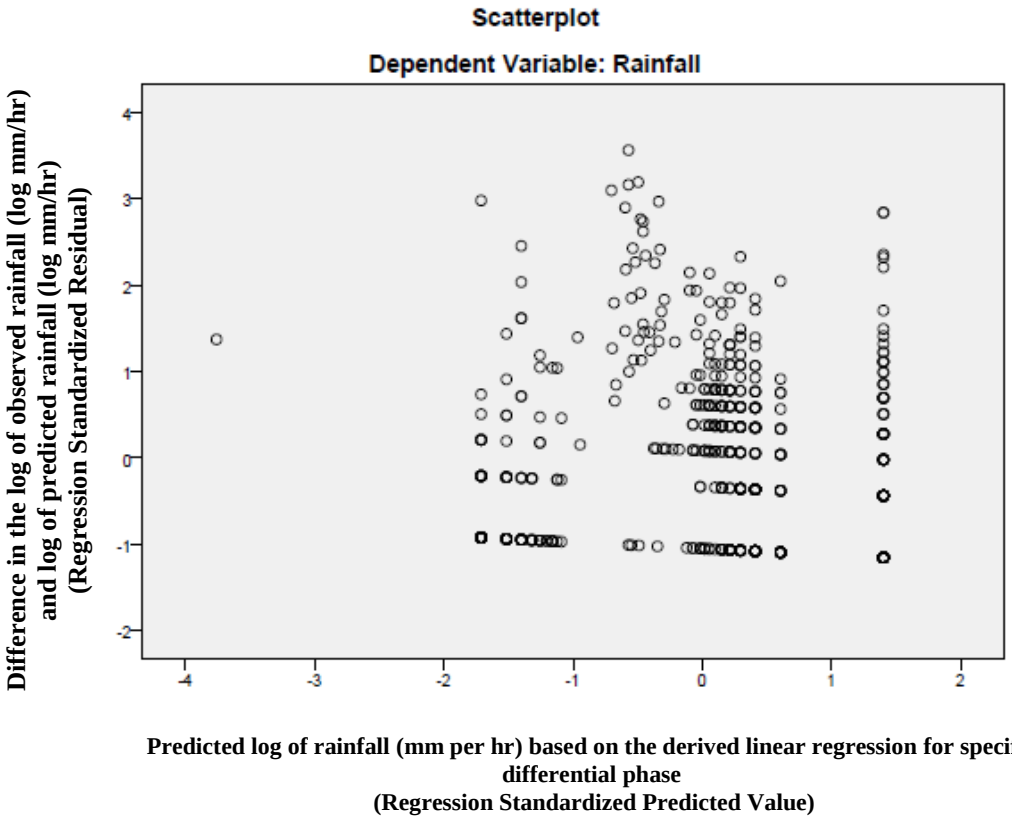


Figure 11. Scatterplot for the singular regression case involving the radar variable specific differential phase for all storm events.

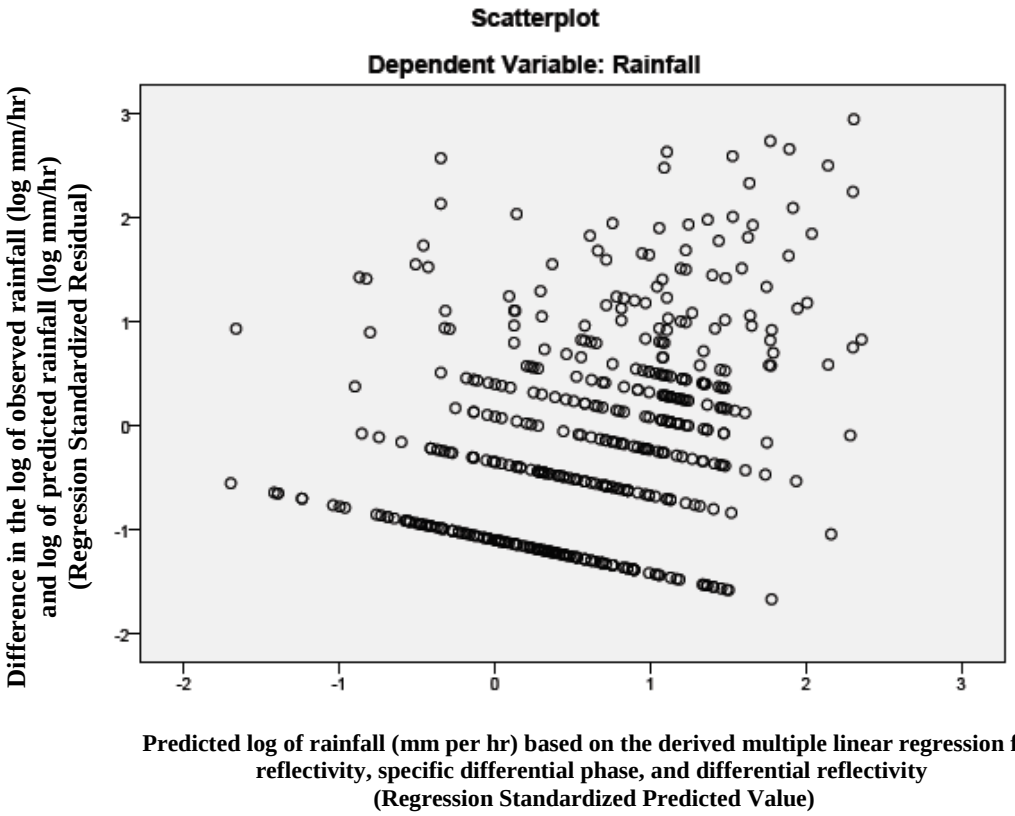


Figure 12. Scatterplot for the multiple regression case involving the radar variables reflectivity, specific differential phase, and differential reflectivity for all storm events.

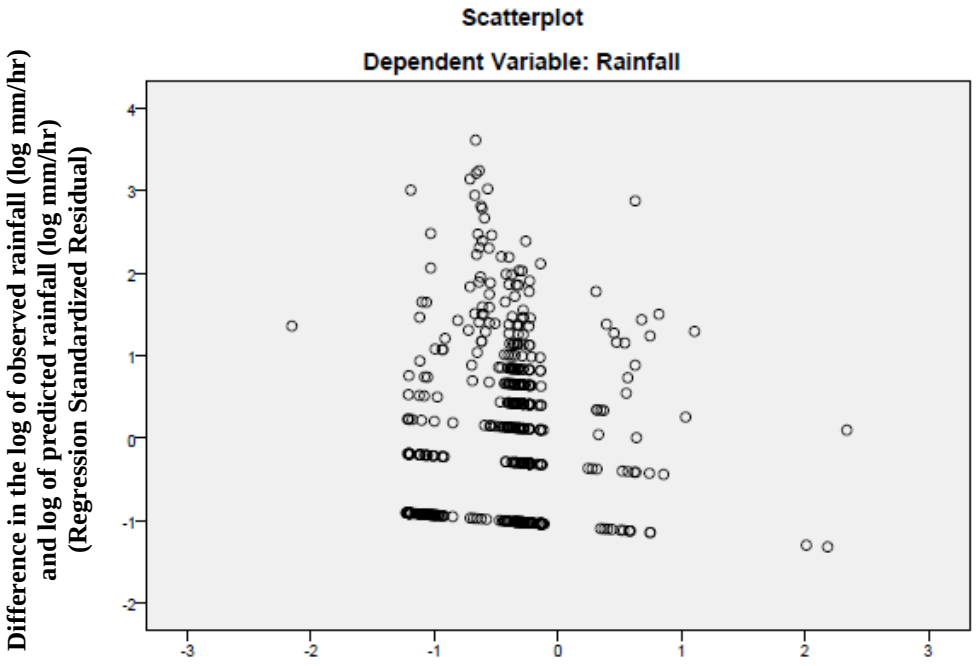


Figure 13. Scatterplot for the multiple regression case involving the radar variables correlation coefficient, specific differential phase, and differential reflectivity for all storm events.

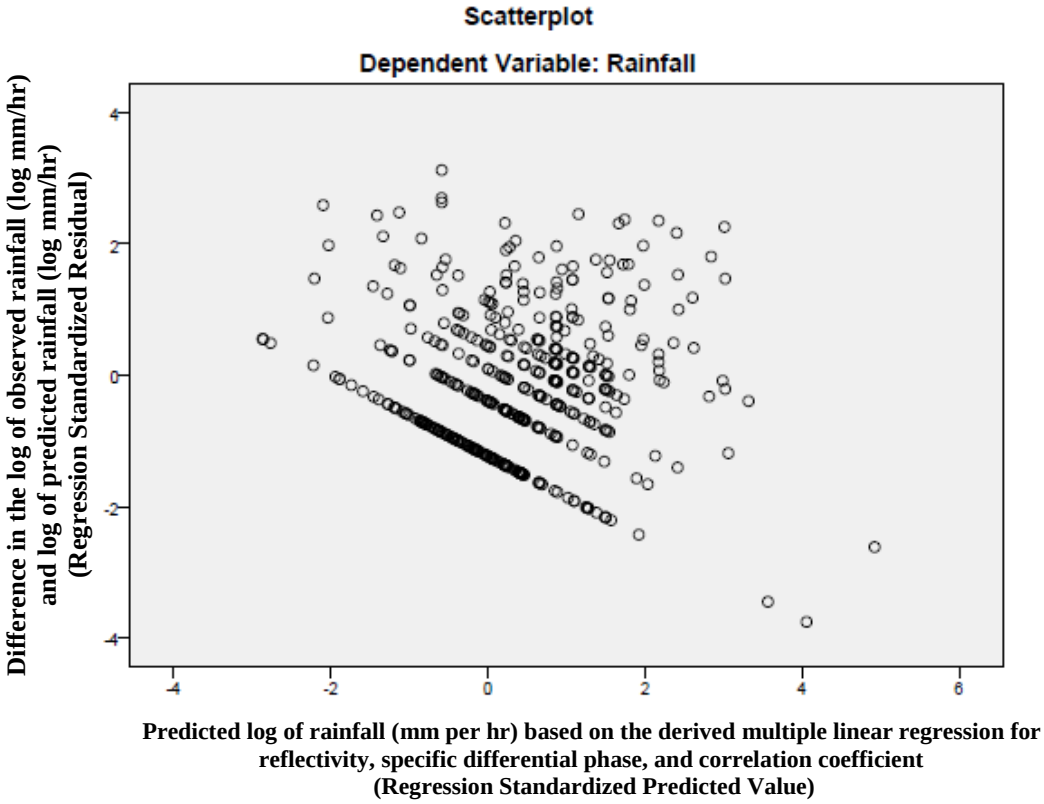


Figure 14. Scatterplot for the multiple regression case involving the radar variables reflectivity, specific differential phase, and correlation coefficient for all storm events.

