THESIS

IMPROVING RADAR QUANTITATIVE PRECIPITATION ESTIMATION THROUGH OPTIMIZING RADAR SCAN STRATEGY AND DEEP LEARNING

Submitted by
Liangwei Wang
Department of Electrical and Computer Engineering

In partial fulfillment of the requirements
For the Degree of Master of Science
Colorado State University
Fort Collins, Colorado
Summer 2024

Master's Committee:

Advisor: Haonan Chen

Venkatchalam Chandrasekaran
Haonan Wang

ABSTRACT

IMPROVING RADAR QUANTITATIVE PRECIPITATION ESTIMATION THROUGH OPTIMIZING RADAR SCAN STRATEGY AND DEEP LEARNING

As radar technology plays a crucial role in various applications, including weather forecasting and military surveillance, understanding the impact of different radar scan elevation angles is paramount to optimize radar performance and enhance its effectiveness. The elevation angle, which refers to the vertical angle at which the radar beam is directed, significantly influences the radar's ability to detect, track, and identify targets. The effect of different elevation angles on radar performance depends on factors such as radar type, operating environment, and target characteristics.

To illustrate the impact of lowering the minimum scan elevation angle on surface rainfall mapping, this article focuses on the KMUX WSR-88D radar in Northern California as an example, within the context of the National Weather Service's efforts to upgrade its operational Weather Surveillance Radar. By establishing polarimetric radar rainfall relations using local disdrometer data, the study aims to estimate surface rainfall from radar observations, with a specific emphasis on shallow orographic precipitation. The findings indicate that a lower scan elevation angle yields superior performance, with a significant 16.1% improvement in the normalized standard error and a 19.5% enhancement in the Pearson correlation coefficient, particularly for long distances from the radar.

In addition, conventional approaches to radar rainfall estimation have limitations, recent studies have demonstrated that deep learning techniques can mitigate parameterization errors and enhance precipitation estimation accuracy. However, training a model that can be applied to a broad domain poses a challenge. To address this, the study leverages crowdsourced data from NOAA and SFL, employing a convolutional neural network with a residual block to transfer knowledge learned from one location to other domains characterized by different precipitation properties. The experimental results showcase the efficacy of this approach, highlighting its superiority over conventional fixed-parameter rainfall algorithms.

Machine learning methods have shown promising potential in improving the accuracy of quantitative precipitation estimation (QPE), which is critical in hydrology and meteorology. While significant progress has been made in applying machine learning to QPE, there is still ample room for further research and development. Future endeavors in machine learning-based QPE will primarily focus on enhancing model accuracy, reliability, and interpretability while considering practical operational applications in hydrology and meteorology.

ACKNOWLEDGEMENTS

I would like to begin by expressing my profound gratitude to Dr. H. Chen. It has been a tremendous honor to work under his guidance and supervision. Without his invaluable mentorship and encouragement, this work would not have been possible. Dr. Chen has not only been my research mentor but also a role model in life.

I would also like to extend my sincere thanks to Drs. V. Chandrasekar and H. Wang for their service on my graduate committee.

I am deeply grateful for the assistance provided by my colleagues in the Artificial Intelligence and Remote Sensing Lab at Colorado State University.

Last but not least, I would like to thank my family and friends in China for their unwavering support, which has helped me overcome the challenges I have faced in the past several years.

DEDICATION

I would like to dedicate this thesis to my parents , advisor and friends.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Research Background and Purpose

Precipitation estimation is a crucial element in weather forecasting and climate studies as it provides vital information on the distribution, amount, and timing of precipitation. Precipitation is a fundamental component of the water cycle, which drives various Earth's ecosystems and supports human activities such as agriculture, transportation, and energy production. Thus, accurate precipitation estimation is crucial for various applications that depend on water resources, such as agriculture, water management, flood forecasting, and climate modeling [1].

Precipitation estimation is vital for agricultural practices, as precipitation is essential for crop growth and development. Accurate precipitation estimation can help farmers make informed decisions on when to plant crops, which crops to plant, and when to harvest them. In regions where water resources are limited, accurate precipitation estimation can help farmers manage water resources effectively, increasing crop yields, and reducing water usage. Furthermore, precipitation estimation can help in the prediction of droughts, enabling farmers to plan and prepare for drought conditions [2].

Precipitation estimation is also essential for water resource management, as it is the primary source of water for human activities, such as drinking, cooking, and hygiene. Precipitation estimation helps in forecasting water availability and water demand, allowing water managers to make informed decisions on water allocation, storage, and distribution [3]. In addition, precipitation estimation is crucial for the management of hydroelectric power plants, as it provides information on the water flow in rivers and reservoirs. Accurate precipitation estimation can help energy producers optimize the operation of hydroelectric power plants and reduce the risk of power shortages.

Another crucial application of precipitation estimation is flood forecasting and warning. Heavy rainfall can cause floods, resulting in loss of life and property damage. Precipitation estimation provides information on the amount and distribution of rainfall, enabling meteorologists to predict flood conditions and issue flood warnings to the public. Accurate and timely flood warnings can help authorities and individuals prepare for floods, evacuate affected areas, and reduce the risk of loss of life and property damage.

Precipitation estimation is also essential for climate modeling, as it provides input data for models that simulate the Earth's climate system. Precipitation is one of the most challenging variables to simulate in climate models, and accurate precipitation estimation is crucial for validating the models and improving their accuracy. Climate models are used to predict future climate conditions, including precipitation patterns, which have significant implications for water resources, agriculture, and human activities.

In conclusion, precipitation estimation plays a vital role in various aspects of human activities and the Earth's ecosystems. Accurate precipitation estimation is crucial for agricultural practices, water resource management, flood forecasting, and climate modeling. It provides information that is essential for decision-making, planning, and preparedness, enabling individuals, authorities, and businesses to mitigate the risks associated with precipitation extremes. Therefore, continued research and development in precipitation estimation methods and technologies are necessary to improve our understanding of precipitation and its impacts on the Earth's ecosystems and human activities [4].

1.1.1 Research Background

Precipitation is a crucial weather phenomenon that plays a significant role in various aspects of human activities and the Earth's ecosystems. Precipitation estimation is the process of quantitatively estimating the amount, distribution, and timing of precipitation. There are several methods available for precipitation estimation, each with its advantages and limitations [4].

The first method for precipitation estimation is station-based observation, which involves real-time monitoring of precipitation amount using precipitation observation equipment installed at stations. This method is highly accurate and has good real-time performance, but it is limited by the location and number of observation points and cannot accurately estimate precipitation over a wide area.

The second method is radar retrieval, which uses weather radar to detect precipitation particles in the atmosphere through the microwave signals emitted by the radar. Precipitation intensity and distribution can be inferred from the changes in signal reflection and scattering. This method has high spatiotemporal resolution and wide coverage, but it is less effective in detecting light precipitation such as drizzle and snow.

The third method is satellite remote sensing, which uses satellite remote sensing technology to analyze the reflection and emission of radiation from clouds, enabling the inference of precipitation amounts in the cloud layer. This method has the advantages of wide coverage and real-time performance, but it is prone to errors in low cloud layers and complex terrain.

The fourth method for precipitation estimation is model-based simulation, which uses numerical models to simulate the physical processes of precipitation formation and evolution. This method can provide high-resolution precipitation estimates with good accuracy, but it requires high computational resources and data inputs, and it is subject to uncertainties in model physics and parameterization.

In recent years, machine learning techniques have been applied to precipitation estimation, leveraging the vast amounts of available data and advanced algorithms to improve the accuracy and efficiency of precipitation estimation. These techniques include artificial neural networks, decision trees, support vector machines, and deep learning models, which can learn complex patterns and relationships in precipitation data and provide more accurate and reliable precipitation estimates.

Precipitation estimation plays a critical role in various fields, including weather forecasting, hydrology, agriculture, and environmental management. Accurate precipitation estimation can improve the understanding and prediction of severe weather events, mitigate the impacts of droughts and floods, and support the sustainable management of water resources and ecosystems.

1.1.2 Research Purpose

The purpose of this thesis is improving radar QPE to enhance the accuracy, reliability, and timeliness of precipitation measurements, which has significant implications for various sectors and applications.

In the former part of this thesis, we focus on studying how the reduction of radar scanning elevation angles enhances the accuracy of precipitation estimation.

The National Weather Service (NWS) has implemented an upgrade to its Weather Surveillance Radar – 1988 Doppler (WSR-88D) scan strategy, aiming to improve the radar's hydrometeorological capabilities. Some WSR-88D stations have already reduced their lowest scan elevation angle from 0.5° to 0°, with the possibility of lowering it further in the future. This change can affect the accuracy of surface rainfall mapping, particularly for shallow orographic precipitation. This article focuses on the KMUX WSR-88D radar located in the mountainous region of Northern California, using it as an example to analyze the effects of the lower scan elevation angle on surface rainfall mapping.

To estimate surface rainfall from radar observations, the article establishes polarimetric radar rainfall relations using local disdrometer data. The researchers apply these relations to the KMUX radar observations at 0° and 0.5° scan elevation angles to derive rainfall estimates. Comparison of these radar-based rainfall estimates with surface rain gauge measurements shows that the lower scan elevation angle improves the accuracy of surface rainfall mapping. Specifically, the 0° scan angle yields an improvement of 16.1% and 19.5% in the normalized standard error (NSE) and the Pearson correlation coefficient ($CORR$), respectively, compared to the 0.5° scan angle when the radar is observing within or above the melting layer at a long distance from the radar.

Then, in the later part of this thesis. The traditional approaches to using polarimetric radar measurements have limitations in extracting information, and conventional radar rainfall estimation algorithms heavily rely on the size distribution of raindrops. Fixed-parameter radar rainfall relations may contain inherent parameterization errors that are difficult to remove. Recent research suggests that deep learning techniques can improve radar-based precipitation estimation by reducing these errors.

To address this issue, this study leverages crowdsourced data from the NOAA and SFL, which provide weather and hydrologic observations feeding NWS operational models. This paper uses a convolutional neural network with a residual block as a benchmark to combat model degradation caused by increased depth. The proposed technique manipulates the training process to transfer knowledge learned at one location to other domains with different precipitation properties. Experimental results demonstrate that this approach improves precipitation estimation compared to conventional fixed-parameter rainfall algorithms.

1.2 Outline of Thesis

The entire thesis is divided into five chapters, and the specific content arrangement of each chapter is as follows:

Chapter 1: The focus of this chapter is to provide a comprehensive overview of the background and significance of quantitative precipitation estimation and the research conducted using radar-based quantitative precipitation estimation. Firstly, it provides a detailed explanation of the historical context and development of this field of study. Secondly, it presents the necessity and purpose of the research, highlighting the gaps in the existing knowledge and the potential implications for various sectors. Lastly, it outlines the research direction of the paper, along with its specific objectives, and gives an overview of the thesis structure, enabling the reader to gain a clear understanding of the subsequent chapters. The overall aim of this chapter is to establish a solid foundation for the research, setting the stage for the in-depth exploration of the topic.

Chapter 2: This chapter provides a comprehensive overview of the key methods employed in radar-based quantitative precipitation estimation, as well as the current state of research in this area. It also explores the potential applications and advancements in machine learning techniques for improving precipitation estimation accuracy. Additionally, the chapter highlights the limitations of the conventional radar-precipitation relationships used in this field, emphasizing the need for more advanced methods to address the challenges associated with complex precipitation patterns and data processing. By covering these aspects in detail, this chapter aims to provide the reader with a thorough understanding of the current state of research and the opportunities and challenges that lie ahead in this field.

Chapter 3: This chapter focuses on the efforts by the National Weather Service to enhance the hydrometeorological applications of the operational Weather Surveillance Radar – 1988 Doppler (WSR-88D), by upgrading its scan strategy. Specifically, the impact of reducing the minimum scan elevation angle from 0.5° to 0° (or even lower in the future) is investigated using the KMUX WSR-88D radar located in mountainous terrain in Northern California. The chapter presents an approach for surface rainfall mapping, emphasizing shallow orographic precipitation, using polarimetric radar rainfall relations derived from local disdrometer data. The performance of the rainfall estimates obtained from the KMUX radar observations at 0° and 0.5° scan elevation angles is compared using rainfall measurements from surface rain gauges, with the results indicating the superior performance of the lower scan elevation angle. By providing a detailed analysis of the impact of the upgraded scan strategy on rainfall estimation accuracy, this chapter contributes to the ongoing efforts to improve the hydrometeorological applications of the WSR-88D radar. In addition, this chapter provides a theoretical basis for selecting radar observation data at the two lowest elevation angles as input data for the next chapter.

Chapter 4: The primary focus of this chapter is to propose an advanced radar-based model for quantitative precipitation estimation using deep separable convolution and Transformer structures for dual-polarized radar data. The input data is derived from the Next Generation Weather Radar System dual-polarized radar data of the National Oceanic and Atmospheric Administration (NOAA) WSR-88D Doppler weather radar network, while the rain gauge data used as reference truth comes from the Global Precipitation Measurement Ground Validation (GPM GV) data provided by the National Aeronautics and Space Administration (NASA). To benchmark the proposed model, the traditional Z-R relationship algorithm and a quantitative precipitation estimation model based on multi-layer perceptron are used for comparison. The main approach of the proposed model involves inputting the dual-polarized radar data within a specific range into the model, which integrates the spatial and dual-polarized information to output the precipitation intensity value at the center point of the range. The model's performance is evaluated by accumulating precipitation amounts over an hour to assess its effectiveness. Overall, this chapter contributes to the development of more accurate and efficient models for quantitative precipitation estimation, which is crucial for various hydrometeorological applications.

Chapter 5: This chapter serves as a comprehensive summary of the research conducted in this thesis, providing an overview of the proposed model for radar-based quantitative precipitation estimation using deep learning techniques. It analyzes the feasibility and practical value of the model and discusses its advantages and limitations. In addition, it highlights the limitations of the model and the current research status of quantitative precipitation estimation, while considering the development trends in the field of deep learning. Based on this analysis, future research prospects and potential directions for further exploration are proposed. This chapter aims to offer a critical assessment of the proposed model and identify areas of improvement, as well as highlight opportunities for future research that could lead to more accurate and efficient models for precipitation estimation, which can play a critical role in hydrometeorological applications.

Chapter 2

Radar Network

2.1 Radar Basics

2.1.1 Weather Radar Development

The development of weather radar can be traced back to the early 20th century, with significant advancements made over the years. Here is a brief history of weather radar development:

In the 1930s, scientists began experimenting with radar technology to detect precipitation in the atmosphere. The first successful weather radar experiment took place in the late 1940s when researchers detected echoes from thunderstorms using a radar system.