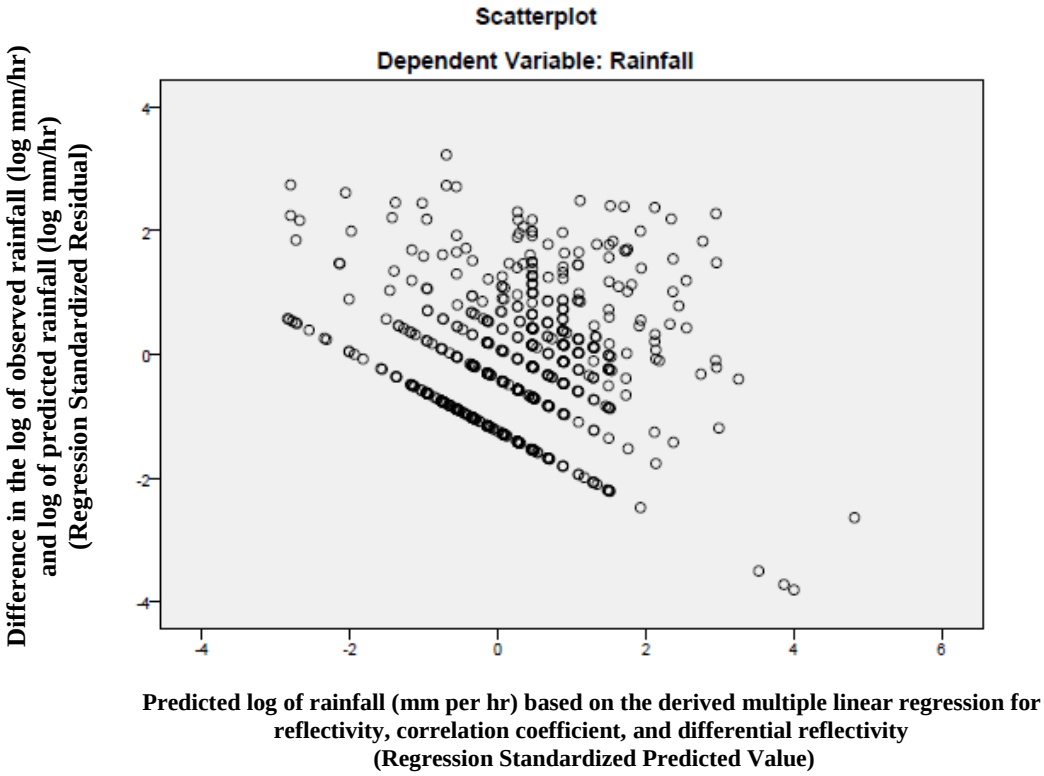


Figure 15. Scatterplot for the multiple regression case involving the radar variables reflectivity, correlation coefficient, and differential reflectivity for all storm events.

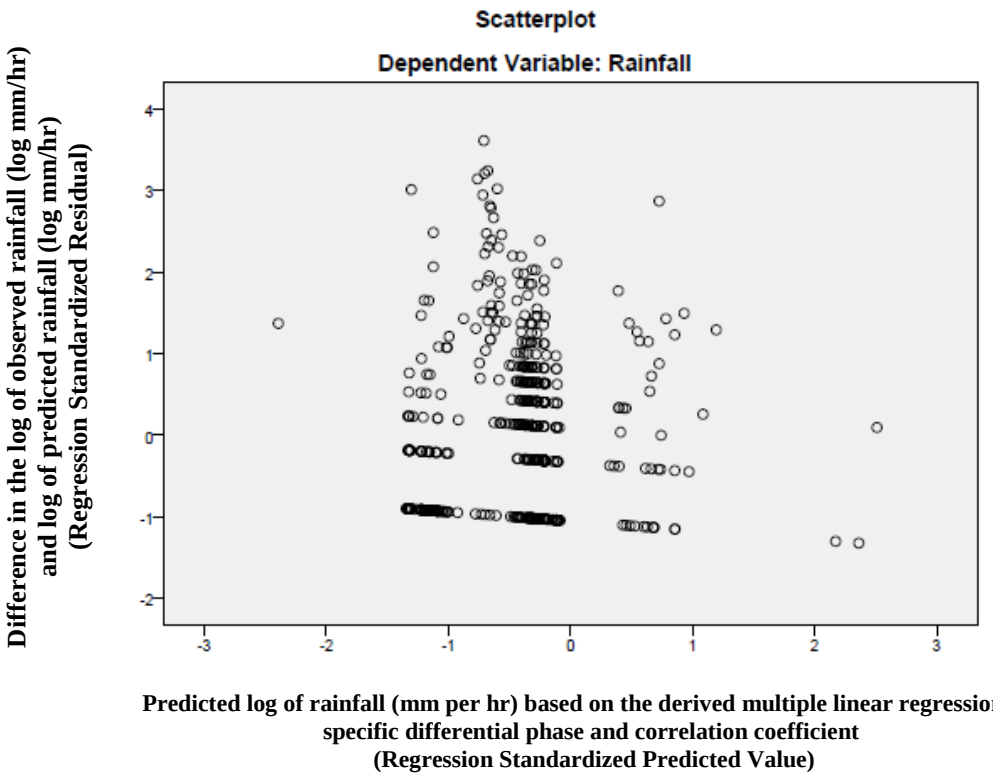


Figure 16. Scatterplot for the multiple regression case involving the radar variables specific differential phase and correlation coefficient for all storm events.

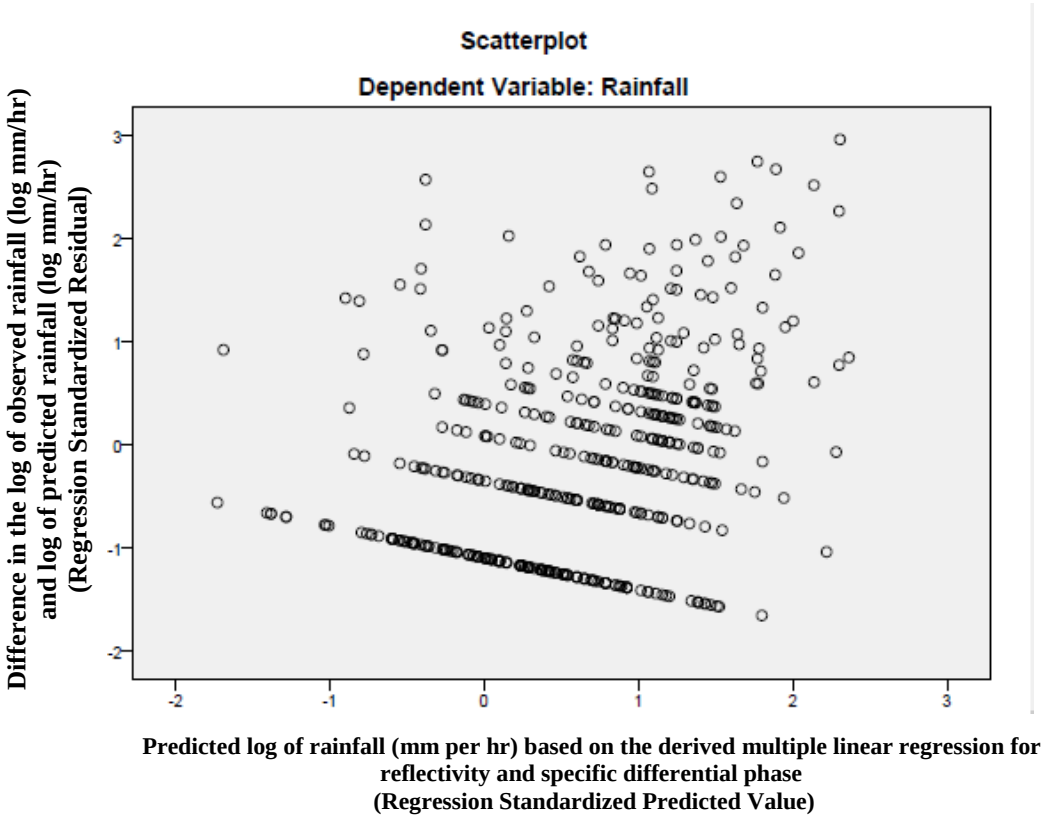


Figure 17. Scatterplot for the multiple regression case involving the radar variables reflectivity and specific differential phase for all storm events.