In the 1950s, the Weather Bureau (now known as the National Weather Service) deployed the WSR-57 radar network across the United States. These early radars provided limited coverage and had limited capabilities. In the 1970s, the WSR-57 was replaced by the more advanced WSR-74 radars, which improved data quality and coverage.

The most significant milestone in weather radar development came with the deployment of the Weather Surveillance Radar-1988 Doppler (WSR-88D) network in the late 1980s and early 1990s. The WSR-88D, also known as the NEXRAD (Next Generation Weather Radar), revolutionized weather forecasting and severe weather detection. It provided higher resolution, Doppler velocity data, and improved scanning strategies compared to its predecessors.

In the 2000s, the implementation of dual-polarization radar technology became a major advancement in weather radar systems. Dual-polarization radars transmit and receive both horizontal and vertical polarizations, providing valuable information about the shape, size, and composition of precipitation particles. This technology greatly enhanced precipitation estimation, severe weather detection, and hydrometeor classification.

In recent years, phased array radar systems have emerged as the next generation of weather radar technology. Phased array radars use electronically controlled beams to scan the atmosphere, allowing for faster scanning, improved data quality, and the ability to track multiple storms simultaneously. These radars offer enhanced severe weather detection capabilities and are being gradually deployed across the globe.

The continuous evolution of weather radar technology has significantly improved our understanding and prediction of weather patterns, leading to better forecasts, severe weather warnings, and overall public safety. Ongoing research and development efforts continue to refine radar systems, enabling more accurate and timely weather information for various applications.

2.1.2 Radar-related Parameters

Radar-related parameters refer to the various measurements and characteristics used in radar systems to detect, analyze, and interpret weather phenomena. These parameters are essential in understanding and forecasting weather patterns, particularly in the context of precipitation, wind, and other atmospheric conditions.

One crucial radar-related parameter is "reflectivity," which quantifies the intensity of the energy returned to the radar from various objects in the atmosphere. Reflectivity is used to estimate the presence and intensity of precipitation, as well as distinguish between different precipitation types, such as rain, snow, or hail.

Another significant parameter is "velocity," (Z) which measures the speed at which objects in the atmosphere are moving towards or away from the radar. By analyzing the Doppler shift of the radar signal, velocity data can be used to identify wind patterns and infer the presence of rotating features like tornadoes.

"Differential reflectivity"(Z_{dr}) is a parameter that compares the strength of the horizontally and vertically polarized radar signals. It helps in identifying the shape and type of precipitation particles, such as raindrops or snowflakes, which provides valuable information about the precipitation processes.

"Correlation coefficient" is another crucial radar-related parameter that indicates the uniformity of hydrometeors (precipitation particles) in the radar volume. It helps in distinguishing between meteorological targets and non-meteorological clutter, like birds, insects, or ground clutter.

"Polarimetric variables" encompass a set of parameters that provide additional information about the shape, orientation, and composition of precipitation particles. These variables include specific differential phase (K_{dp}), specific attenuation (A), and cross-correlation coefficient (ρ_{hv}), among others.

2.2 WSR-88D radar network

Modern radar networks refer to a collection of radar systems that are interconnected and work together to provide comprehensive surveillance and tracking capabilities over a particular region or area. These networks are typically used for various applications, including weather monitoring, air traffic control, defense and security, and maritime surveillance.

One prominent example of a radar network is the WSR-88D (Weather Surveillance Radar 1988 Doppler) network, also known as the NEXRAD (Next Generation Radar). The WSR-88D radar network is operated by the National Weather Service (NWS) in the United States and is used primarily for weather monitoring and forecasting.

2.2.1 What is WSR-88D radar network

The NWS employs the WSR-88D radar, denoting Weather Surveillance Radar-1988 Doppler, with its name reflecting the year of its inception. This radar operates based on Doppler principles, allowing it to not only pinpoint the location and intensity of precipitation but also track movements towards or away from the radar. The continuous advancements in radar technology have equipped NWS meteorologists with the tools to meticulously study storms, offering heightened accuracy in identifying rotation or hail within the precipitation zone of a thunderstorm. Radar algorithms play a pivotal role, aiding meteorologists in gauging rotation strength and depth, as well as discerning precipitation type and intensity, thereby facilitating informed decisions for issuing warnings.

2.2.2 Operation of WSR-88D Radars

The WSR-88D acquires weather data, including precipitation and wind information, by emitting bursts of energy from its 28-foot diameter antenna located inside the radome (the white, soccer ball covering). This energy is transmitted as a signal and, when it encounters objects such as raindrops, snow, hail, bugs, birds, or dust, it scatters in all directions. A small portion of this scattered energy is then directed back towards the radar.

The radar's antenna, which initially sent the signal, receives the reflected signal during its listening period. This received signal is then sent to a computer system situated in a small building at the base of the radome. These computers analyze the strength of the returned pulse, the time it took for the pulse to travel to the object and back, and the phase shift of the pulse. This process of emitting a signal, listening for the returned signal, and then emitting the next signal happens rapidly, up to approximately 1300 times per second.

The majority of the time, the WSR-88D is "listening" for the returning signals it sent. When the total time of all the transmitted pulses within an hour is calculated (the time the radar is actually transmitting), the radar is "on" for approximately 7 seconds each hour. The remaining 59 minutes and 53 seconds are spent in listening mode, waiting for any returned signals.

The WSR-88D is considered a Doppler radar due to its ability to detect the "shift in the phase" of the energy pulse. The phase of the returning signal typically changes based on the motion of the raindrops (or other objects like bugs, dust, etc.). This phenomenon is known as the Doppler effect, named after the Austrian physicist, Christian Doppler, who discovered it. The Doppler effect is experienced, for example, when a train passes by, and the pitch of its whistle changes from high to low as it approaches and then moves away from an observer.

Similarly, in the atmosphere, when the radar's energy pulse strikes an object and is reflected back towards the radar, the radar's computers measure the phase change of the reflected pulse. This change in phase is then converted into a velocity of the object, either towards or away from the radar. By analyzing the movement of objects relative to the radar, the radar can estimate wind speed. This ability to detect wind patterns enables the National Weather Service to identify the formation of tornadoes, thereby allowing them to issue tornado warnings with more advanced notice.

2.3 Radar QPE Algorithm Description

The estimation of rainfall rate and amount has been crucial throughout human history. Radar technology has proven to be a powerful tool for this purpose, as it allows us to observe precipitation over large areas in a relatively short time, and it provides advanced warning of impending precipitation systems that might affect a region. Radar systems have been utilized to detect precipitation echoes since the early stages of World War II. The concept of radar QPE can be traced back to the pioneering work of Wexler and Marshall, among others. Cifelli and Chandrasekar further categorized the factors influencing radar QPE into two main groups: the basic science aspect and the applied engineering aspect. To accurately estimate rainfall, both these aspects must be addressed. A comprehensive account of the various factors related to basic and physical science that impact radar quantitative precipitation estimation was presented in a previous study.

In this subsections, we will provide an overview of the current radar QPE algorithms, including descriptions of different raindrop size distribution (DSD) models and various parameter-based rainfall algorithms.

2.3.1 Raindrop Size Distribution (DSD)

The raindrop size distribution (DSD) refers to the probability density distribution function of raindrop sizes, which serves as the foundation for deriving physically based rainfall algorithms. Analyzing the variability of DSD properties and determining whether they arise randomly or are associated with specific physical processes can offer valuable insights into the microphysical and dynamic cloud processes involved in the formation of precipitation particles. An accurate understanding of DSD within a precipitating system is crucial for radar-based quantitative precipitation estimation (QPE) and quantitative precipitation forecasting (QPF). Different rainfall algorithms can be developed based on diverse DSD models, utilizing radar observations such as reflectivity, differential reflectivity, and/or specific propagation phase. DSD information is also vital for research in simulating rainfall fields.

2.3.2 Different Rainfall Algorithms

2.3.2.1 Traditional $Z - R$ Algorithm

Traditional rainfall measurements from radars, especially single-polarization radars, often start with this reflectivity-rainfall relation, commonly referred as the $Z - R$ relation. If we let N(D) denote the DSD, where D is the equivalent drop size in diameter, the reflectivity factor Z can be defined as

$$Z = \int_D D^6 N(D) dD \tag{2.1}$$

It is commonly expressed in the units of $\mathrm{mm}^6\mathrm{m}^{-3}$. However, the logarithmic transformation $10\log_{10}(Z)$ is often used since the precipitation particles can vary in diameter over many orders of magnitude, and its units are in decibels of Z relative to 1 $\mathrm{mm}^6\mathrm{m}^{-3}$ (which is 0 dBZ).

The still-air rainfall rate (in mm/h) can be written as a relationship with the DSD as,

$$R = 0.6\pi \times 10^{-3} \int v(D) D^3 N(D) dD \tag{2.2}$$

Again, D is diameter in mm, $N(D)dD$ is the number of drops m^{-3} in the interval from D to $D + dD$, and $v(D)$ is the drop terminal velocity (in $\mathrm{m/s}$) at sea level, which depends on parameters such as air density, size and shape of raindrops, etc. The terminal velocity of raindrops can be expressed as a form given by Atlas et. al,

$$v(D) = 9.65 - 10.3 \exp(-0.6D) \tag{2.3}$$

or a power law form as

$$v(D) = \alpha D^{\beta} \tag{2.4}$$

where the coefficient α varies from 3.6 and 4.2, and β from 0.6 to 0.67 for precipitation events. For rainfall rate estimations, the most commonly used sea-level values for α and β are 3.78 and 0.67, respectively. Based on this model, radar-based rainfall algorithms can be developed. All the rainfall algorithms described in this chapter are assumed to apply at sea level.

From equations 2.1 and 2.2, it can be seen that both Z and R can be approximated as moments of the DSD. Subsequently, the $R(Z)$ relation can be expressed in a parametric form as

$$Z = aR^b \tag{2.5}$$

where A and b are empirically derived constants. A list of coefficients recommended by the National Weather Service for use in the U.S. can be found in 2.1

Table 2.1: Coefficients for usage of $Z - R$ realtions

Name of formula	a	b	Optimum for:
Marshall-Palmer Stratiform	200	1.6	General stratiform rainfall
Cool Season (East) Stratiform	130	2.0	East of continental divide
Cool Season (West) Stratiform	75	2.0	West of continental divide
Summer Deep Convection	300	1.4	WSR-88D Convective
Rosenfeld Tropical	250	1.2	Tropical convective systems

2.3.2.2 $R - K_{dp}$ Approach

The advancement of dual-polarization techniques, exemplified by the upgrade of the NWS WSR-88D to dual-pol capabilities, has significantly improved our ability to discern the type and intensity of rainfall. While challenges in radar rainfall estimation persist, dual-polarization observations offer several advantages compared to single polarization radar rainfall estimation, as they provide more comprehensive information about the drop size distribution (DSD). Additionally, dual-polarimetric estimators are better equipped to account for the presence of ice in precipitation.

It was shown that the specific differential propagation phase K_{dp} (in degree km^{-1}) is proportional to the product of rainwater content and the mass-weighted mean diameter of raindrops,

$$K_{dp} \approx \left(\frac{180}{\lambda}\right) 10^{-3} CW \left(0.062 D_m\right) \tag{2.6}$$

Again, D_m is the mass-weighted mean diameter, whereas W is the rainwater content in units of g/m^3 , and λ is radar wavelength in m.

Reusing the relationship between rainfall rate and DSD (equation 2.2), a general R(K dp) estimator can be derived in the following form,

$$R\left(K_{dp}\right) = aK_{dp}^{b} \tag{2.7}$$

where the constants a and b are dependent on different climatic and synoptic properties.

2.3.2.3 Hybrid Algorithm

With the aid of dual-polarimetric observations, researchers often use more than one radar parameter to conduct rainfall estimations. In this thesis, rainfall algorithms based on more than one polarimetric radar parameter such as reflectivity (Z), differential reflectivity (Z_{dr}), and specific differential phase (K_{dp}) are all classified as hybrid algorithms. Several hybrid algorithms, including R(Z, Z_{dr}) and R(Z_{dr}, K_{dp}) based approaches, will be described in this section.

These variables eventually lead to the derivation of the relations between new radar parameters and precipitation as

$$R\left(Z, Z_{dr}\right) = cZ^{a} Z_{dr}^{b} \tag{2.8}$$

$$R\left(Z_{dr}, K_{dp}\right) = cZ_{dr}^{a} K_{dp}^{b} \tag{2.9}$$

2.4 Radar Research Status and Methodology

Weather radar observations offer high spatiotemporal resolution of precipitation information, allowing for the identification of regional heavy rainfall in advance. With a single weather radar device covering an expansive area of several hundred kilometers, it becomes possible to anticipate such rainfall. Furthermore, the three-dimensional data provided by weather radar facilitates the study of precipitation system dynamics and vertical structure. However, directly observing precipitation intensity through weather radar poses a challenge, resulting in extensive research in meteorology. The objective of this research is to accurately estimate precipitation from weather radar observations. Currently, two primary approaches are pursued for quantitative precipitation estimation based on weather radar. The first approach employs various methods to estimate precipitation using observed weather radar data, while the second approach involves adjusting radar precipitation estimation through data fusion, incorporating data from meteorological satellites and rain gauges. The subsequent sections will delve into a comprehensive overview of the ongoing research on radar based quantitative precipitation estimation in the United States.

Research on radar-based quantitative precipitation estimation (QPE) has been ongoing for several decades in the United States. One of the key challenges in QPE research is to accurately estimate precipitation intensity, which is not directly observable by weather radar. To overcome this challenge, various methods have been developed and used in practice.

Quantitative Precipitation Estimation (QPE) employs different methods to accurately assess precipitation levels. Among these methods, the Z-R relationship is widely utilized, establishing a correlation between radar reflectivity (Z) and rain rate (R). This approach assumes the consistency of the Z-R relationship across time and space, proving effective for light to moderate rainfall. However, it may not accurately estimate heavy rainfall events, necessitating alternative techniques.

The dual-polarization radar technique is another method employed for QPE. It utilizes both horizontal and vertical polarization to measure precipitation particle size, shape, and orientation. This technique provides more precise information regarding precipitation intensity and type, particularly for heavy rainfall. The National Weather Service (NWS) in the United States has adopted this method within its operational radar network.

Deep learning has emerged as a popular tool for improving radar-based QPE accuracy. Convolutional neural networks (CNNs) are employed in one approach to extract features from radar images, aiding in the distinction of precipitation from other atmospheric objects. These features subsequently contribute to generating precipitation estimates.