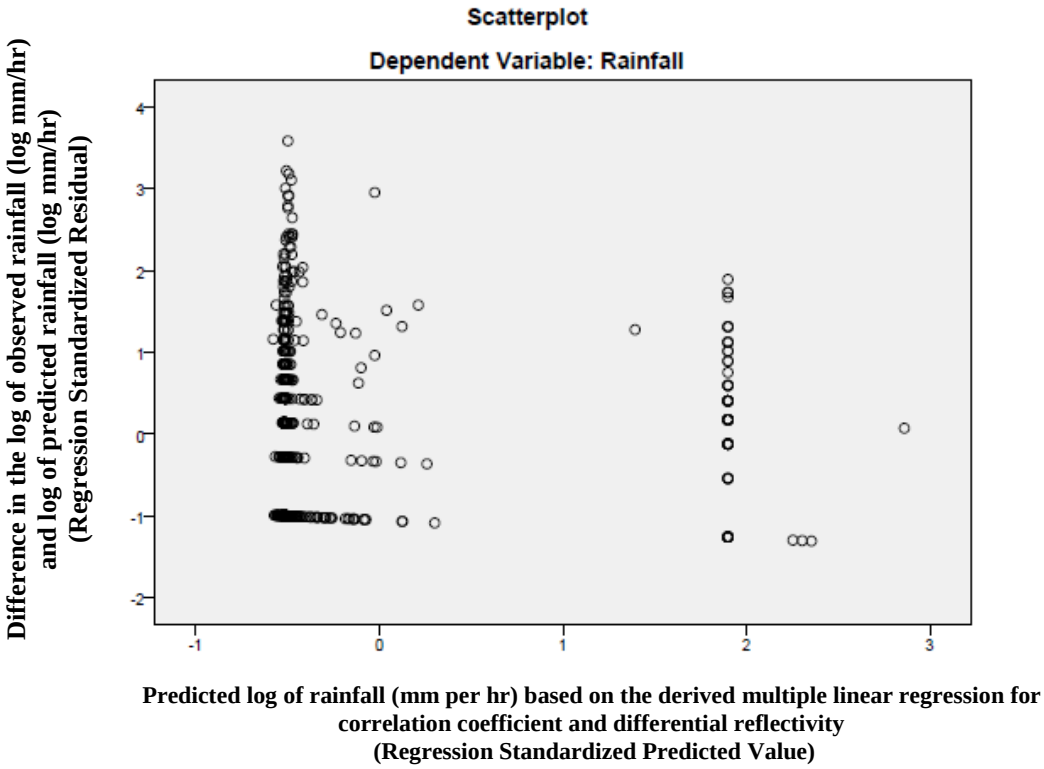


Figure 18. Scatterplot for the multiple regression case involving the radar variables correlation coefficient and differential reflectivity for all storm events.

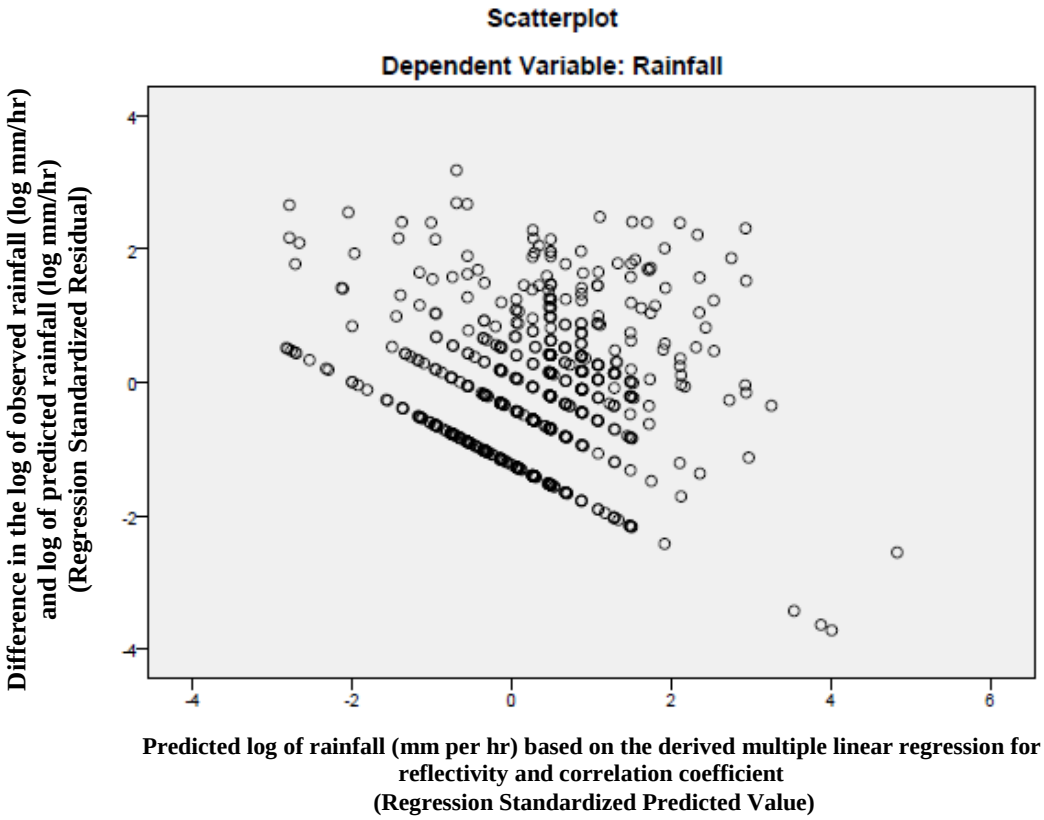


Figure 19. Scatterplot for the multiple regression case involving the radar variables reflectivity and correlation coefficient for all storm events.

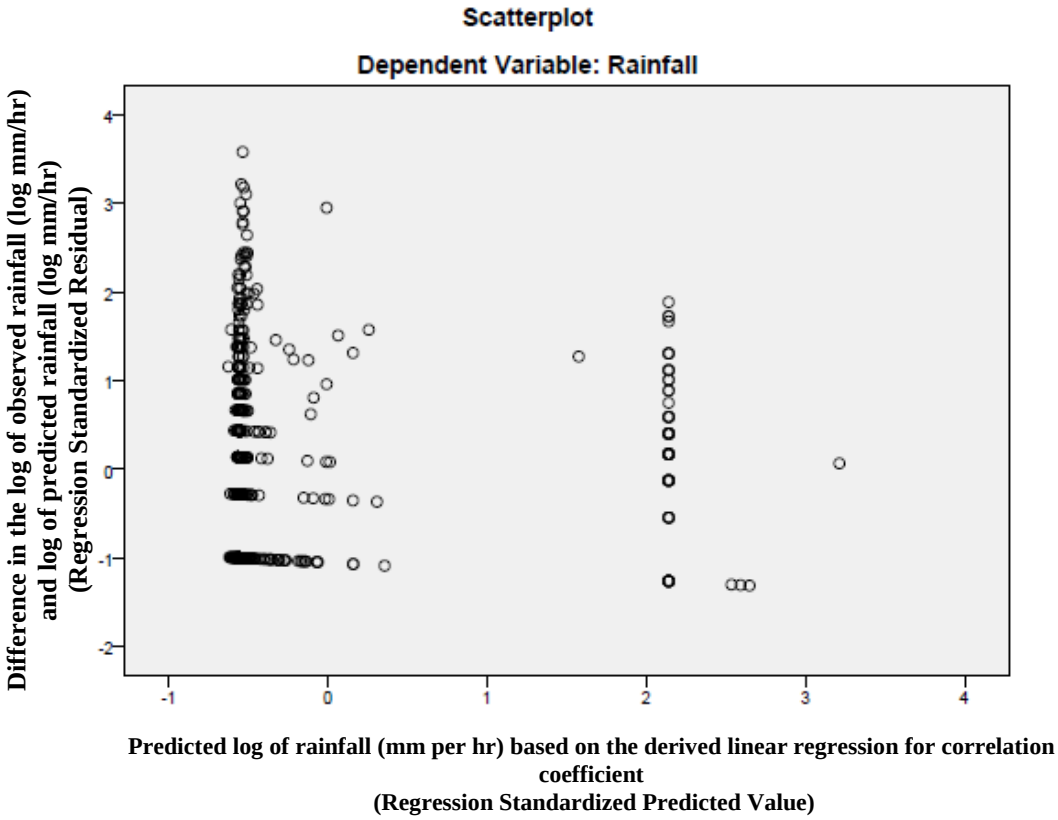


Figure 20. Scatterplot for the linear regression case involving the radar variable correlation coefficient for all storm events.