Deep learning offers an advantage in capturing complex relationships between input data and output precipitation values, especially for the highly nonlinear association between radar reflectivity and precipitation rate.

Several studies have demonstrated the effectiveness of deep learning in enhancing radar-based QPE. For instance, one study employed a CNN to predict hourly precipitation amounts based on radar reflectivity data, yielding a higher correlation coefficient compared to traditional methods. Another study utilized a deep neural network (DNN) to estimate precipitation using radar and rain gauge data, highlighting improved accuracy compared to conventional approaches.

However, challenges exist when utilizing deep learning for radar QPE. Acquiring substantial amounts of training data is necessary to ensure accurate model performance. Additionally, deep learning models can be computationally intensive, requiring significant computing resources.

Overall, ongoing research on radar-based QPE in the United States encompasses the development and application of various methods. It also involves integrating diverse data sources to enhance accuracy and reliability in precipitation estimation.

2.4.1 Dual-polarization Weather Radar based QPE

Quantitative Precipitation Estimation (QPE) through dual-polarization weather radar is a sophisticated technique that enhances the accuracy of measuring and assessing precipitation levels. This method involves the utilization of weather radar systems that employ both horizontal and vertical polarization signals. Unlike traditional single-polarization radars that solely measure the horizontal orientation of precipitation particles, dual-polarization radars also consider the vertical alignment of these particles.

Precise, punctual, and dependable measurements of precipitation are imperative for hydrological predictions and early warning systems. In scenarios involving convective precipitation, conventional networks of rain gauges frequently fail to capture the highest levels of precipitation due to limitations in density and the substantial spatial fluctuations within rainfall patterns. Despite challenges such as signal attenuation or partial obstruction, the incorporation of C-band weather radar has been implemented by most European meteorological services. Historically, quantitative precipitation estimation (QPE) based on weather radar has been derived from horizontal reflectivity data. However, the utilization of dual-polarization weather radar can address numerous deficiencies inherent in the conventional approach reliant on horizontal reflectivity. With the expansion of weather radar archives, their significance is growing not only for operational purposes but also for climatological analyses.

Quantitative precipitation estimation (QPE) using dual-polarization weather radar has a long and rich history. It all started in the late 1970s when researchers first realized that the differential reflectivity (Z_{dr}) measured by a dual-polarization radar could provide valuable information about the type of precipitation. The first significant breakthrough came in the early 1990s when researchers realized that Z_{dr} could be used to estimate the equivalent radar reflectivity factor (Z) of rain and snow. This paved the way for the development of dual-polarization radar-based QPE algorithms that took into account the type of precipitation.

In the mid-1990s, several researchers, including James Zrnic and V. N. Bringi, made significant contributions to the field by developing algorithms that could estimate rainfall rate using Z and Z_{dr}. They showed that this approach was more accurate than the traditional Z-based QPE algorithms and could be used to estimate precipitation over a wide range of rainfall rates and precipitation types.

In the early 2000s, researchers continued to refine the dual-polarization radar-based QPE algorithms. In 2003, Bringi and Chandrasekar introduced a new parameter, specific differential phase (K_{dp}), which could be used to estimate the rainfall rate even in regions of heavy attenuation.

In 2017, Haonan Chen contributed to the development of an improved algorithm for quantitative precipitation estimation using dual-polarization radar data. The algorithm, called DROPS2.0, utilizes differential reflectivity and specific differential phase measurements to estimate rainfall rates with higher accuracy than previous methods.

Chen's contributions include modifying the original DROPS algorithm to account for the effects of horizontal reflectivity gradients, improving the estimation of specific differential phase using a sliding-window technique, and introducing a new rain type classification scheme based on reflectivity and differential reflectivity measurements. These improvements resulted in more accurate rainfall estimates, particularly in areas with complex terrain or where stratiform and convective precipitation mix.

Since then, many researchers have made significant contributions to the field of dual-polarization radar-based QPE, including improving the accuracy of the algorithms, developing new techniques to deal with the effects of beam broadening, and developing new methods to estimate precipitation type. Today, dual-polarization radar-based QPE is widely used by meteorologists and hydrologists around the world to estimate rainfall rates and help predict floods and other weather-related hazards.

2.4.2 Machine Learning Based QPE

In recent years, deep learning techniques, particularly those based on neural networks, have been increasingly applied to weather radar quantitative precipitation estimation (QPE). Compared to traditional QPE algorithms, neural network-based methods can improve the accuracy of precipitation estimation, particularly in regions with complex topography or sparse rain gauges.

In the early 2010s, researchers began exploring the application of neural networks to QPE. These initial studies focused on using artificial neural networks (ANNs) to estimate precipitation using reflectivity measurements from single-polarization radar data. ANNs were found to be effective in addressing some of the limitations of traditional QPE algorithms, such as the inability to account for non-linear relationships between reflectivity and precipitation, or to accurately estimate rainfall in regions with complex terrain. However, ANNs still had limitations in terms of their ability to account for variability in precipitation processes, particularly in regions with a wide range of precipitation types.

To address these limitations, researchers began exploring the use of dual-polarization radar data in neural network-based QPE algorithms. Dual-polarization radar data provides additional information on the size, shape, and orientation of precipitation particles, which can be used to more accurately estimate precipitation intensity and type. In 2013, a study by K. H. Lee et al. demonstrated that a neural network-based QPE algorithm using dual-polarization radar data could achieve higher accuracy than traditional Z-R relationships, particularly in regions with mixed-phase precipitation.

In the years following, neural network-based QPE algorithms continued to improve. Researchers developed more sophisticated neural network architectures, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), to better account for spatial and temporal variability in precipitation processes. CNNs are particularly effective in processing large volumes of data, such as those generated by weather radar scans. RNNs, on the other hand, are well-suited to modeling temporal relationships in precipitation processes, such as the evolution of precipitation systems over time.

In addition to improving neural network architectures, researchers also developed new techniques for training neural networks using large volumes of radar data. One technique, known as transfer learning, involves pre-training a neural network on a large dataset of radar data before fine-tuning it on a smaller, more specific dataset. Another technique, known as adversarial training, involves training a neural network to generate realistic synthetic radar data that can be used to augment existing datasets.

In recent years, the application of neural network-based QPE has been further enhanced by the development of more advanced radar technologies, such as phased array radar (PAR) and dual-frequency radar (DFR). PAR systems use an array of small antennas to scan the atmosphere, providing high-resolution data that can be used to more accurately estimate precipitation intensity and location. DFR systems, on the other hand, use two radar frequencies to better estimate the size and shape of precipitation particles.

In conclusion, neural network-based QPE has made significant strides in the past decade, driven by advancements in neural network architectures and training techniques, as well as the development of new radar technologies. These advances have enabled more accurate and reliable estimation of precipitation intensity and type, particularly in regions with complex terrain or sparse rain gauge networks. Moving forward, continued research into neural network-based QPE will likely focus on improving the accuracy and efficiency of algorithms, as well as integrating additional data sources, such as satellite data and ground-based sensors, to further enhance precipitation estimation.

2.4.3 Limitation of Radar QPE

Radar quantitative precipitation estimation (QPE) has been widely used for precipitation monitoring, forecasting, and hydrological applications due to its high spatial and temporal resolution. However, radar QPE also has several limitations that should be taken into consideration.

Firstly, radar QPE is subject to various uncertainties caused by the complex nature of precipitation. Precipitation can vary in intensity, duration, and spatial extent, making it difficult to accurately estimate the precipitation rate. The attenuation of the radar signal as it passes through the precipitation also introduces errors in the QPE, as the signal can be absorbed or scattered by the precipitation particles, leading to underestimation of the precipitation rate. In addition, the radar beam can be affected by terrain, atmospheric conditions, and the presence of other obstacles, which can further affect the accuracy of the QPE.

Secondly, radar QPE has limitations in detecting and estimating certain types of precipitation. For example, light precipitation, such as drizzle and light snow, may not be detectable by the radar due to their small size and low reflectivity. Similarly, frozen precipitation, such as sleet and freezing rain, can also be difficult to detect and estimate due to their different scattering properties compared to liquid precipitation. In addition, radar QPE may not be able to accurately estimate precipitation in mountainous regions due to the complex terrain and the effects of the orography on the radar beam.

Thirdly, radar QPE has limitations in estimating the precipitation rate over long periods of time. The accuracy of the QPE may degrade over time due to changes in the precipitation type and intensity, as well as changes in the atmospheric conditions. Moreover, the accuracy of the QPE may also be affected by the temporal and spatial resolution of the radar data, as well as the accuracy of the radar calibration.

Finally, radar QPE is limited by the availability and quality of the radar data. The accuracy of the QPE depends on the quality of the radar data, including the sensitivity, calibration, and data processing techniques. Moreover, the availability of the radar data may be limited in some regions due to the lack of radar coverage, technical issues, or data transmission problems.

In conclusion, while radar QPE is a useful tool for precipitation monitoring and forecasting, it also has limitations that should be taken into consideration. These limitations include uncertainties in the QPE caused by the complex nature of precipitation, limitations in detecting and estimating certain types of precipitation, limitations in estimating the precipitation rate over long periods of time, and limitations in the availability and quality of the radar data. Researchers and practitioners should be aware of these limitations and continue to develop and improve radar QPE techniques to overcome these challenges.

2.5 Introduction of Estimation Formulas

In this thesis, we have utilized numerous evaluation metrics, we will introduce them one by one in this section.

2.5.1 Mean Absolute Error

MAE, which stands for Mean Absolute Error, is a widely used metric in statistics and machine learning to evaluate the accuracy of a predictive model or estimate. It measures the average absolute difference between the predicted values and the actual values in a dataset.

The formula for calculating MAE is straightforward: the absolute difference between each predicted value and its corresponding actual value is calculated, and then the average of these absolute differences is computed. The result gives us a single value that represents the average magnitude of the errors made by the model.

The formula to calculate the Mean Absolute Error (MAE) is as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2.10}$$

where: n is the number of data points or samples. y_i represents the actual or observed value of the i-th data point. $\hat{y}_i$ represents the predicted value of the i-th data point.

MAE is particularly useful when dealing with continuous data or regression problems, where the goal is to predict numerical values. It provides a simple, intuitive measure of how well the model performs on average in terms of absolute prediction errors. A lower MAE indicates better accuracy, as it means the model is making smaller absolute errors in its predictions.

One advantage of MAE is its resistance to outliers, as it considers the absolute magnitude of errors rather than the squared differences (as in the case of Mean Squared Error or MSE). This makes MAE more robust in situations where extreme values can heavily influence the model's performance.

2.5.2 Root Mean Square Error

Root Mean Square Error (RMSE) is a statistical metric commonly used to measure the accuracy and performance of a predictive model or an estimation method. It is widely applied in various fields, including data analysis, machine learning, and scientific research.

The RMSE represents the square root of the average of the squared differences between the predicted values and the actual observed values. By calculating the RMSE, one can quantify how closely the predicted values align with the true values, providing an indication of the model's or method's predictive accuracy.

In mathematical terms, the RMSE is expressed as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2.11}$$

where: n represents the number of data points or observations. y_i is the actual observed value for the i-th data point. $\hat{y}_i$ is the predicted value for the i-th data point.

A lower RMSE value indicates that the predictions are closer to the true values, signifying higher accuracy. Conversely, a higher RMSE value implies that the model or method has more significant prediction errors and may require improvement.

RMSE is a valuable tool for comparing different models or techniques and selecting the one that best fits the data. It helps researchers and practitioners assess the quality of their models and make informed decisions based on the level of accuracy they require for their specific applications.

2.5.3 Pearson's Correlation Coefficient

The correlation coefficient (CORR) is a statistical measure that quantifies the strength and direction of the linear relationship between two variables. In the context of radar meteorology, the correlation coefficient is used to assess the uniformity of hydrometeors (precipitation particles) within the radar volume. It is an essential parameter for distinguishing between meteorological targets, such as rain or snow, and non-meteorological clutter, such as birds, insects, or ground objects.

The correlation coefficient takes values between -1 and 1, where:

CORR=1 indicates a perfect positive linear relationship between the two variables. In radar terms, this means a high level of uniformity among the hydrometeors, suggesting genuine meteorological targets.

CORR=-1 indicates a perfect negative linear relationship between the two variables. In the context of radar, this might indicate strong interference or clutter, resulting in a low correlation of the signals.

CORR≈0 suggests a weak or no linear relationship between the variables, implying non-uniform or mixed hydrometeors, which might include both meteorological and non-meteorological targets.

The formula to calculate the correlation coefficient (CORR) between two variables X and Y can be expressed as:

$$\text{CORR} = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\cdot\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \cdot \sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}} \tag{2.12}$$

Where: X_i and $Y_i$are individual data points of variables X and Y, respectively. $\bar{X}$ and $\bar{Y}$are the mean values of X and Y, respectively. n is the total number of data points in the dataset.

Correlation is a valuable tool in data analysis and can provide insights into relationships between variables in various fields, including economics, finance, social sciences, and natural sciences. However, it is essential to remember that correlation does not imply causation, and additional analysis is required to establish any cause-and-effect relationships between variables.

2.5.4 Normalized Standard Error

In statistics, the normalized standard error (NSE) is a measure used to assess the precision or accuracy of an estimate relative to its magnitude. It is particularly useful when comparing estimates that have different scales or units of measurement. The NSE is expressed as a percentage, making it easier to interpret and compare across different estimates.

The formula for calculating the normalized standard error is as follows:

$$NSE = \left(\frac{SE}{|\bar{x}|}\right) \times 100 \tag{2.13}$$

Where: NSE is the normalized standard error (expressed as a percentage). SE is the standard error of the estimate. $|\bar{x}|$is the magnitude of the estimate, represented by the absolute value of the estimate.

The standard error (SE) measures the variability or spread of the estimate's sampling distribution. It indicates how much the estimate is likely to differ from the true population parameter. The absolute value of the estimate $|\bar{x}|$ is used to make sure the NSE is positive and to handle both positive and negative estimates.

A lower NSE value indicates a more precise estimate, meaning that the estimate is likely to be closer to the true value. Conversely, a higher NSE value suggests a less precise estimate with potentially larger variability from the true value.

The NSE is especially useful when comparing estimates from different datasets or when assessing the accuracy of predictions from different models, as it standardizes the precision assessment across different contexts.

2.5.5 Bias Ratio

Bias Ratio is a statistical measure used to assess the accuracy of a model or estimator by quantifying the difference between the expected or average predicted values and the actual values. It helps evaluate whether the model is systematically overestimating or underestimating the target variable.

The formula for Bias Ratio is as follows:

$$\text{Bias Ratio } = \frac{\text{Average Predicted Value}}{\text{Average Actual Value}} \tag{2.14}$$

Where: Average Predicted Value: The average of all predicted values generated by the model or estimator. Average Actual Value: The average of all actual values from the dataset used for training or testing the model.

The Bias Ratio can be greater than 1, indicating that the model tends to overestimate the target variable. Conversely, a value less than 1 suggests an underestimation by the model. Ideally, a Bias

Ratio close to 1 signifies that the model's predictions are on average quite accurate with respect to the actual values.

Bias Ratio is just one of the many metrics used in model evaluation. When combined with other evaluation metrics like Mean Squared Error (MSE) or Mean Absolute Error (MAE), it provides a more comprehensive understanding of the model's performance and helps in identifying potential biases in the predictions. It is crucial to consider the Bias Ratio, especially in situations where biased predictions could have significant real-world consequences.

Chapter 3

Improving Surface Rainfall Mapping through Lowering the Minimum Scan Elevation Angle

3.1 Aim of this Chapter

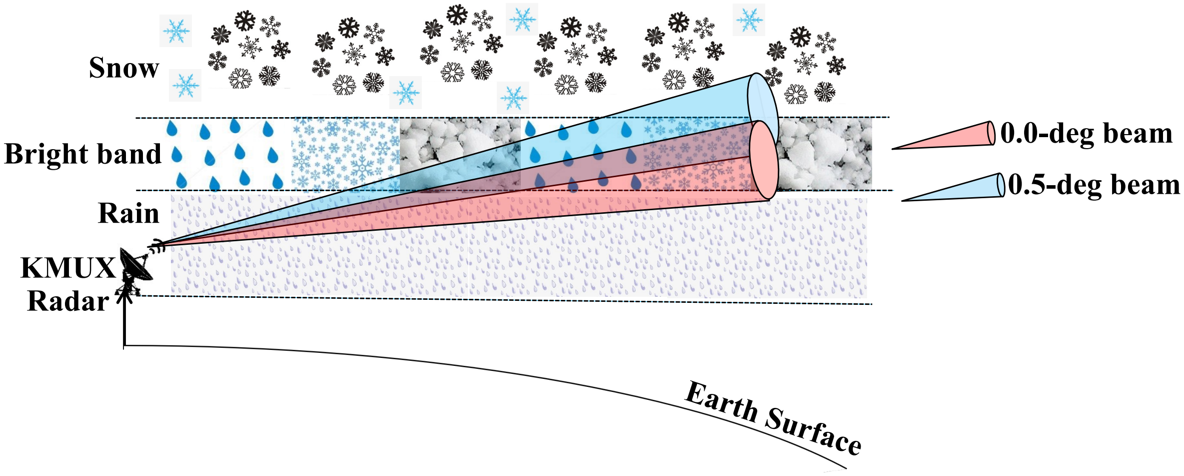

Figure 3.1: Diagram illustrating the sampling limitations of KMUX WSR-88D in Northern California.

The Weather Surveillance Radar–1988 Doppler (WSR-88D) network, also known as NEXRAD, is a critical component of severe weather warning and forecast infrastructure in the United States, providing essential monitoring of hydrometeorological conditions [5]. However, due to current regulations, the operational WSR-88D network is limited to a minimum scan elevation angle of 0.5°, which combined with the Earth's curvature, restricts its ability to observe the lower troposphere where many hazardous weather events occur [6] [7] [8].

In order to overcome this constraint and enhance the WSR-88D's capacity for surface rainfall mapping, the National Weather Service (NWS) has initiated a project to enhance the scan strategy of the WSR-88D. This involves incorporating lower scan elevation angles, with the potential inclusion of a 0° angle in the future. A pioneering implementation of this strategy can be observed at the KMUX radar, situated atop Mount Umanhum in the Santa Cruz Mountains of northern CA. The KMUX radar has been conducting scans at a 0° elevation angle since late 2018.

In the orographic precipitation system of Northern California, the melting layer is typically located at an altitude of 1-1.5 km above sea level (ASL) [9]. Since the KMUX radar is situated at an altitude of around 1 km ASL, the 0.5° beam often observes mixed-phase hydrometeors in the bright band or snowflakes above the bright band, even when it is raining on the ground (as illustrated in Fig. 3.1). This paper aims to quantify the improvements brought about by the enhanced WSR-88D scan strategy, using the KMUX radar as an example. Specifically, the study derives and compares surface rainfall estimates based on KMUX radar observations at both 0.5° and 0° scan elevation angles. To estimate rainfall using radar data, the study derives polarimetric radar rainfall relations based on local raindrop size distribution (DSD) data, which are then applied to both 0.5° and 0° scan elevation angles. The results of different radar scan elevation angles are investigated at varying distances from the radar to quantify the added value of the 0° scan for surface rainfall mapping.

3.2 Study Domain and Selected Precipitation Events

The study domain is centered around the KMUX radar in the San Francisco Bay Area, CA, which supports one of the most prosperous economies in the United States. Due to the unique geography and their exposure to atmospheric rivers, there are many flood-prone regions in this area [9] [10] [11]. Hence, accurate rainfall mapping is critical to balancing the competing needs of water supply and flood mitigation [8].

In this article, we select two typical precipitation events, namely, 27-28 January 2021 and 24-25 October 2021, to quantify the hydrometeorological impacts of lowering the minimum scan elevation angle of KMUX radar. The two events occurred after November 2018, when the 0° scan was implemented by the KMUX radar. Both events are associated with atmospheric rivers (AR), thus can represent the local heavy precipitation characteristics to a large extent [9] [12].

Polarimetric radar observations from the KMUX radar at both 0.5° (the minimum scan angle before November 2018) and 0° scan elevation angles during these two precipitation events are investigated and are used to derive rainfall estimates. The rainfall estimates are comparatively evaluated using surface rain gauge measurements at different distances from the KMUX radar (see Fig. 3.2) to highlight the improvement brought by the lower scan elevation angle at longer distances.

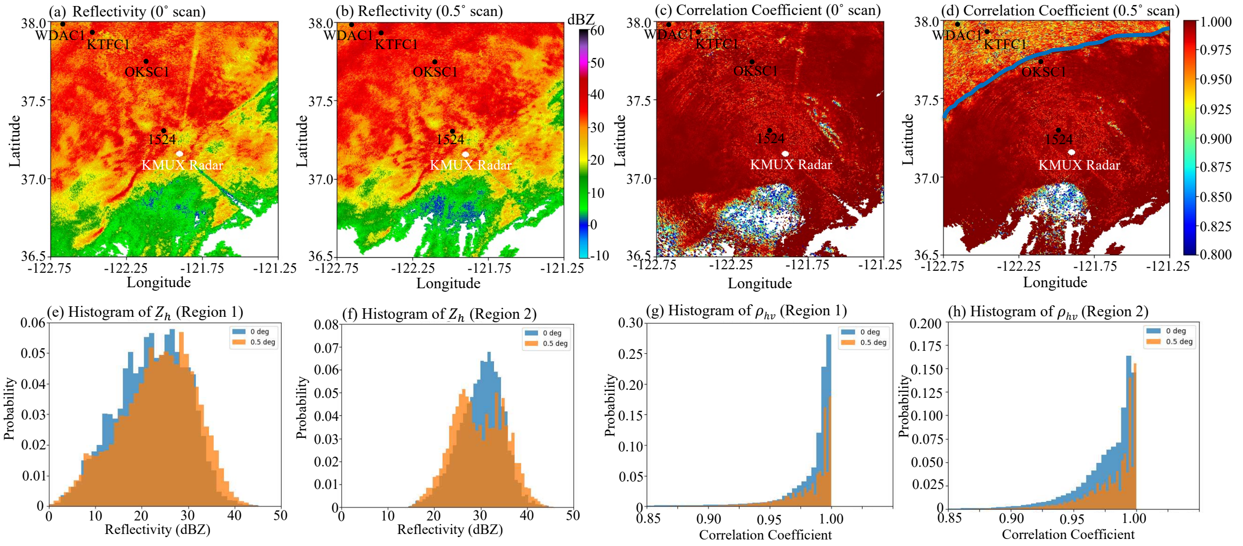
(a) Reflectivity (0° scan)
(b) Reflectivity (0.5° scan)
(c) Correlation Coefficient (0° scan)
(d) Correlation Coefficient (0.5° scan)
(e) Histogram of Z_h (Region 1)
(f) Histogram of Z_h (Region 2)
(g) Histogram of ρ_{hv} (Region 1)
(h) Histogram of ρ_{hv} (Region 2)

Figure 3.2: KMUX radar observations at 0215UTC, October 25, 2021: reflectivity Z_h of (a) 0° scan and (b) 0.5° scan; correlation coefficient ρ_{hv} of (c) 0° scan and (d) 0.5° scan. Histogram of KMUX radar observations over two subdomains (Region 1 is within 70 km from the KMUX radar, and Region 2 is 70-100 km from the KMUX radar): reflectivity Z_h in (e) Region 1 and (f) Region 2; correlation coefficient ρ_{hv} in (g) Region 1 and (h) Region 2.

3.3 Methodology and Results

3.3.1 Methodology

In this study, the polarimetric radar observations, including reflectivity Z_h, differential reflectivity Z_{dr}, specific differential phase K_{dp}, and correlation coefficient ρ_{hv}, from the two scan elevation angles are investigated and are used to produce surface rainfall estimates. For rainfall estimation, this article derives the locally optimized radar rainfall relations based on the raindrop size distribution (DSD) measurements collected during November 2018 through May 2019, from a disdrometer deployed in Santa Clara, CA. In particular, using the T-matrix scattering approach, the DSD data are utilized to simulate the polarimetric radar variables at S-band (i.e., WSR-88D operating frequency). The corresponding rainfall rates R are also calculated using the DSD data. Then, 70% of the data are used to obtain the DSD-fitted radar rainfall relations, i.e., $R(Z_h)$, $R(K_{dp})$, $R(Z_h, Z_{dr})$, and $R(K_{dp}, Z_{dr})$, based on nonlinear least squares regression.

In order to evaluate the performance of the locally fitted radar rainfall relations, they are compared with a set of operational algorithms and a few conventional relations used in different geographica regions (e.g., a convective rainfall estimation algorithm in Oklahoma and north Texas; [13]), using the 30% independent testing DSD dataset. The normalized standard error (NSE, %), the Pearson correlation coefficient ($CORR$, unitless), and root mean square error ($RMSE$, mm/hr) for these radar rainfall relations are computed.

$$CORR = \frac{\sum[(R_R - <R_R>)(R - <R>)]}{\sqrt{\sum(R_R - <R_R>)^2}\sqrt{\sum(R - <R>)^2}} \tag{3.1a}$$

$$NSE = \frac{<|R_R - R|>}{<R>} \tag{3.1b}$$

$$RMSE = \sqrt{<(R_R - R)^2>} \tag{3.1c}$$

where R_R and R represent rainfall estimates from radar parameters and directly calculated from DSD data, respectively. The angle brackets stand for sample average.

The DSD-fitted radar rainfall relations are then applied to the KMUX radar observations at 0° and 0.5° scan elevation angles to produce surface rainfall estimates during the two precipitation events. To compare and evaluate the radar-derived rainfall estimates, a similar set of metrics to Eq. 3.1 is used. However, instead of using rain rates directly computed from DSD as references, rainfall measurements from four rain gauges (black dots in Fig. 3.2) are used as references in practical applications.

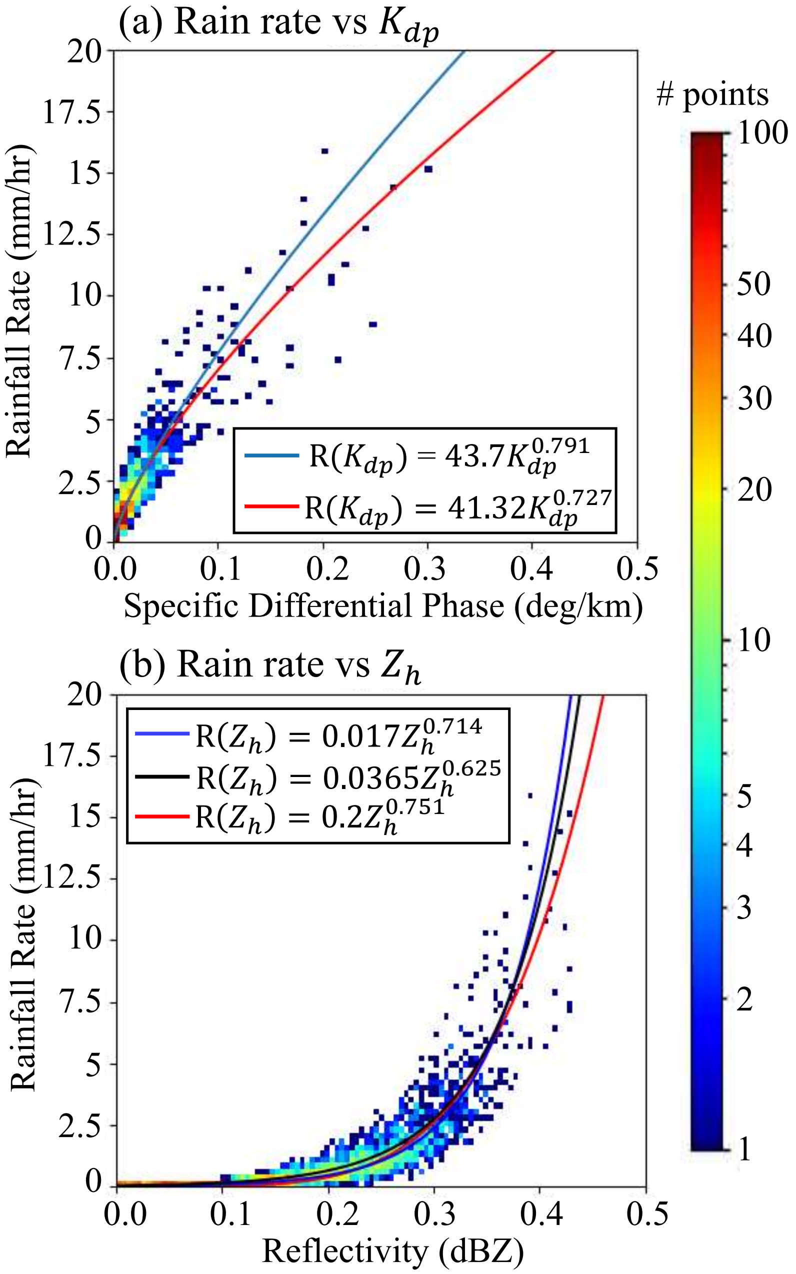

Figure 3.3: (a) Scattergram of rainfall rate R versus K_{dp}. The red line indicates the locally fitted radar rainfall relation, whereas the blue line shows a selected $R(K_{dp})$ relation used for (convective) rainfall estimation in Oklahoma and north Texas. (b) Scattergram of rainfall rate R versus Z_h. The red line indicates the locally fitted relation, whereas the blue and black curves show the WSR-88D $Z - R$ relation for convective rain and Marshall-Palmer $Z - R$ relation for stratiform rain, respectively. Note that the color-coded rain rates and radar variables are all computed from disdrometer data in Northern California.