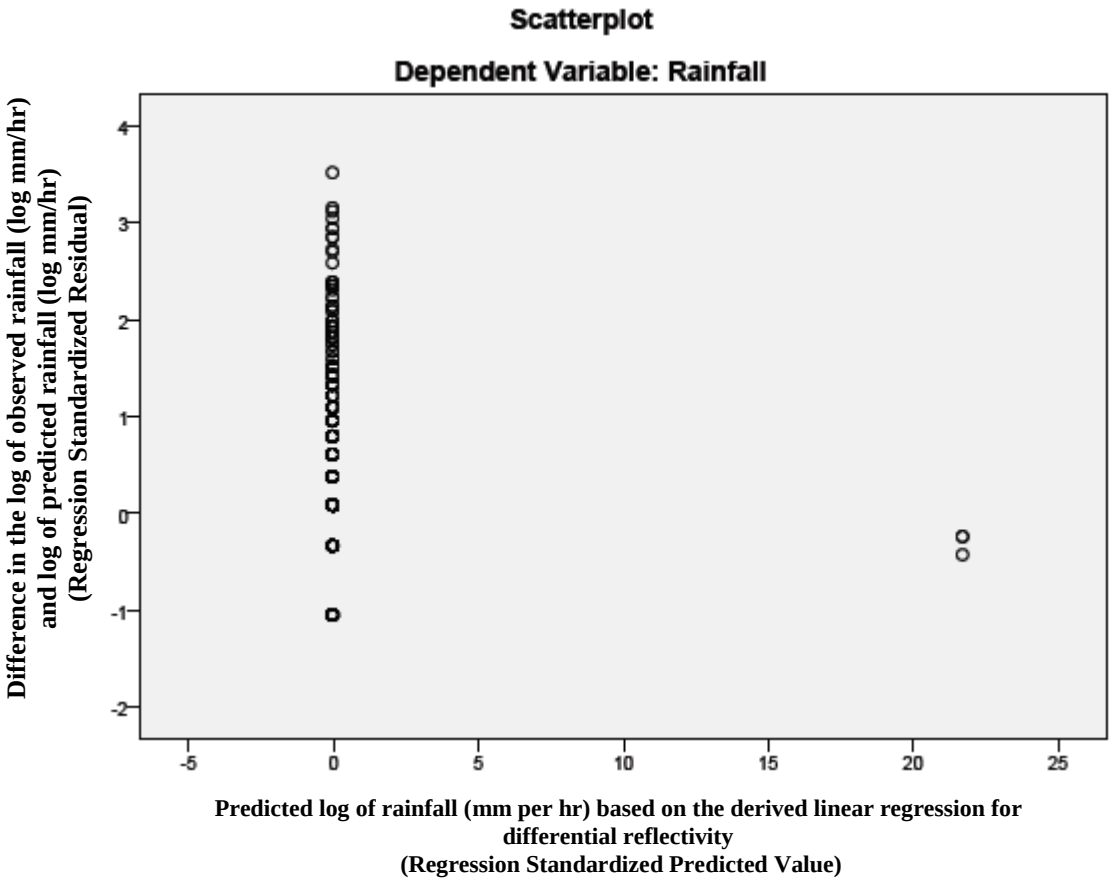
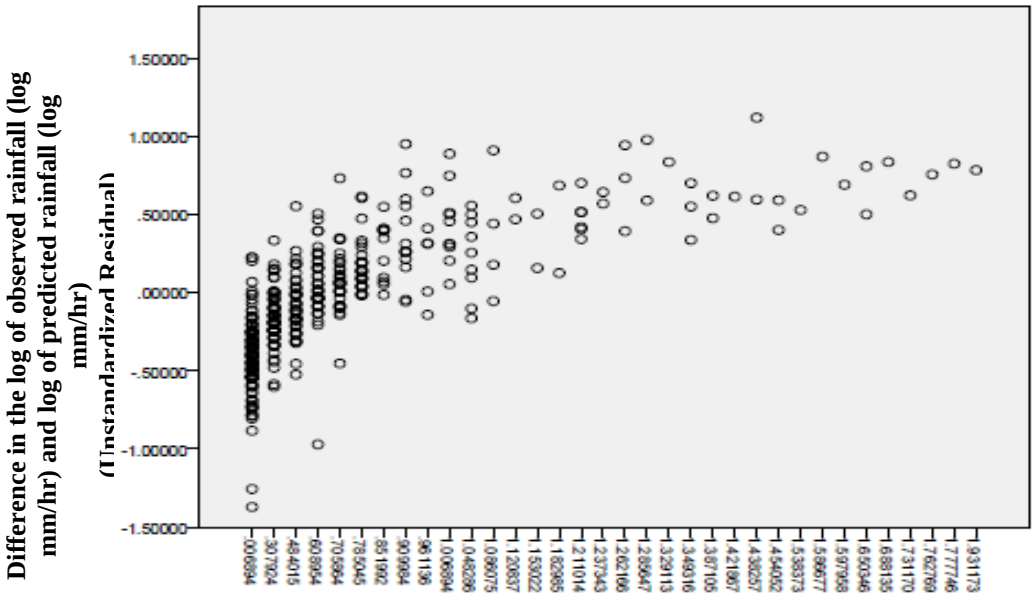
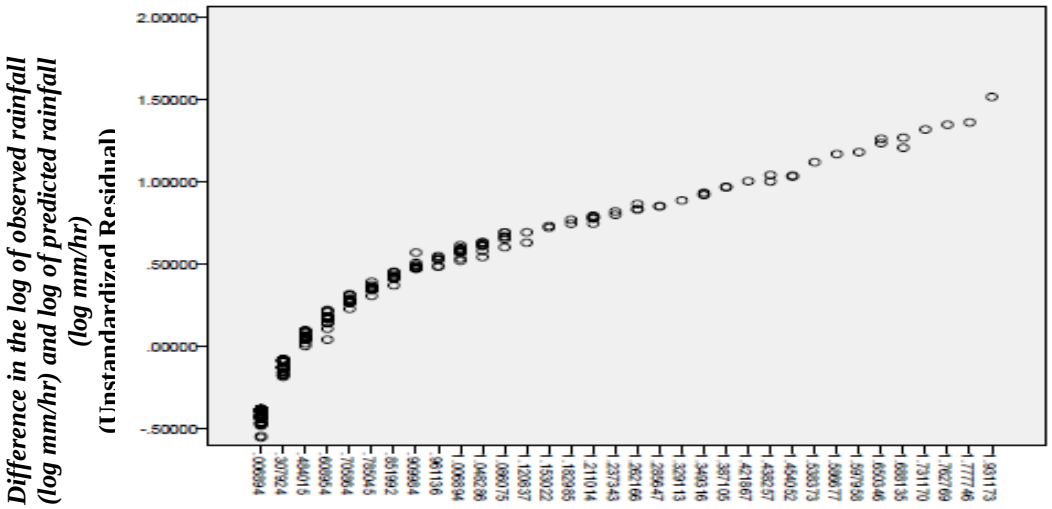


Figure 21. Scatterplot for the linear regression case involving the radar variable differential reflectivity for all storm events.



Observed log of rainfall (mm per hr) based on the derived linear regression for reflectivity, specific differential phase, correlation coefficient, and differential reflectivity

Figure 22. Scatterplot for log of the rainfall (mm per hour) vs the unstandardized residual of the multiple linear regression model of reflectivity, specific differential phase, correlation coefficient, and differential reflectivity.



Observed log of rainfall (mm per hr) **based on the derived linear regression for specific differential phase, correlation coefficient, and differential reflectivity**

Figure 23. Scatterplot for log of the rainfall (mm per hour) vs the unstandardized residual of the multiple linear regression model of specific differential phase, correlation coefficient, and differential reflectivity.

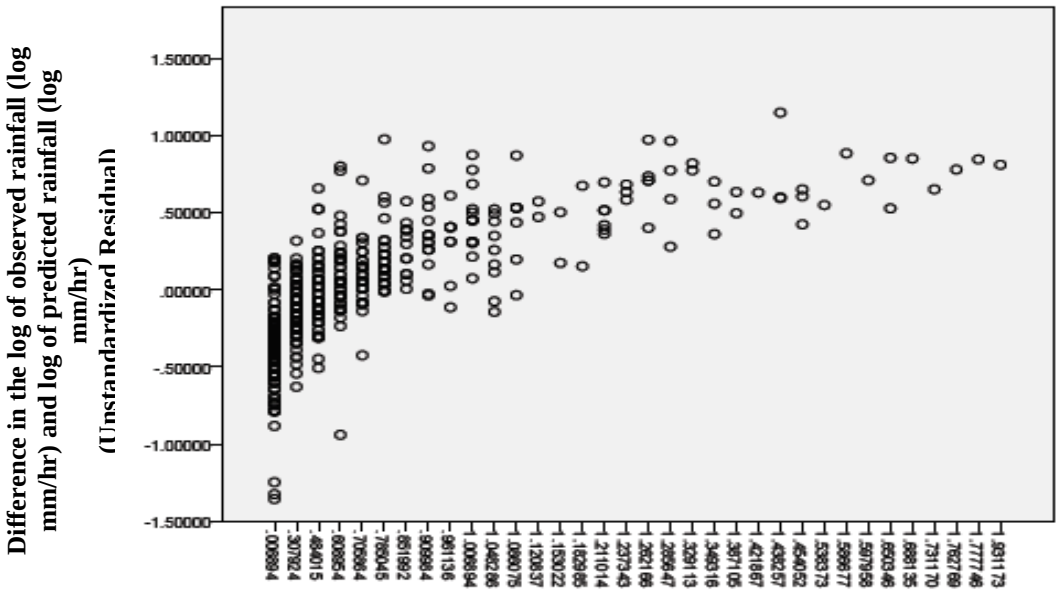


Figure 24. Scatterplot for log of the rainfall (mm per hour) vs the unstandardized residual of the multiple linear regression model of reflectivity, correlation coefficient, and differential reflectivity.

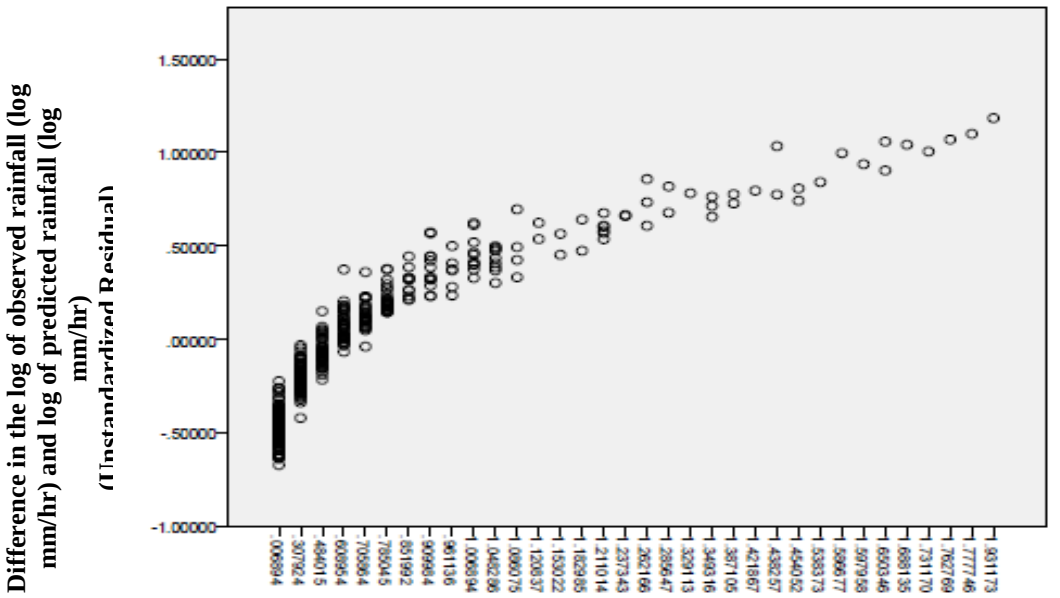


Figure 25. Scatterplot for log of the rainfall (mm per hour) vs the unstandardized residual of the multiple linear regression model of specific differential phase, reflectivity, and differential reflectivity.