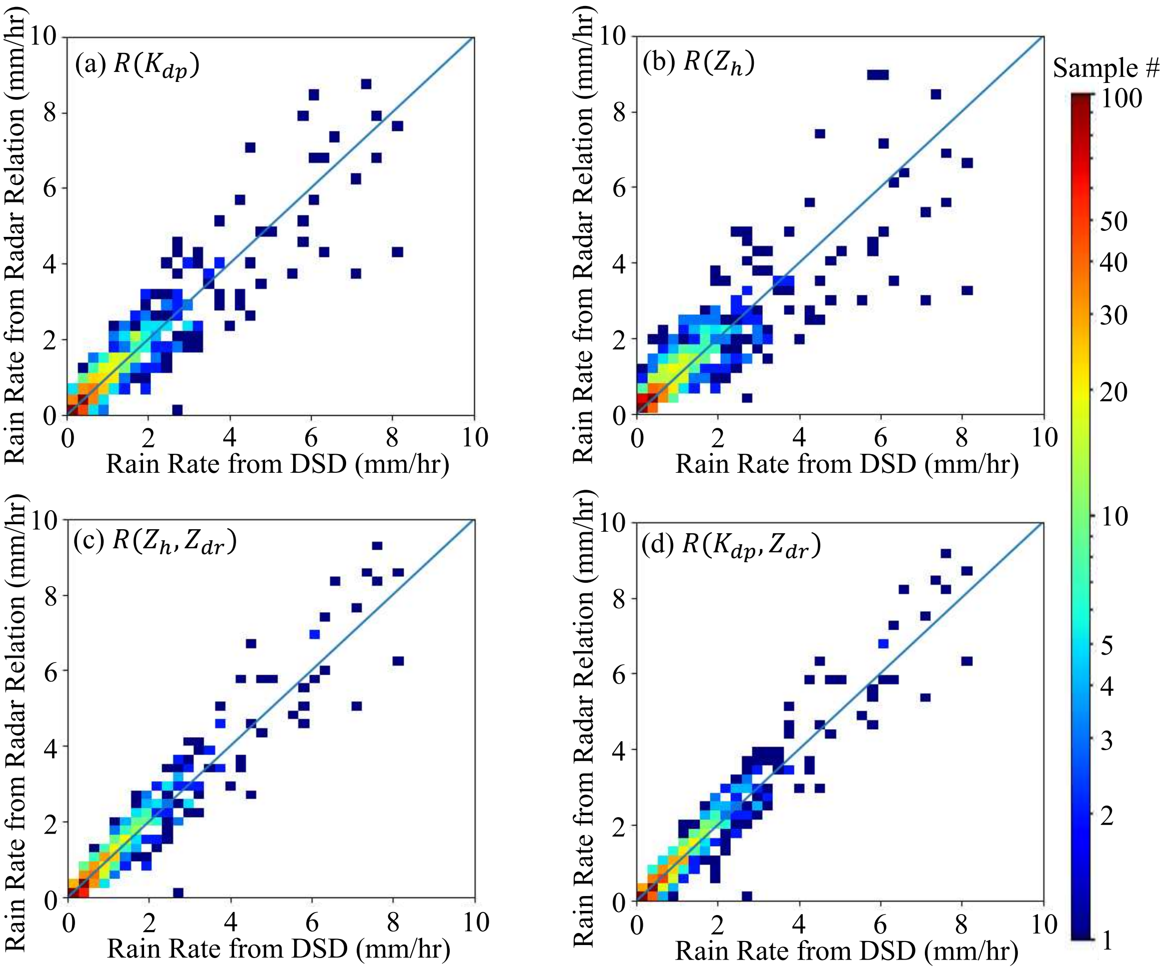

Figure 3.4: Scattergram of rainfall rates estimated from (a) $R(K_{dp})$, (b) $R(Z_h)$, (c) $R(Z_h, Z_{dr})$, and (d) $R(K_{dp}, Z_{dr})$ versus rainfall rates directly computed using the independent testing DSD dataset.

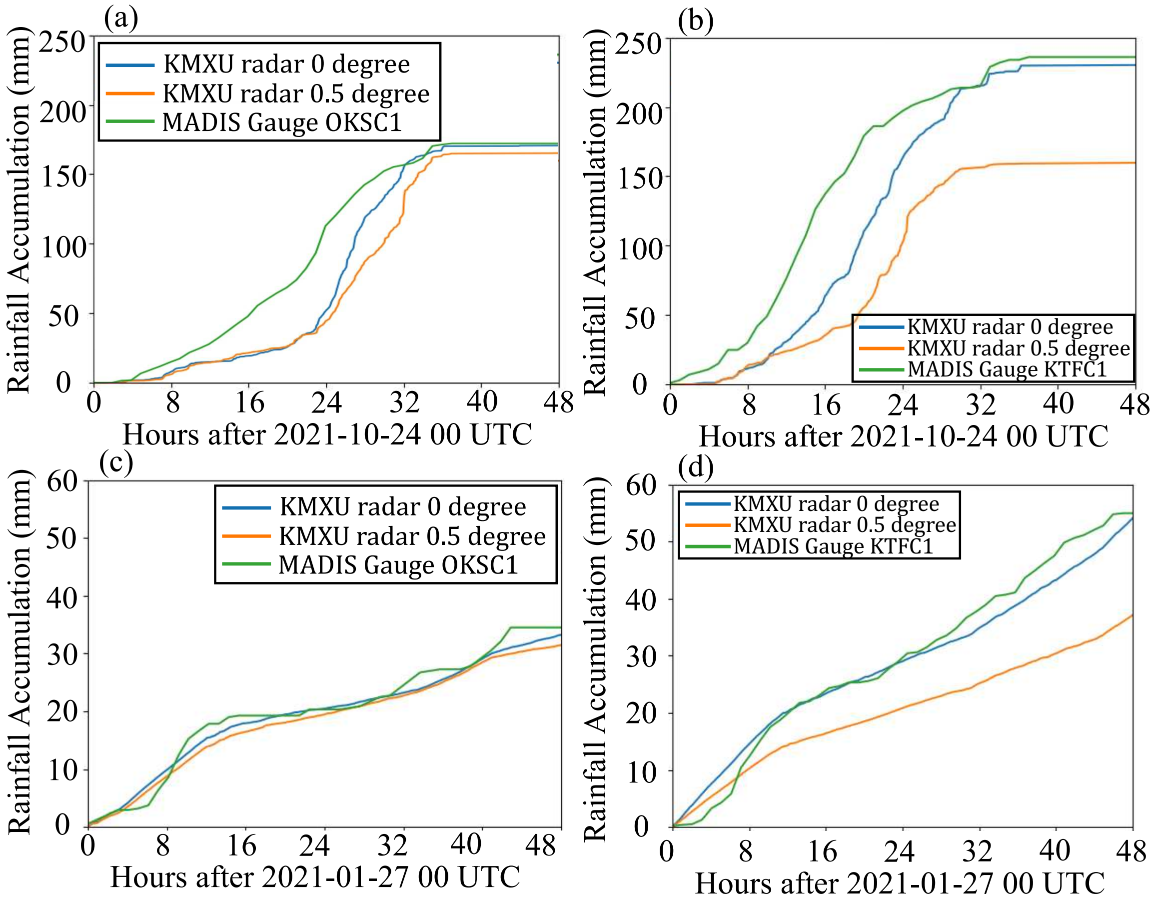

Figure 3.5: Comparison of rainfall accumulations of two precipitation events: (a-b) from 0000UTC, October 24 to 0000UTC, October 26, 2021; (c-d) from 0000UTC, January 27 to 0000UTC, January 29, 2021. Both radar rainfall estimates derived from 0° and 0.5° elevation angles of KMUX radar, and the corresponding surface rainfall measurements at two gauge locations are illustrated.

3.3.2 Results

As an example, Fig. 3.2 presents the KMUX radar reflectivity and correlation coefficient from two elevation angles at 0215UTC, 25 October 2021. Comparison between 0° and 0.5° data indicates that there is a regional boundary around 70 kilometers (as outlined by the blue line in Fig. 3.2(d)) from the radar station, beyond which the correlation coefficient from the 0.5° scan is noticebly lower, suggesting that the 0.5° beam was observing within or above the melting layer. To gain more insights into the observation difference, we further separate the radar coverage domain into two subregions depending on the distance from the radar site. Region 1 is within 70 km from the radar, whereas Region 2 is 70-100 km from the radar. Figs. 3.2 (e)-(h) show the histograms of KMUX radar reflectivity and correlation coefficient for the two subregions, which highlight that the reflectivity is narrower and that there is a higher concentration of cc closer to 1 for the 0.0 deg scan, especially over Region 2. In addition, both the narrower reflectivity and the more prevalent correlation coefficient close to 1 is consistent with the 0.0 deg scan sampling in the liquid region of the storm compared to the 0.5 deg scan. The higher liquid concentration of 0° scan observations indicates positive impacts of lowering WSR-88D scan elevation angle on surface rainfall estimation, which will be demonstrated in this section.

Based on the DSD data, the locally-fitted radar rainfall relations are obtained as follows:

$$R(K_{dp}) = 41.32K_{dp}^{0.727} \tag{3.2a}$$

$$R(Z_h) = 0.02Z_h^{0.751} \tag{3.2b}$$

$$R(Z_h, Z_{dr}) = 0.0118Z_h^{0.881}Z_{drl}^{-4.103} \tag{3.2c}$$

$$R(K_{dp}, Z_{dr}) = 103.9K_{dp}^{0.891}Z_{drl}^{-2.235} \tag{3.2d}$$

Here, Z_h ($\mathrm{mm^6m^{-3}}$) is the reflectivity factor at horizontal polarization and $Z_{drl} = 10^{Z_{dr}/10}$ is differential reflectivity in linear scale.

For comparison, four commonly used radar rainfall relations are included, including a $R(K_{dp})$ relation used for convective rainfall estimation ([14] [15]), Marshall-Palmer $Z - R$ relation for stratiform rainfall estimation (Eq. 3.4), and operational algorithms used by WSR-88D (Eqs. 3.5 and 3.6).

$$R(K_{dp}) = 43.7K_{dp}^{0.791} \tag{3.3}$$

$$R(Z_h) = 0.0365Z_h^{0.625} \tag{3.4}$$

$$R(Z_h) = 0.017Z_h^{0.714} \tag{3.5}$$

$$R(Z_h, Z_{dr}) = 0.0142Z_h^{0.770}Z_{drl}^{-1.67} \tag{3.6}$$

Figure 3.3 shows the scattergrams of the DSD-derived rainfall rates versus K_{dp} and reflectivity (Z_h; in dBZ). The locally fitted relation (in red) and other commonly used radar rainfall relations are also indicated in Fig. 3.3. Overall, it can be seen that the locally fitted relations demonstrate a better performance in fitting the DSD-based observations, especially at high rain rate values ($>$ 5 mm/hr) [14] [15]. Nevertheless, since this study domain is often characterized by shallow stratiform rain with low rain rates, we expect that the difference between different radar rainfall relations may not be significant in practical applications.

To further quantify the parameterization errors of various radar rainfall relations, this article applies Eqs. 3.2-3.6 to the simulated radar observables based on the independent testing DSD data. For illustration purposes, Fig. 3.4 shows the rainfall estimates derived using Eq. 3.2 and those directly computed from the independent testing DSD data. As detailed in Section 3.3.1, $CORR$, NSE, and $RMSE$ scores are computed for all the applied radar rainfall relations, and the results are summarized in Tbl. 3.1. According to these statistics, the locally fitted radar rainfall relations are superior to other conventional radar rainfall relations. In addition, the $R(K_{dp}, Z_{dr})$ relation shows the best performance, with the highest $CORR$ and the lowest $RMSE$ values. Therefore, this new radar rainfall relation (Eq. 3.2d) is applied to the KMUX radar observations from 0° and 0.5° scans to quantify the benefit brought by lowering the scan elevation angle for surface rainfall estimation.

Table 3.1: Evaluation results of the parameterization errors of different radar rainfall relations based on the independent DSD Data: NSE, $CORR$, and $RMSE$.

Radar rainfall relation	NSE	$CORR$	$RMSE$
$R\left(K_{dp}\right)=41.32K_{dp}^{0.727}$	25.17%	0.969	0.383
$R\left(Z_h\right)=0.02Z_h^{0.751}$	26.28%	0.936	1.217
$R\left(Z_h,Z_{dr}\right)=0.0118Z_h^{0.881}Z_{drl}^{-4.103}$	33.09%	0.965	0.436
$R\left(K_{dp},Z_{dr}\right)=103.9K_{dp}^{0.891}Z_{drl}^{-2.235}$	18.88%	0.985	0.279
$R\left(K_{dp}\right)=43.7K_{dp}^{0.791}$	26.38%	0.967	0.465
$R\left(Z_h\right)=0.017Z_h^{0.714}$	27.35%	0.738	2.258
$R\left(Z_h\right)=0.0365Z_h^{0.625}$	26.72%	0.809	1.574
$R\left(Z_h,Z_{dr}\right)=0.0142Z_h^{0.770}Z_{drl}^{-1.67}$	36.14%	0.820	1.127

After obtaining rainfall estimates using Eq. 3.2d based on the 0° and 0.5° scan data during the two selected precipitation events, rainfall measurements from four rain gauges are used as references to quantify the radar QPE performance. Two rain gauges are located in each subregion (OKSC1 and 1524 in Region 1; WDAC1 and KTFC1 in Region 2); selection of these gauges can be used to help us quantify both the QPE impact of lowering the KMUX radar scan elevation angle and the assessment of QPE performance at different ranges from the KMUX radar.