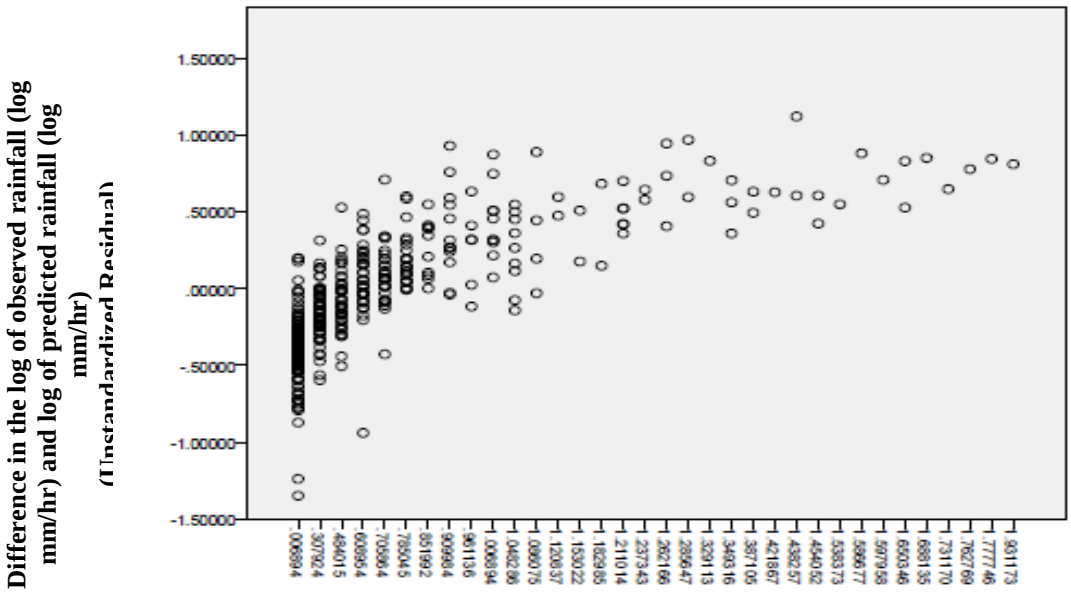


Figure 26. Scatterplot for log of the rainfall (mm per hour) vs the unstandardized residual of the multiple linear regression model of specific differential phase, reflectivity, and correlation coefficient.

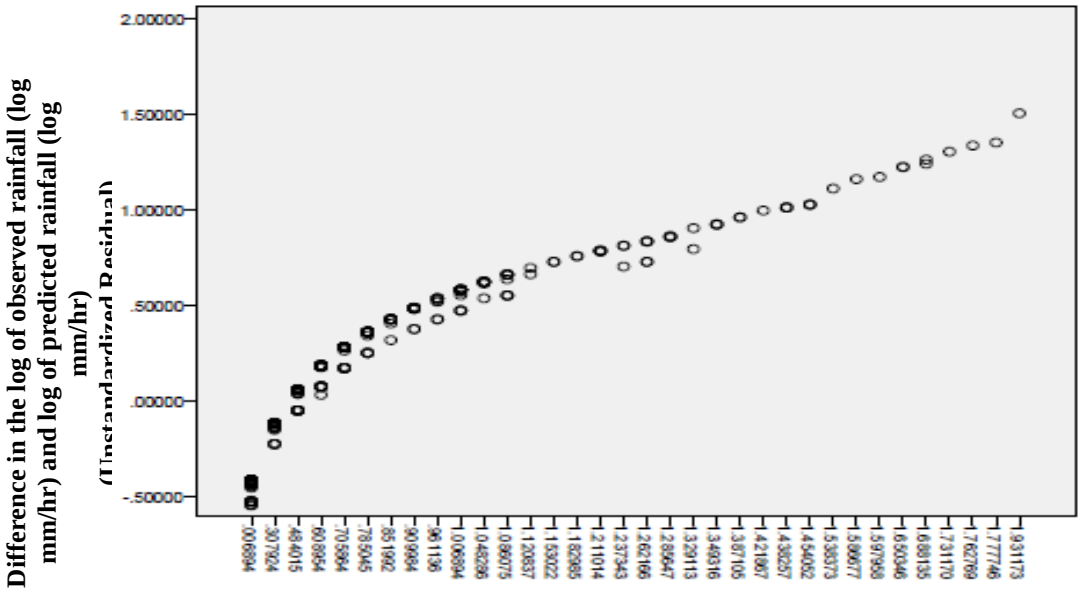


Figure 27. Scatterplot for log of the rainfall (mm per hour) vs the unstandardized residual of the multiple linear regression model of differential reflectivity and correlation coefficient.

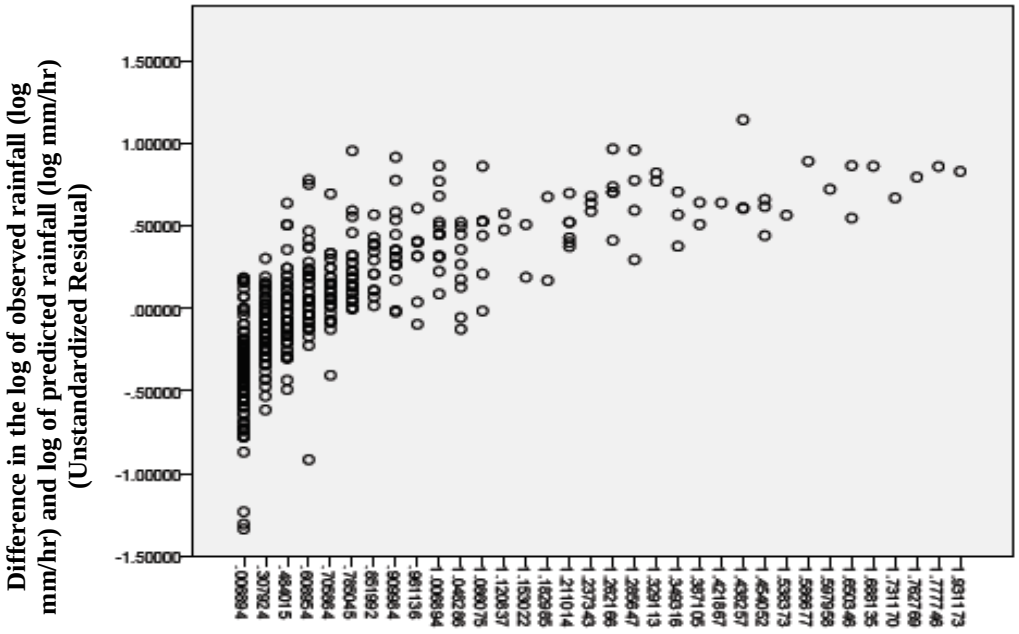


Figure 28. Scatterplot for log of the rainfall (mm per hour) vs the unstandardized residual of the multiple linear regression model of reflectivity and correlation coefficient.

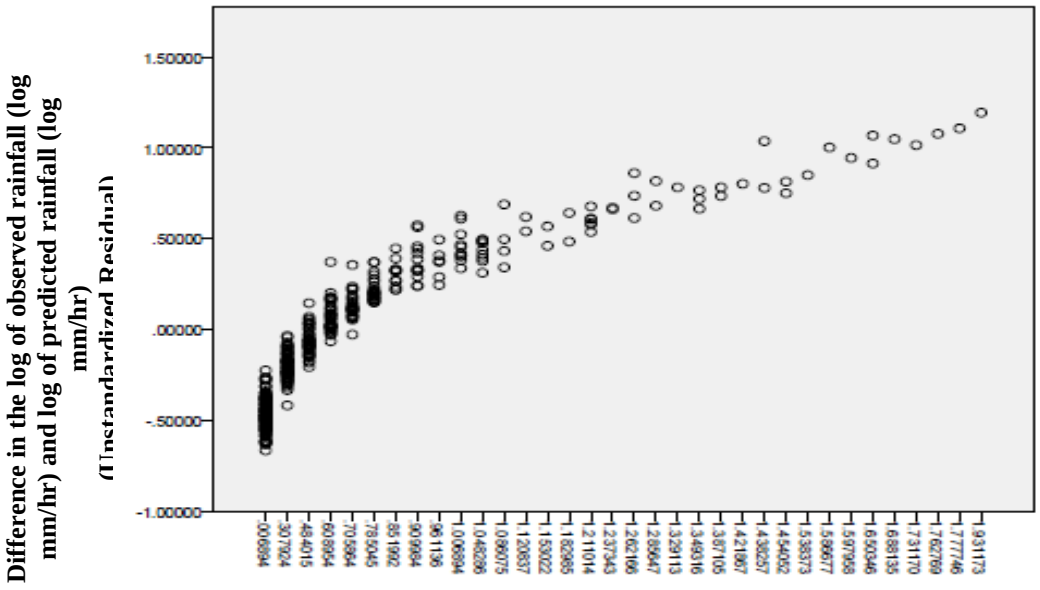


Figure 29. Scatterplot for log of the rainfall (mm per hour) vs the unstandardized residual of the multiple linear regression model of reflectivity and specific differential phase.