Fig. 3.5 illustrates the accumulated rainfall at two validation gauge locations from 0000UTC, October 24 to 0000UTC, October 25, 2021, including rainfall estimates from the KMUX radar observations at 0° and 0.5° scan elevation angles, as well as rainfall measurements from the gauges. Here, we show the estimates from one gauge in Region 1 (OKSC1) and one in Region 2 (KTFC1) to highlight the relative performance of rainfall estimates at different distances from the radar. Also, the performance for the other gauges which are not shown in Fig. 3.5 are similar. In general, radar rainfall estimates from the 0° scan have the most pronounced that those from the 0.5° scan, and the difference is large at long distances (see Fig. 3.5b), consistent with the 0.0 deg beam remaining in the liquid region below the melting layer over a larger portion of the domain compared to the 0.5 deg beam (see Fig. 3.1). For a more comprehensive verification, Fig. 3.6 shows the quantitative evaluation results of hourly QPE derived from 0° and 0.5° scan data during the two selected precipitation events, based on rainfall measurements from all the four surface gauge stations. Fig. 3.6 clearly shows the errors in radar rainfall estimates are reduced comapred to 0.5 deg data using the 0° scan data, especially at long ranges from the radar (i.e., at gauge location WDAC1 and KTFC1 in Region 2). This is consistent with the radar data distributions in Fig. 3.2 and result comparisons in Fig. 3.5.

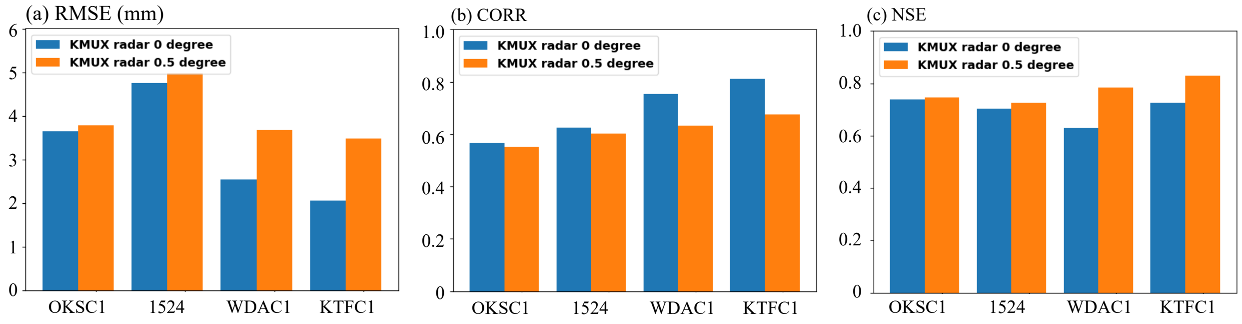

Figure 3.6: Quantitative evaluation results of radar hourly QPE derived from 0° and 0.5° scan elevation angles during the two selected precipitation events: (a) $RMSE$ (mm/hr), (b) $CORR$ (unitless), and (c) NSE (%). These results are based on rainfall measurements from four validation gauges illustrated in Fig. 3.2.

3.4 Summary and Conclusion

Using the KMUX radar deployed in the San Francisco Bay Area as an example, this study quantifies the QPE impacts of upgrading the scan strategy of WSR-88D by lowering the minimum scan elevation angle from 0.5° to 0°. New radar rainfall relations were derived using DSD observations from a local disdrometer, and the new relations are compared with several conventional radar rainfall algorithms. Then the optimal relation is applied to the KMUX radar data at 0° and 0.5° scan elevation angles during two typical AR-induced precipitation events in Northern California. Major conclusions are summarized as follows:

(1) The evaluation results show that the DSD-fitted local radar rainfall relations are superior to conventional radar relations in this study area. The intercomparison further indicates that the $R(K_{dp}, Z_{dr})$ relation performs the best with the highest $CORR$ and lowest $RMSE$ scores compared with the other three relations, i.e., $R(K_{dp})$, $R(Z_h)$, and $R(Z_h, Z_{dr})$.

(2) The radar QPE based on KMUX observations at different scan elevation angles are derived using the best locally fitted radar rainfall relation and the QPE products are evaluated using rainfall measurements from four surface rain gauges. It is found that radar QPE from 0° scan data is generally more accurate than that derived from 0.5° scan data.

(3) The relative improvement on radar QPE brought by lowering the KMUX WSR-88D radar scan elevation angle at different detection ranges is further assessed, by grouping the gauges into two categories in terms of their distance from the radar site (within 70 km versus beyond 70 km). This conditional analysis highlights that the lower scan elevation angle (i.e., 0°) data has significantly improved the radar QPE performance at long distances from the radar, where the radar beams are likely intersecting or overshooting the melting layer.

However, it should be noted that radar QPE uncertainty is determined by many physiographic factors [11]. Incorporating the vertical structure of radar observations and topographical information in complex terrain [9] to further improve radar QPE performance should be investigated in future. In addition, similar demonstration study with WSR-88D stations at other geophysical regions and different seasons should be conducted to fully understand when and where the upgraded scan strategy can help with radar-based quantitative precipitation estimation.

Chapter 4

A Machine Learning Perspective on Polarimetric Radar Rainfall Estimation

4.1 Study Domains and Datasets

This section will provide a detailed introduction to the WSR-88D dual-polarization Doppler radar observation data and the GPM GV ground-based rain gauge data used in this study, and describe the process of constructing the dataset as well as the normalization methods applied to it.

4.1.1 Data Introduction

The black circle in figure 1 shows the study area of this chapter, which is within 100 kilometers of the radar site KMLB. The black dot in the figure represents the coordinates of the KMLB radar station, which is located at 80.65° W longitude and 28.11° N latitude. The red dots represent the distribution of the 40 rain gauges within 100 kilometers of KMLB, which are provided by the South Florida Water Management District (SFL). Only radar and rain gauge data within 100 kilometers of KMLB were selected for this study, and only data from the lowest two elevation angles of the radar were used for three main reasons.

First, as the distance of the detected target increases, the beam width of the electromagnetic wave emitted by the radar widens, and the unit volume of the detection also increases. As a result, the radar echo contains more information on backscatter from different particles in the atmosphere, which leads to a decrease in data quality.

Second, the energy of the electromagnetic wave emitted by the radar gradually weakens as the propagation distance increases, which is called attenuation of the electromagnetic wave in the atmosphere. The fundamental reason for the attenuation of electromagnetic waves in the atmosphere is due to the absorption and scattering of electromagnetic wave energy by media such as water vapor. The attenuation of electromagnetic waves reduces the power of the radar echo, which leads to a decrease in data quality.

Third, the radar detection process is performed in the air, while the rain gauge observation is on the surface of the ground. The height difference between the two observation methods can cause a time error in the data between them. When the elevation angle is fixed, the height of the detection also increases with the increase of the electromagnetic wave propagation path, so the time error between the two data types also increases. Moreover, precipitation can move laterally a considerable distance between the radar observation position and the ground, or evaporate before reaching the ground, which can cause errors in subsequent rainfall estimates.

Also, there are two main reasons why this area has been chosen as the research region for this chapter.

To begin with, it is situated in the southeastern part of the continental United States and is characterized by a plain terrain with an average altitude of approximately 35 meters. The highest point in this area does not exceed 120 meters. Moreover, a plethora of equipment, such as radars and rain gauges, have been deployed in this region. Given its flat terrain and low altitude, radar observation data quality is more precise and dependable, and the impact of radar obscuration is relatively minimal.

Additionally, this area is located on the southeastern coast of the Florida Peninsula in North America and comprises predominantly subtropical climate regions. However, the southern coastal areas mostly belong to tropical climate regions. Precipitation is abundant and predominantly in liquid form, with snowfall being infrequent. As such, the rain gauge can obtain adequate and distinctive precipitation data.

4.1.1.1 Radar Data

In this chapter, the reflectivity factor and differential reflectivity in the dual-polarization radar data are used as input features for the model. The radar data used in this study were obtained from the KMLB dual-polarization Doppler radar station located in Melbourne, Florida, USA. This radar belongs to the NEXRAD radar network, which consists of 160 high-resolution Doppler weather radars operating in the S-band. It is jointly operated by the National Oceanic and Atmospheric Administration (NOAA), the Federal Aviation Administration (FAA), and the United States Air Force. Its technical name is WSR-88D (Weather Surveillance Radar, 1988, Doppler). The KMLB radar station operates in the S-band with an electromagnetic frequency of 2700-3000MHz and transmits circularly polarized electromagnetic waves with a wavelength of 10.0-11.1 centimeters. The dual-polarization mode is dual-transmit, dual-receive, and the basic operational modes are the clear air mode and precipitation mode. The maximum detection range is 460 kilometers, with a azimuthal range of 0-360 degrees, a maximum antenna scan elevation of 19.5 degrees, and a minimum of 0.5 degrees, which can be lowered to 0.2 degrees or even lower in coastal areas. The spatial resolution of the radar data varies depending on the type of radar product data. Level-I data is the digital receiver recorded data, Level-II data is super-resolution data, and the radar range resolution (range gate) is 250 meters, and the azimuth resolution is 0.5 degrees, with a resolution of 0.25km × 0.5° for each range gate. The range gate resolution for Level-III data is 1.0km × 1.0°.

The reason why weather radar is used to observe precipitation systems is that when the electromagnetic waves emitted by the radar pass through precipitation particles in the atmosphere, some of the waves will continue to propagate in the forward direction, while others will be scattered back in the opposite direction due to the obstruction of precipitation particles, and the ability of different types of precipitation particles to scatter back the electromagnetic waves is also different. The difference in the strength of the backscatter produced by different precipitation particles is called the radar echo intensity. The echo intensity of precipitation particles per unit volume is usually measured using reflectivity and reflectivity factor.

The reflectivity observed by the radar represents the total backscattering cross section of precipitation particles per unit volume, which reflects the strength of the backscattering produced by the particles. This strength is determined by the physical properties of precipitation particles per unit volume. Generally, the backscattering cross section of precipitation particles increases with the increase of particle size. Therefore, the larger the size or the greater the number of precipitation particles per unit volume, the stronger the backscattering produced and the higher the reflectivity [16]. However, the backscattering cross section produced by precipitation particles not only depends on the particles themselves but also depends on the wavelength of the electromagnetic waves emitted by the weather radar. It is independent of other parameters of the radar. Therefore, the reflectivity observed by weather radar at the same wavelength can be compared to determine the strength of backscattering of different target particles.

Due to the drawback that radar reflectivity measured at different wavelengths cannot be directly compared, the reflectivity factor (Z_h) is more commonly used in the field of meteorological radar to measure the strength of backscattering of precipitation particles [17] [18]. It is defined as the sum of the sixth power of the diameter of precipitation particles that meet the Rayleigh scattering conditions per unit volume, with units of mm6/m3. Because the reflectivity factor only depends on the physical characteristics of the number and size of precipitation particles per unit volume and is independent of the working parameters and detection distance of the radar, the reflectivity factors observed by radars at different wavelengths can also be compared with each other. However, since the range of variation of the reflectivity factor Z_h is very large and can even span multiple orders of magnitude, for convenience, dBZ is commonly used to indicate the magnitude of the reflectivity factor, that is:

dBZ=10×lgZ_h/Z_0 (Z_0=1 mm^6/m^3).

Generally speaking, as the particle size increases, the backward scattering ability of precipitation particles also increases rapidly, but the process is not monotonic, but rather has fluctuations. When the precipitation particles are not too large, the backward scattering ability of water droplets is greater than that of ice particles of the same size, while when the precipitation particles are very large, the backward scattering ability of water droplets is smaller than that of ice particles of the same size [19].

A single-polarization weather radar emits horizontally polarized electromagnetic waves to observe precipitation particles through the radar echo in the horizontal direction. However, a dual-polarization radar can emit both horizontally and vertically polarized electromagnetic waves. As shown in the figure, the differential reflectivity can be calculated by the ratio of the power of the two directions of echoes. The differential reflectivity was first introduced for precipitation measurement by Seliga and Bringi in 1976.

The differential reflectivity is the logarithm of the ratio of the radar reflectivity factors for horizontal and vertical polarization waves of the same volume. The formula for calculating it is as follows:

$Z_{dr} = 10\ \mathrm{lg}(Z_h/Z_v)$

Where Z_h represents the horizontal reflectivity factor and Z_v represents the vertical reflectivity factor. The physical meaning of the differential reflectivity can be clearly obtained by analyzing this formula, and the type of precipitation particles can be distinguished by the value of Z_{dr}.

For larger liquid raindrops, due to factors such as air resistance, they typically assume an oblate spheroid shape as they fall, with the sum of the horizontal diameters of particles per unit volume greater than the sum of the vertical dimensions [20]. The reflectivity factor is the sum of the sixth power of the particle diameter per unit volume. Calculation shows that the horizontal reflectivity factor is greater than the vertical reflectivity factor, resulting in a positive Z_{dr}. In contrast, for ice crystals and other types of hydrometeors, the vertical dimensions are larger than the horizontal dimensions, resulting in a greater vertical reflectivity factor than the horizontal reflectivity factor, or a negative Z_{dr} value. For small raindrops, the horizontal and vertical reflectivity factors are approximately equal, and the Z_{dr} value is around 0 dB. Therefore, the introduction of differential reflectivity in quantitative precipitation estimation can reduce the uncertainty caused by differences in the raindrop spectrum when using only reflectivity factor estimation, thereby improving the accuracy of precipitation estimation.

The Specific Differential Phase (K_{dp}) is a radar variable that characterizes the differential attenuation between the horizontal and vertical polarizations of radar waves as they travel through precipitation. By measuring the rate of change of the differential phase shift of the radar wave as it propagates through the precipitation medium, K_{dp} can provide information on the size and shape of hydrometeors and the intensity of precipitation [21].

K_{dp} is particularly useful in detecting and quantifying heavy rainfall rates, because it is less sensitive to measurement noise and artifacts than other radar variables such as reflectivity and Doppler velocity. In addition, K_{dp} can help to correct for the attenuation of radar signals by rain, which can cause significant errors in the estimation of precipitation. The introduction of K_{dp} in radar measurements has led to significant improvements in precipitation estimation, especially in regions with high rainfall rates or complex precipitation structures [22].

This study used 2016-2021 KMLB dual-polarization radar data, which is in a polar coordinate system, with Level-II data type and a resolution of (250m x 0.5°). The radar variables used in this study were reflectivity factor, differential reflectivity, and Specific Differential Phase. Only radar scan data from the lowest two elevation angles were used when creating the dataset.

4.1.1.2 Rain Gauge Data

The rain gauge is the most accurate instrument for measuring precipitation intensity compared to other measurement instruments. Therefore, it has always been the primary source of quantitative precipitation estimation and is still one of the most widely used instruments for measuring accumulated precipitation. The rain gauge measures the amount of rainfall at a given location in a unit of time, so the main goal of the rain gauge is to obtain representative rainfall measurements for the measured area. Therefore, in this study, the rain gauge is used as the true value for quantitative precipitation estimation. The rain gauge data used in this paper is from the SFL rain gauge network in Florida, USA, which is ground validation data for NASA's GPM project. The rain gauge used in this network is a tipping bucket rain gauge, with a time resolution of 1 minute and a precipitation intensity resolution of 0.01 inch/min, or 0.254mm/min. This study uses valid data from rain gauge stations within 100 kilometers of the main radar KMLB in the SFL rain gauge network from 2016 to 2021 (See Fig. 4.1).

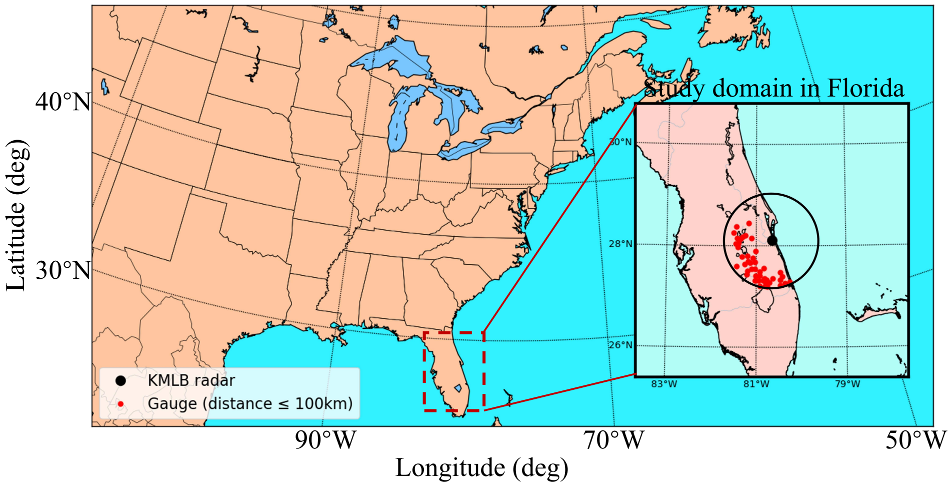

Figure 4.1: Location of the KMLB WSR-88D (red dot) in Melbourne, FL, USA and its 100-km coverage domain. The black dots indicate rain gauges within 100 km from the KMLB radar. The gauges are manged by the South Florida Water Management District (SFL).

4.1.2 Data Preprocessing

4.1.2.1 Dataset Construction

In deep learning, the training process of a supervised learning model involves providing the model with an input and a corresponding expected output, and then continuously adjusting the model's weights and biases using gradient descent propagation algorithms to gradually make the model's output approach the correct result. Each sample in the training data of a supervised learning algorithm consists of an input object and an expected output value. The supervised learning algorithm analyzes this training data and establishes a mapping relationship from input to expected output, which can be used to infer new instances. Therefore, the preparation of the dataset is crucial, and this section will introduce the process of dataset preparation.

As explained in the previous section on radar and rain gauge data and their data formats, radar data is a two-dimensional matrix with spatial location information, while rain gauge data is a scalar fixed in space with latitude and longitude. The two types of data have different spatial and temporal resolutions, so it is necessary to match the two types of data in space and time.

Regarding spatial matching, a sample in this paper's dataset should consist of a rain gauge and two layers of radar data within a certain range of coordinates above it. In this study, radar observation data within a certain area above the rain gauge is used as input features because precipitation at a certain point is not only related to information about precipitation particles above it, but also related to precipitation information in the surrounding areas due to factors such as wind and the lateral movement of precipitation particles. Therefore, this study hopes to construct a fixed range of radar data and use it as input to allow the deep learning precipitation estimation model to learn the mapping relationship between the precipitation intensity observed by the rain gauge and the precipitation data in the area surrounding the rain gauge, and output the precipitation estimation value. The size of the fixed area should not be too large because the deep learning model is a data-driven algorithm and is extremely sensitive to data quality. At the same time, data further away from the target area will have lower correlation with precipitation in the target area. Therefore, using too large of a fixed area will introduce noise and low correlation precipitation spatial information in the quantitative precipitation estimation process, leading to errors in the deep learning model's estimation. Therefore, the input of this study's precipitation estimation model is a (6x9x9) matrix consisting of two layers of KMLB radar data in a spatial resolution of (250m x 0.5°) rectangular window, and each layer of radar data has three variables: reflectivity factor, differential reflectivity factor, and specific differential phase.

Regarding time matching, a sample in this paper's dataset consists of a radar file and multiple data points of given precipitation intensity from the corresponding time range of the rain gauge. In this study, since the time span of KMLB radar files is not the same and the time resolution of precipitation intensity data (mm/h) from the rain gauge is 1 minute, the dataset is constructed based on the time span of KMLB radar data.

After the dataset is created, it will be randomly divided into training set, validation set, and test set according to a ratio of 7:1:2. The training set and validation set will be used during the training process. The training set will be continuously adjusted by the loss function and backpropagation algorithm to adjust the deep learning model's parameters to fit the data distribution. The validation set will be used to adjust the hyperparameters of the deep learning model during the training process. The test set will be used to evaluate the effectiveness of the model.

4.1.2.2 Dataset normalization

Deep learning algorithms learn the high-dimensional mapping relationship of input data based on the statistical distribution characteristics of the input data and the distribution patterns of the target values. This enables them to perform nonlinear fitting of the target values for the task. Therefore, the effectiveness of the algorithm depends on how well the input sample data represents its underlying patterns. Deep learning models sometimes incorporate multiple task-related features, which may have different scales. It is important that different types of features have a comparable impact on the distances between different samples. If the features are not in the same scale, features with higher values and larger ranges will dominate the neural network's influence, which can hinder the network's fitting capability. Additionally, inputting data with a unified scale facilitates faster network convergence.

In this study, the data used consists of radar reflectivity factor, differential reflectivity, and Specific Differential Phases. Due to the differing scales of these three features, it is necessary to scale the data to a unified standard range. Common normalization methods in machine learning include Max-Min normalization and standard deviation normalization.

The former, also known as Min-Max normalization, is the concept of linearly scaling the data. By finding the maximum and minimum values in the variable, the data is linearly transformed using the formula to the range of [0, 1]:

X_{norm}=(X-X_{min})/(X_{max}-X_{min})

In the equation, X represents the original data, X_{norm} represents the normalized data, and X_{max} and X_{min} represent the maximum and minimum values of the corresponding variable. In this method, if there are outliers at the boundaries, they should be removed in advance. If the distribution itself is too discrete, it may not effectively reflect the data characteristics. However, this method only requires knowledge of the maximum and minimum values of the samples, making it computationally simple.

Standard deviation normalization, also known as z-score normalization or centering and scaling, utilizes the statistical characteristics of the data, namely the mean and standard deviation. It can linearly scale the data without altering its distribution, resulting in data with a mean of 0 and a standard deviation of 1, following a standard normal distribution.

In this study, the normalization method used is the Min-Max normalization, which is specific to radar data. The normalization process can create a barrier between regions with and without data, which helps the network better learn boundary information.

4.2 Radar Quantitative Precipitation Estimation Model

4.2.1 Convolutional Neural Network

The idea of depthwise separable convolution is derived from the concept of regular convolution, which has a long history and occupies an important position in the field of image processing. Its inspiration comes from the hierarchical processing of visual information in the visual cortex of the biological vision system, where recognition is a process of grading from low-dimensional pixel features to high-dimensional abstract features and from local edges to global contours.

In 1988, Yann LeCun et al. first proposed the LeNet-5, a handwritten digit recognition neural network based on backpropagation, which became the foundation of convolutional neural network models [23]. In 2012, the performance of the convolutional neural network AlexNet in the ImageNet competition demonstrated its tremendous potential in computer vision [24]. Over the next decade, with the emergence of many large public datasets and the development of hardware that improved computational capabilities, convolutional neural networks continued to develop in computer vision tasks such as image segmentation and object detection. During this time, deeper and more effective convolutional neural networks such as VGG, GoogleNet, ResNet, DenseNet, HRNet, and EfficientNet were proposed [25] [26], further advancing the application of deep learning in computer vision.

With the advancement and deepening of these works [27] [28] [29], convolutional neural networks have developed rapidly, and at the same time, these works have also demonstrated the superiority of convolution in image processing tasks. The advantages of convolution are mainly reflected in: (1) Local perception. Considering a pixel in an image, its adjacent areas have higher correlation, so the convolution operation captures the local features of the surrounding area of the point to be calculated, just as a certain neuron in biology only responds to a specific stimulated area. The region that a network focuses on is measured using the concept of effective receptive field [55]. The receptive field represents the range of the input image that a point can see, and this area will correspondingly expand with the deepening of the network layers, thereby achieving the concept of from local to global. (2) Weight sharing. Based on the position-independent statistical features of the image, a convolutional kernel can be used to capture a certain feature of the entire image, such as a horizontal edge, through a sliding window. Multiple convolutional kernels can capture multiple different feature information. Because parameter sharing is the effect of the parameters of a convolutional kernel on the entire image, features appearing at different positions can be captured. (3) Sparse connection. Its graph-to-graph mapping method is not as dense as the fully connected operation, which means that the connection between adjacent neurons is not dense. Fewer parameters are suitable for extracting redundant information from images, and less dense connections leave more room for improvement of lightweight neural units in the network. The convolution operation can be regarded as an image feature extractor, and using different convolutional kernels will highlight different features of the image. Compared with manually constructing convolutional kernels in image processing, convolution operations in deep learning avoid manually designing features, and the stacking of multiple convolution layers completes the extraction of low-level features to complex abstract features, making convolutional neural networks more suitable for image processing.

Convolution can perform feature extraction, where each layer's output feature map becomes the input for the next layer. Through training, a feature extractor suitable for a specific task can be obtained. When images are used as inputs to neural networks, they are often represented as three-dimensional data. In addition to representing the image dimensions (h, w), there is also a dimension representing the number of input channels, denoted as d_{in}. For grayscale images, the input channel is 1, while for color images, the input channel is 3, representing RGB. In traditional multi-channel convolution, the number of channels in the convolutional kernel is usually the same as the input image (or feature), and the number of output channels depends on the number of convolutional kernels, as shown in Figure 4.2. Therefore, for an input with N channels and an output with M channels, a single convolution operation with a kernel size of f × f requires N × M × f × f parameters.

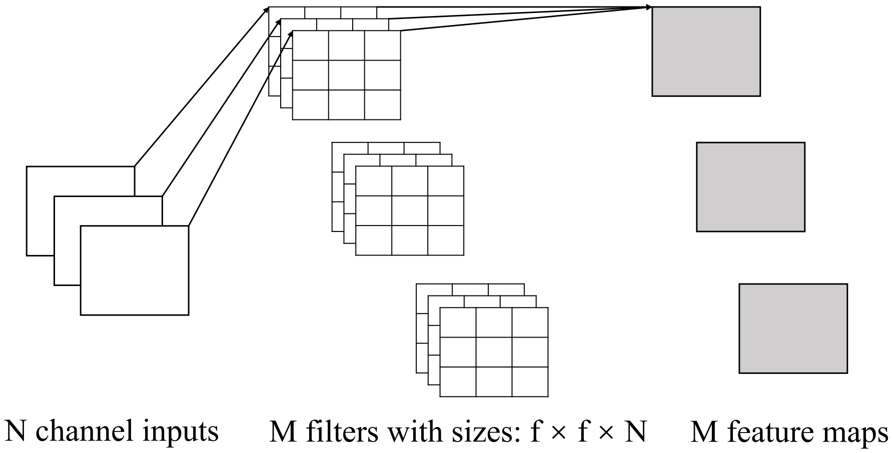

Figure 4.2: Schematic diagram of conventional multi-channel convolution.

4.2.2 Depthwise Separable Convolution

The design purpose of Depthwise Separable Convolution is to reduce the model size without sacrificing model accuracy [30] [31]. The core operation for model size reduction is based on group convolution. Group convolution refers to dividing N-channel feature maps into n groups (N can be divided by n), performing convolution on each group separately, and then stacking the resulting n groups of output feature maps along the channel dimension. The specific operation of Depthwise Separable Convolution, as shown in Figure 4.3, replaces traditional convolution with depthwise convolution and pointwise convolution. The depthwise convolution (also known as channel-wise convolution) is a grouped convolution with the number of groups equal to the number of channels.

In depthwise separable convolution, the input with N channels is divided into N groups, and each group is convolved separately to obtain intermediate feature maps with N channels. Then, M sets of 1 × 1 × N convolution kernels are used for pointwise convolution to obtain M sets of output feature maps. Depthwise separable convolution provides a feature extraction method that considers both spatial and channel interactions.

Compared with traditional convolution, depthwise separable convolution reduces the number of parameters, which allows for more flexible design of deeper and larger network architectures with the same computational resources. However, if the model size is already small, replacing it with depthwise separable convolution may result in a decrease in accuracy. Therefore, depthwise separable convolution is a trade-off between sacrificing unit performance for overall model performance.

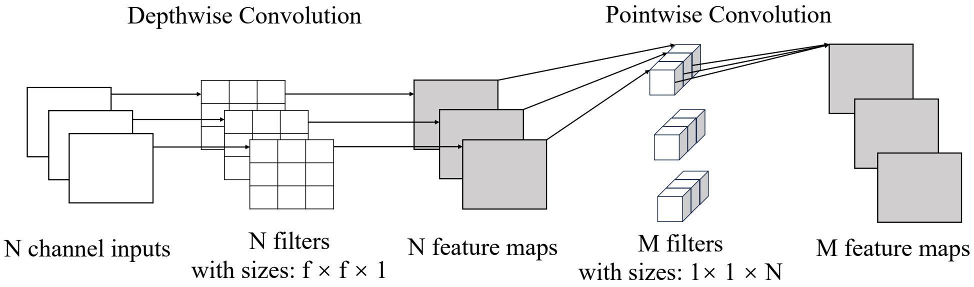

Figure 4.3: Schematic diagram of depthwise separable convolution.

4.2.3 ShuffleNet Based on Depthwise Separable Convolution

ShuffleNet V1 is an efficient convolutional neural network for mobile devices proposed by Face++ in 2017 [32]. The network was developed based on the observation that when hardware resources are limited, the number of channels is often changed through 1×1 convolutional operations. However, a large number of 1×1 convolutional operations can lead to a significant decrease in model performance. To avoid using a large number of 1×1 convolutions, ShuffleNet is based on depthwise convolution and proposes the Pointwise Group Convolution (1×1 group convolution) [33]. Additionally, it uses Channel Shuffle to alleviate the performance degradation caused by group convolution. Channel Shuffle is the core operation proposed in V1, which addresses the issue that during group convolution, the features of each group are processed separately without communication between different groups. To address this, the position information of channels is shuffled at the group level to obtain interaction between different groups, as shown in Figure 4.4.