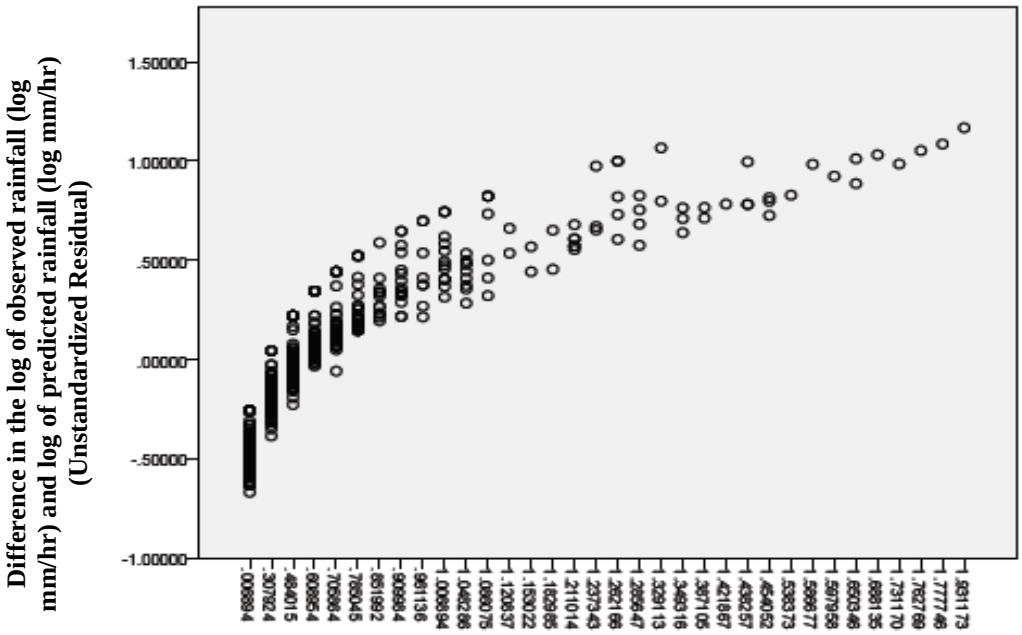


Figure 30. Scatterplot for log of the rainfall (mm per hour) vs the unstandardized residual of the multiple linear regression model of reflectivity and differential reflectivity.

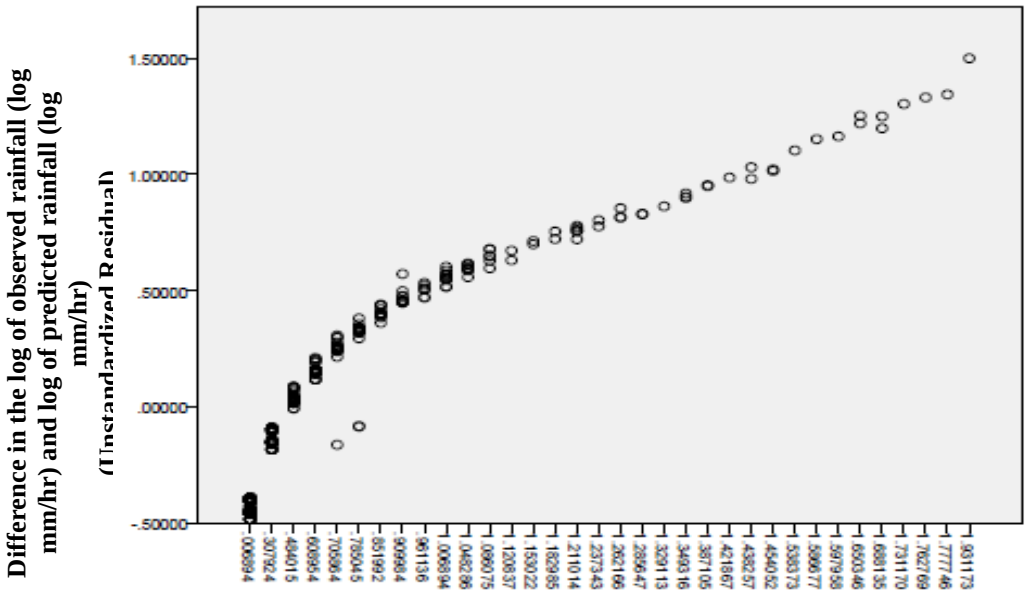


Figure 31. Scatterplot for log of the rainfall (mm per hour) vs the unstandardized residual of the multiple linear regression model of specific differential phase and differential reflectivity.

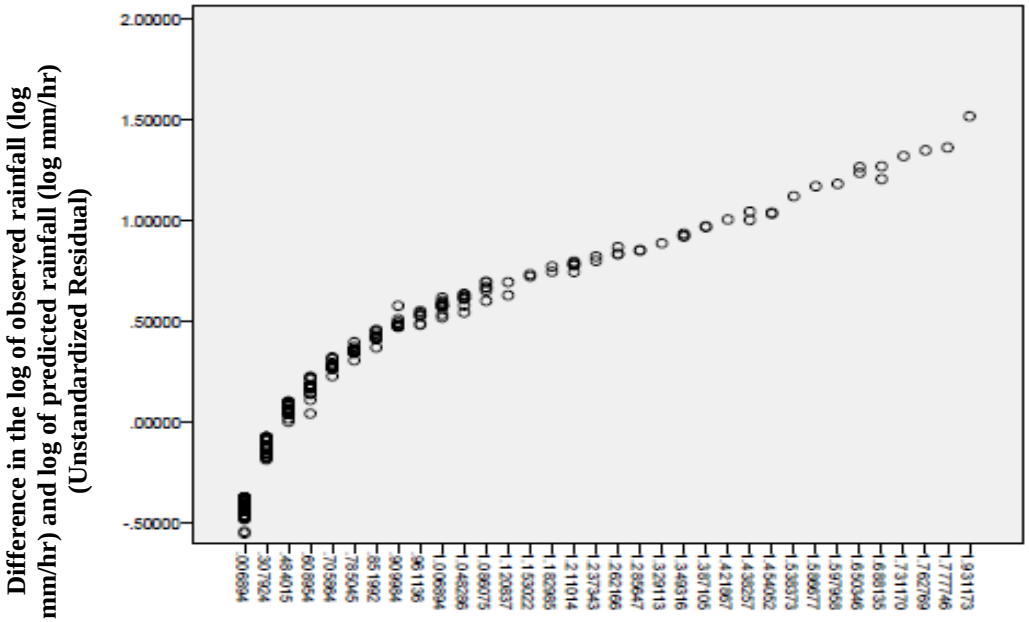


Figure 32. Scatterplot for log of the rainfall (mm per hour) vs the unstandardized residual of the multiple linear regression model of specific differential phase and correlation coefficient.

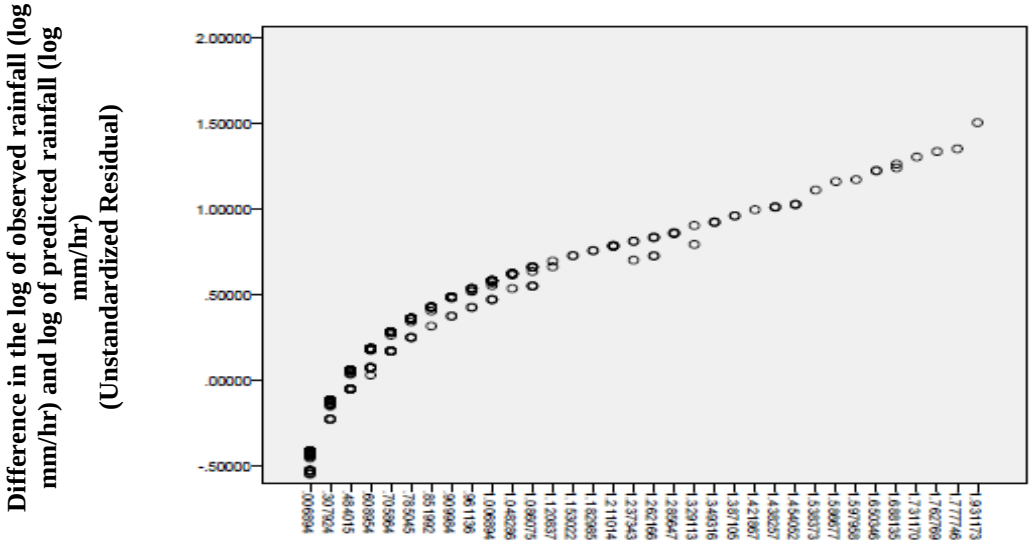


Figure 33. Scatterplot for log of the rainfall (mm per hour) vs the unstandardized residual of the linear regression model of correlation coefficient.

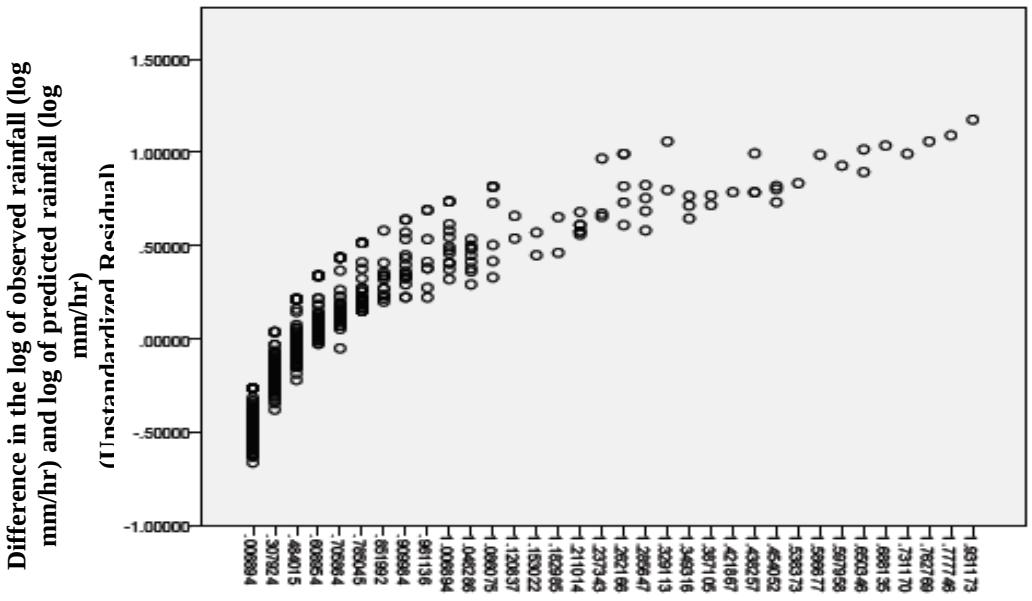


Figure 34. Scatterplot for log of the rainfall (mm per hour) vs the unstandardized residual of the linear regression model of reflectivity.

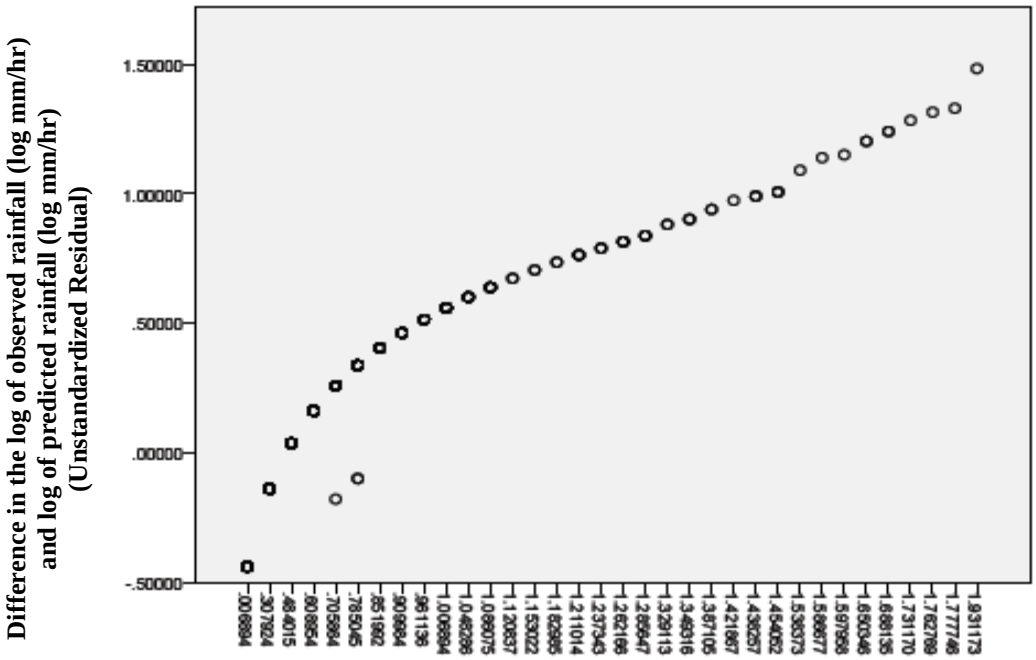


Figure 35. Scatterplot for log of the rainfall (mm per hour) vs the unstandardized residual of the linear regression model of differential reflectivity.

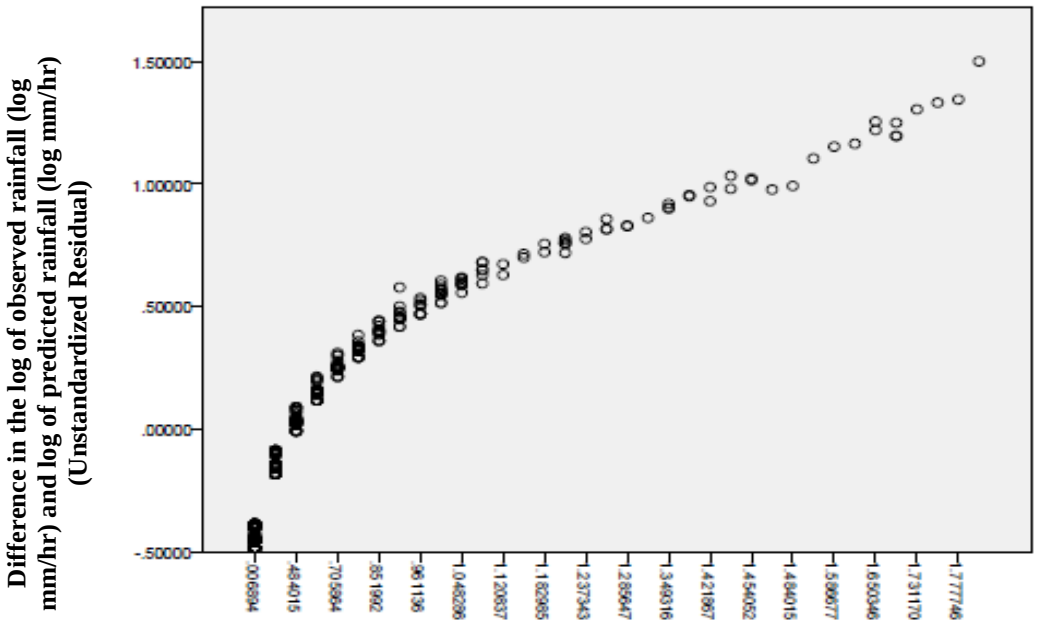


Figure 36. Scatterplot for log of the rainfall (mm per hour) vs the unstandardized residual of the linear regression model of specific differential phase.

5.2 Limitations of Results

In applying the R^2AIn-GIS tool, limitations arose. For instance there are several occurrences in the various storm events where there was radar data but no precipitation indicated with the rain gauges. It is possible that the rain could be falling around the gauge and just not in it to be measured. Wind and distance between radar scan height and the ground can contribute to differences in the gauge to radar measurement of precipitation as discussed in the next section.

Further limitations are associated with the seasonality of the storm events. Following Giangrande and Ryzhkov (2003), this study focused on the warm season from May through August during the first two seasons sampled with the polarimetric KMVX radar. To develop an adequate MLR that is representative of the region, a full-season

MLR should be further developed or at the very least a cold-season MLR should also be developed. By computing various statistical tests for the gauges operated by the USGS for the Glacial Ridge Prairie, limitations arise in having station E05 missing from the 2012 storm cases. This gauge is crucial in providing a thorough depiction of rainfall within the Glacial Ridge Prairie.

Following Giangrande and Ryzhkov (2003), the number of observation hours utilized in this study was 114 hours compared to their 52 hours. This increase in observation hours does provide a larger data collection for analysis. However, a limitation arises in the number of gauges for this study area, with seven gauges versus the 108 gauges of the Oklahoma Mesonet. It should be noted that the region of the Glacial Ridge Prairie is significantly smaller than the domain covered under the Oklahoma Mesonet. The average spacing between each Oklahoma Mesonet gauge is 35 km while this study site has an average gauge spacing of 9.6 km. Another limitation is the raster calculator function of the R^2AIn-GIS. Again, because the radar does not always have three consecutive radar scans ending at 00, 15, 30, and 45, the scans are not always an identical representation of the fifteen-minute gauge data. The average of the sum of three consecutive radar scans that end near or at the rain gauge intervals listed above was used to calculate a value for each radar variable. Occasional radar grids missing from various storm events occurred during analysis.

5.3 Limitations and Errors from Data Sources

It is evident that limitations exist when analyzing storm events on a spatial and temporal scale. Tipping bucket rain gauges serve as a great point indicator of rainfall on

a temporal scale. Although these gauges are used in many locations across the United States, flaws exist. Commons errors that can cause rainfall estimates to be incorrect are: turbulent airflow around the gage, melting hail, evaporation, adhesion of drops to gauge, and water splashing (Speltz 1992). With each bucket tip containing 0.01 inches of water in the bucket, the effects of evaporation, drop adhesion and splashing can cause underestimates in rainfall totals. Dahlstrom (1973) found that less than 2% of underestimated rainfall totals were from the contributions of evaporation, drop adhesion and splashing. Very hard rainfall in the presence of wind can cause rain droplets to not fall within the small orifice of the gauge, leading to underestimates of rainfall. Average deficits at 5 m s^{-1} and 10 m s^{-1} are 12% and 19% respectively (Larson and Peck 1974).

Meteorologists have utilized windshields around gauges to reduce the effects of wind as have also strategically placed gauges in open areas with no structural obstructions to eliminate shadowing effects. Another issue with the tipping bucket style of rain gauge is with the bucket itself. If there is insufficient rain to tip the bucket, the gauge will not report precipitation until the bucket tips (at which point it sending an electronic signal to the receiver). This delay affects the reporting time of the precipitation. An error that can cause overestimates in tipping bucket rain gauges are faulty reed switches. This occurs when the full bucket tips, releasing the collected water and returning back to the collection position quickly with very little to no time for the other bucket to collect rain (Clement 1995). This problem can result in the collecting bucket to inefficiently remove water, meaning when it tips again the precipitation amount will overestimate by counting some of the precipitation that fell more than once.

During heavy downpour events, funnel misdirection can occur. Clement (1995) defines this as "when a sudden onset of precipitation partially fills the collector and a vortex action is created within the funnel." The vortex action forces some or all of the water away from the collecting bucket, leading to underestimates in rain total.

With the rain gauges in the study area located out in fields, the potential for small animals and insects to make homes in rain gauges exist. A remedy for this pest issue is regular maintenance and inspection of the gauges.

A factor that could hinder further precipitation analysis over the region is the period of the rain gauge observation record. With the rain gauges recording since 2003 or 2004, the capability of doing point to area frequency of precipitation is reduced with no coverage prior to installation. With continued maintenance and operation, the gauges in the GRNWR will provide further precipitation data to contribute to rainfall analysis and further hydrologic modeling.

Radars also have significant limitations. No drop size distribution data were available for estimation of a refined *Z-R* herein. The distance between KMVX and GRNWR gauge network is ~93 km, meaning that the 0.5° is at an altitude of ~700 m AGL. Thus, evaporation is possible, which in the extreme could leading to non-missing radar variables but no ground precipitation. Winds aloft—similar to winds at the surface—near the gauge can lead to underestimates of accumulated precipitation. Strong winds aloft can steer the precipitation so that when a radar scans the rain aloft it is over a grid cell but could it reaches the surface outside the grid cell. Conversely, the radar can scan rain aloft and away from the gauge but strong winds aloft can steer the rain into the gauge with clear radar scans above the gauge.

CHAPTER VI

CONCLUSION

The R^2AIn-GIS is an additional data processing and analysis tool that uses GIS software to effectively combine radar and rain gauge information in a convenient manner. By taking parts of the concepts of Zhang and Srinivasan (2010), R^2AIn-GIS builds upon their original construct with the incorporation of dual-polarization radar capabilities as well as updated scripting languages for current use in the ArcGIS® software suite. The R^2AIn-GIS tool was used to analyze warm season (May through August) storm events over the GRNWR in an effort to determine the validity of the tool as an effective means to analyze storms in a GIS environment. The tool was used on a variety of storm events that occurred in the warm season of 2012 and 2013 to utilize data from the recently upgraded dual-pol KMVX radar. The average fifteen minute radar grid was paired with the fifteen minute rainfall to allow further statistical analysis. Including statistical frequency, gauges and radar values alike, data were inputted into SPSS, a statistical analysis software, to determine a multiple linear regression relationship between the variables of rainfall, reflectivity, correlation coefficient, differential reflectivity, and specific differential phase. The results of the statistical analysis show that based on the number of storm cases and resultant fifteen minute radar, gauge pairings that there are models that can be used that predict better than the standard Z-R relationship.

Based off the findings of the various statistical tests, it appeared that the overall best models included the radar variables of reflectivity and correlation coefficient as seen in Table 10. These two variables had significant p-values and their Durbin-Watson scores were among the highest even compared with the other radar variables. Models including the radar variables reflectivity and correlation coefficient were found to be heteroscedastic along with the highest R Squared values. Future work on the implementation of a reflectivity and correlation coefficient equation that is truly representative for more than just the area in GRNWR would greatly benefit the hydrometeorological and radar meteorology literature.

It is suggested that future work focus on additional storm analysis and operation of the gauges within the GRNWR as the region undergoes a transformation from farmland back to wetland. More storm events and a growing period of record for gauges will enable continuous updating of the MLR and statistical values of the regression models. It would be further beneficial to include the Goldfeld-Quandt test to also address the assumption of homoscedasticity and apply this to the previous cases with more advanced statistical software.

Future research should focus on utilizing precipitation data that is more robust and covers a larger geographic landscape. The data for this research study had very little precipitation to offer the best results. Thus, it is critical that any future research conducted includes larger amounts of precipitation. Perhaps, since the study area was so refined the results were not as ideal as in other instances where researchers would assess the precipitation over a larger geographic landscape. It is also important that future researchers provide specific details on methodologies and results. In so many instances,

such as Giangrande and Ryzhkov (2003), the authors fail to even report the important components of their methodologies and results such as important issues regarding ANOVA. This research study at least attempted to address each step of the methodologies and results, allowing future researchers to mimic and improve upon the study.

REFERENCES

Arnold, J. G., R. Srinivasan, R. S. Muttiah, J. R. Williams, 1998: Large area hydrologic Modeling and assessment part 1:model development. *Journal of the American Water Resources Association,* **34**, 73-89.

Bannayan, M., G. Hoogenboom, 2008: Weather analogue: a tool for real-time prediction of daily weather data realizations based on a modified k-nearest neighbor approach. *Environmental Modelling and Software*, **23**, 703-713.

Battan, L. J., 1973: *Radar observation and the atmosphere*. University of Chicago Press, Chicago, IL.

Box, G. E. P. and Cox, D. R., 1964: An analysis of transformations, *J. Roy. Stat. Soc., Series B,* **26**, 211-252.

Brandes, E. A., 2000: Dual-polarization radar fundamentals and algorithm prospects. 53 pp. (Available online at http://www.roc.noaa.gov/app/sta/algorithm00.pdf.)

Clement, P. C., 1995: A comparison of radar-derived precipitation and rain gage precipitation in northeastern Colorado., M. S. Thesis, Dept. of Atmospheric Science, Colorado State University, 92 pp (Available from http://discovery.library.colostate.edu/Record/.b19433724)

Cohen, J., 1988: *Statistical power analysis for the behavioral sciences*. Erlbaum, 2nd Edition.

Coulter, J., 1910: *Industrial history of the Valley of the Red River of the North.* Madison, WI. State Historical Society of North Dakota. 146 pp.

Dahlstrom, B., 1973: Investigation of errors in rainfall observations: a continued study. Report 34, Dept. of Meteorology, University of Uppsala. 563-575 pp. (Available fromhttp://journals.ametsoc.org/doi/pdf/10.1175/2007JHM925.1)

Doviak, R. J. and D. S. Zrnić, 1993: *Doppler radar and weather observations*. Academic Academic Press, 562 pp.

Environmental Systems Research Institute, 2015: ArcGIS® help. http://resources.ArcGIS®.com/en/help/ (last accessed 20 October 2015)

Field, A., 2013: *Discovery statistics using IBM SPSS statistics*. Sage Pub., 915 pp.

Fodor, N., and G. J. Kovacs, 2005: Sensitivity of crop models to the inaccuracy of meteorological observations. *Physics and Chemistry of the Earth*, **30**, 53-57.

Gebremichael, M., and W. F. Krajewski, 2004: Assessment of the statistical characterization of small-scale rainfall variability from radar: Analysis of TRMM ground validation datasets. *J. Appl. Meteor.*, **43**:8, 1180-1199.

Giangrande, S.E., and A.V. Ryzhkov, 2003: The quality of rainfall estimation with the polarimetric WSR-88D radar as a function of range. Preprints, *31st Int. Conf. on Radar Meteorology*, Seattle, WA, Amer. Meteor. Soc., 357-360.

Goldfeld, S. M., and R. E. Quandt, 1965: Some tests for homoscedasticity. *J. Amer. Stat. Assoc*. **60**, 539–547.

Goovaerts, P., 2000: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrology*, **228**, 113-129.

Hancock, P.A., and M. F. Hutchinson, 2006: Spatial interpolation of large climate data sets using bivariate thin plate smoothing splines. *Environmental Modelling and Software,* **21**, 1684-1694.

Hardegree, S. P., S. S. Van Vactor, D. H. Levinson, and A. H. Winstral, 2008: Evaluation of NEXRAD radar precipitation products for natural resource applications. *Rangeland Ecology & Management*, **61**, 346-353.

James, L., and S. Korom, 2001: Lessons from Grand Forks: planning nonstructural flood control measure. *Natural Hazards Review.* **2**(4), 182-192.

Jeffrey, S.J., J. O. Carter, K. B. Moodie, and A. R. Beswick, 2001: Using spatial interpolation to construct a comprehensive archive of Australian climate data. *Environmental Modelling and Software,* **16**, 309-330.

Jolly, W.M., J. M. Graham, A. Michaelis, R. Nemani, and S. W. Running, 2005: A flexible integrated system for generating meteorological surfaces derived from point sources across multiple geographic scales. *Environmental Modelling and Software,* **20**, 873-882.

Larson, L.W., and E. L. Peck, 1974: Accuracy of precipitation measurements and hydrologic modeling. *Water Resource Res.*, **10**, 857-863.

Lindhult, M.S., P. J. Godfrey, and J. G. Fabos, 1988: Geographic information systems in water resources research. Water Resources Research Center, University of Massachusetts, Amherst, Massachusetts, 85 pp.

Marshall, J.S., and W. M. K. Palmer, 1948: The distribution of raindrops with size. *J. Meteorology*, **5**, 165-166.

National Centers for Environmental Information, 2014: NEXRAD Data Archive for Grand Forks KMVX. National Oceanic and Atmospheric Administration. [Available online from https://www.ncdc.noaa.gov/nexradinv/]

Nature Conservancy, The, 2015: Minnesota Glacial Ridge Project. http://www.nature.org/ourinitiatives/regions/northamerica/unitedstates/minnesota/placesweprotect/glacial-ridge-project.xml (last accessed 21 October 2015).

Neff, E. L., 1977: How much rain does a rain gauge gage? *J. Hydrol.*, **35**, 213-220.

Prentice, R. A., 2016: Radar and applications course. Warning Decision Training Division, National Weather Service, 1034 pp.

Red Lake Watershed District, 2006: Red Lake Watershed District 10-Year Comprehensive Plan, 309 pp.

Rinehart, R. E., 2010: *Radar for Meteorologists*, Rinehart Publications, 482 pp.

Rogers, P., J. Kaiser, D. Kellenbenz, and M. Ewen. 2013: A comparative hydrometeorological analysis of the 2009, 2010, and 2011 Red River of the North Basin spring floods. *National Weather Service Central Region Tech. Attachment.*

Schilling, W., 1991: Rainfall data for urban hydrology: what do we need? *Atmos. Res.*, **27**, 5-22.

Schwert, D., 2011: North Dakota State University, Geology of the Fargo-Moorhead Region. http://www.ndsu.edu/fargo_geology/flashflood.html (last accessed 20 October 2015).

Seliga, T.A., and V. N. Bringi, 1976: Potential use of radar reflectivity at orthogonal. Polarizations for measuring precipitation. *J. Appl. Meteor.*, **15**, 69-76.

Shepherd, J. M., O. O. Taylor, and C. Garza, 2004: A dynamic GIS-Multicriteria technique for siting the NASA-Clark Atlanta urban rain gauge network. *J. Atmos.Oceanic Tech.,* **21**(9), 1346-1363.

Speltz, D. J., 1992: A comparison of radar rainfall estimates and rain gage measurements during two Denver thunderstorms. M. S. Thesis, Dept. of Atmospheric Science, Colorado State University, 83 pp. (Available from http://discovery.library.colostate.edu/Record/.b17702987)

Stellman, K. M., H. E. Fuelberg, R. Garza, and M. Mullusky, 2001: An examination of radar and rain gauge-derived mean areal precipitation over Georgia watersheds. *Weather and Forecasting*. **16**, 133-144.

Tillman, J. A., 1975: The Power of the Durbin-Watson Test. *Econometrica*., **43**(5/6) 959-974.

Todhunter, P. E., 2001: A hydroclimatological analysis of the Red River of the North snowmelt flood catastrophe of 1997. *J. American Water Resources Assoc.* **37**(5): 1263-1278.

---. 2011: Caveant admonitus (“let the forewarned beware”): The 1997 Grand Forks (USA) flood disaster. *Disaster Prevention and Management.* **20**(2): 125-139.

United States Census Bureau, 2015: Population for Crookston, Minnesota. http://quickfacts.census.gov/qfd/states/27/2713870.html (last accessed 21 October 2015).

Zhang, X., and R. Srinivasan, 2008: GIS based spatial precipitation estimation a comparison of geostatistical Approaches. *J. American Water Resources Assoc.,* **45**, 894-906.

---, 2010: GIS-based spatial precipitation estimation using next generation radar and raingauge data. Environmental Modelling and Software.