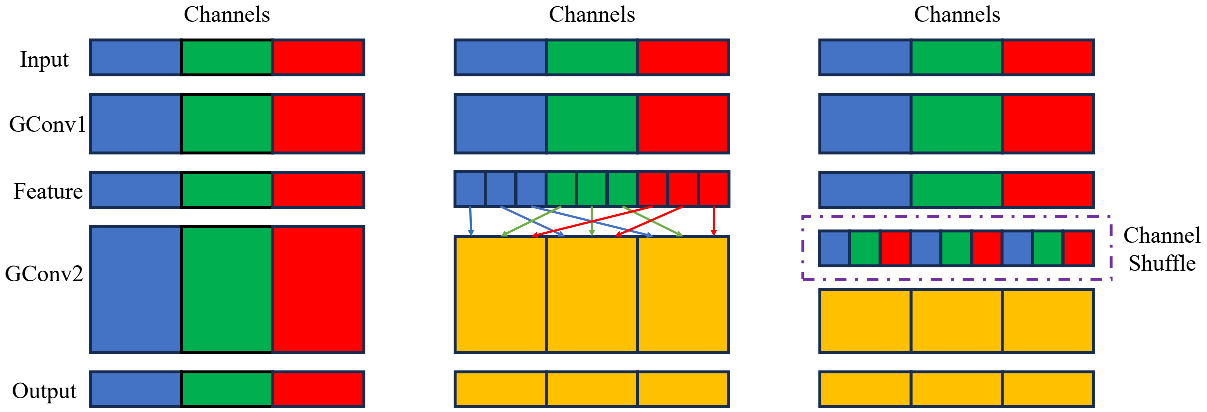

Figure 4.4: Schematic diagram of Channel Shuffle.

4.2.4 Image Processing Model Vision Transformer

The attention mechanism refers to the process of finding local information that needs to be paid more attention to in global information, in order to better process information. Its inspiration comes from the working mechanism of human vision: first scan globally, and then process the key areas that need more attention in the field of view based on feedback signals from the brain [34] [35].

In computer vision, the attention mechanism refers to the process of finding areas that need more attention [36] [37]. In natural language processing, it is the process of finding morphemes with greater relevance.

In 2017, Google's translation team published the work "Attention is All Your Need" at NIPS, marking a revolutionary advance in the application of attention mechanisms in natural language processing [38] [39]. As the title suggests, the Transformer model proposed in this work abandoned traditional CNN and RNN structures and used only self-attention mechanisms, achieving state-of-the-art performance. Self-attention mechanism calculates the correlation between input word vectors and themselves, allowing each word to capture the global relevance regardless of distance, breaking the inherent limitations of CNN, which can only focus on local information, and the inherent drawbacks of RNN's information dependence on long distance. At the same time, the self-attention mechanism obtains the correlation with the global context through matrix multiplication in one step, which is highly parallel compared to the iterative process of RNN, greatly improving the efficiency of model training. This led to a series of excellent Transformer-based improvements. In 2019, Google proposed the BERT (Bidirectional Encoder Representations from Transformers) model [40]. Based on Transformer, this work constructed a powerful pre-training model for natural language processing through self-supervised learning. The use of Transformer as a feature extractor and the development of the GPT series of unidirectional language models by the OpenAI team show that the performance of natural language processing tasks can be greatly improved with a large corpus and model parameters [41].

With the rapid progress of Transformer-based self-attention mechanisms in natural language processing, the use of self-attention mechanisms in computer vision is also constantly catching up and developing rapidly. Its application mainly starts from two perspectives. One is the combination of CNN and self-attention mechanism, such as DETR in 2020, which is used for end-to-end object detection [42]. This work uses CNN to extract features and inputs the flattened features into Transformer to obtain results through object query vectors. This work abandons the concepts of anchors and non-maximum suppression post-processing in traditional object detection and achieves complete end-to-end detection. The second is attempting to completely eliminate CNN and only use Transformer to complete image processing tasks, such as iGPT from ICML20, which flattens the pixel-level feature map of the two-dimensional image into a one-dimensional vector and directly inputs it into Transformer to perform the next frame image prediction task [43]. The most representative work is ViT (Vision Transformer) proposed at ICLR2021, which divides a 256×256 image into 16×16 patches and replaces the step of extracting features by CNN with linear mapping of patches. Patch becomes an important idea when Transformer is combined with computer vision.

Subsequent research mainly focuses on the differences between natural language processing and computer vision tasks to improve the model.

The Transformer architecture abandons traditional deep learning structures and constructs representation models using only attention mechanisms for machine translation tasks, achieving remarkable results. The Transformer is an encoder-decoder structure. Since the model proposed in this article mainly uses the Transformer's encoder structure, as shown in Figure 3-6.

The appearance of Vision Transformer (ViT) marks that in the field of computer vision, it is possible to achieve ideal results in image processing tasks without relying on convolutional neural networks.

4.3 Construction of Neural Network for Quantitative Precipitation Estimation

4.3.1 Algorithm Framework for Quantitative Precipitation Estimation

The overall process of the proposed quantitative precipitation estimation algorithm in this paper mainly consists of three parts: dataset creation, quantitative precipitation estimation model training, and quantitative precipitation estimation model testing. The dataset required for this study is obtained by matching and preprocessing KMLB polarimetric radar data and GPM GV rain gauge data.

The training process involves continuously adjusting and updating the neural network model parameters based on a large number of sample data. In the quantitative precipitation estimation model training process, the sample data from the training set is inputted into the neural network. The ShuffleNet Block extracts local spatial information from the radar data, and then the Transformer-Encoder allocates global attention to extract the global spatial information of the radar data. The features of reflectivity factor, differential reflectivity, and specific differential phase are fused, and then the high-dimensional features in the Transformer-Encoder output feature vector are mapped to the sample space by the fully connected layer for regression calculation. At the same time, the validation set is used to verify the quantitative estimation results to monitor whether the model has overfitting during the learning process.

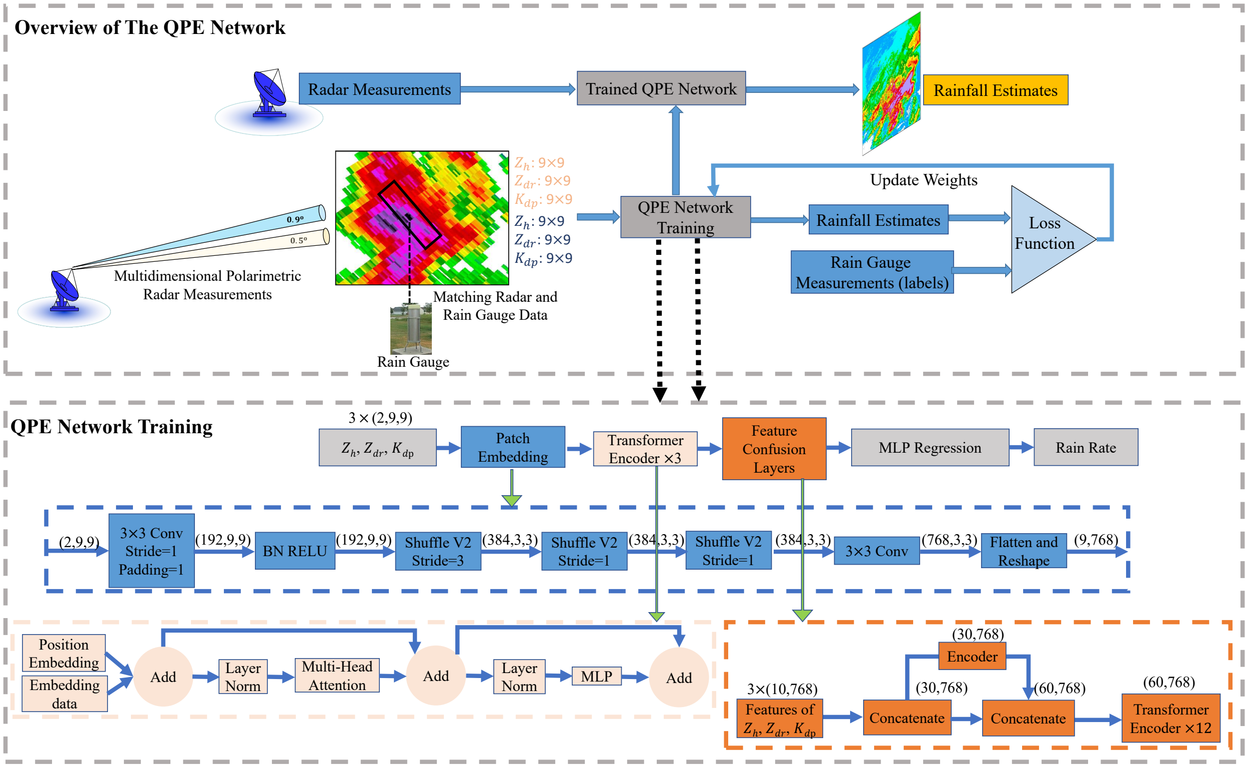

Figure 4.5: Conceptual diagram (upper block) of the machine learning framework for polarimetric radar rainfall estimation. Essentially, the network is trained to produce surface rainfall estimates based on multidimensional multiparameter radar observables. The rain gauge measurements are used as labels in the training phase. Detailed structure of the QPE network is illustrated in the lower block. BN and RELU refer to batch Normalization and rectified linear unit, respectively. MLP stands for multilayer.

Finally, the trained quantitative precipitation estimation model is used to estimate the quantitative precipitation of the test dataset and measure the algorithm's accuracy. The corresponding estimated results for each sample in the radar data from the test dataset are obtained by inputting them into the neural network, and then the evaluation criteria mentioned in Chapter 4 are calculated based on the estimated results and the ground truth data from the rain gauge. The results are visualized using a scatter plot to conduct quantitative and qualitative analysis. The flowchart of the proposed quantitative precipitation estimation model in this paper is shown in Figure 4.5.

4.3.2 Neural Network Structure for Quantitative Precipitation Estimation

This article proposes a neural network structure that combines the characteristics of Vision Transformer and ShuffleNet. The neural network model can be divided into three main parts: feature extraction, feature fusion, and regression calculation. The feature extraction part is completed by the PatchEmbedding layer, whose main structure is composed of ShuffleNet Blocks, as shown in Figure 4.5, with detailed types and parameters annotated (including the dimensions of input/output feature maps, convolution kernel dimensions, pad, and stride).

The role of this layer is to divide the input (4×9×9) radar data matrix into two feature extraction channels according to variable types. The inputs to these two channels are the (2×9×9) reflectivity factor matrix and the differential reflectivity matrix. Although the value ranges of the two input data are scaled uniformly to [-1, 1] after normalization preprocessing, reflectivity factor and differential reflectivity have different physical meanings. This is different from the RGB three-channel data used in computer image processing, so dual-feature extraction channels are used to process the input data separately.

When the (2×9×9) radar data matrix is input to the Patch Embedding layer, each (2×3×3) small region is extracted separately to obtain a high-dimensional feature vector for each Patch, which is implemented by the convolutional structure in Figure 3-11 in this study. The channel reuse, different sizes of convolution kernels, and channel shuffling techniques in the ShuffleNet Block make the feature vector better reflect the local variation information and high-dimensional features in the (3×3) small region, which is an improvement over the original Vision Transformer.

After the Patch Embedding operation is completed, three matrices of different input data types with a size of (9×768) are obtained, where 9 represents 9 (3×3) Patches and 768 represents the feature dimension. This study borrows the idea of Vision Transformer and adds learnable spatial information position encoding and regression vectors (regression token) to each feature extraction channel. After this operation, the size of the three matrices will become (10×768). At this point, the feature extraction stage ends, and the next stage is feature fusion. Since the three feature vector matrices are obtained from three different feature extraction channels, there is no correlation between them, so feature fusion is introduced to fuse the reflectivity factor, differential reflectivity, and Specific Differential Phase. In this stage, the concatenation feature vector matrix with a size of (30×768) is first obtained by concatenating them along the first dimension, and then skip connection is used to avoid gradient vanishing caused by too much depth. In the skip connection, as shown in Figure 4.5, one branch does not perform any operations to preserve the original information, and the other branch uses a Transformer-Encoder layer to perform initial feature fusion on the 30 feature vectors, and then concatenates the outputs of the two branches to obtain the complete feature vector matrix with a size of (60×768). The complete feature vector matrix establishes a relationship model between 60 feature vectors through the attention mechanism of 12 Transformer-Encoder layers, where the first 30 feature vectors contain single-variable features, while the latter 30 feature vectors contain mixed bivariate features. Each feature vector also has its own position information, and these 12 Transformer-Encoder layers rely on position encoding to establish spatial relationships between different Patches. At this point, the feature fusion stage ends, and the relationships between the reflectivity factor, differential reflectivity, and Specific Differential Phase, as well as the relationships between different positions, are established.

During the regression calculation phase, the main structure consists of fully connected layers. The input for this phase is a complete feature vector matrix of size (60x768), but due to the addition of regression vectors in the Patch Embedding phase, the input to the fully connected layers is composed of feature vectors at positions 0, 10, 20, 30, 40, and 50 in the complete feature vector matrix. These four regression vectors form the input to the fully connected layers and have a size of (1x4608). The regression calculation phase uses the fully connected layers to nonlinearly map the regression vectors of size (1x4608) to a scalar value of size (1x1), which represents the estimated precipitation intensity of the final model.

4.4 Results and Practical Performance

In order to demonstrate and evaluate the superior performance of the deep learning-based polarimetric radar rainfall estimates, this chapter compares the deep learning model with three widely used parametric radar rainfall relations with independent testing dataset from 2017 to 2018, including two $Z - R$ relations (WSR-88D convective $Z - R$ relation [44] in Eq. 4.1a and the Marshall/Palmer stratiform $Z - R$ relation [45] in Eq. 4.1b) and a polarimetric relation (Eq. 4.1c) used by WSR-88D in operatinal environment [46].

In order to better assess the accuracy of quantitative precipitation estimates (QPE) from various networks, this section will quantitatively analyze the estimation results for individual precipitation cases in the study area. The temporal resolution for accumulated precipitation is 1 hour. The data for the four precipitation cases are radar and rain gauge data collected on May 1, 2017, July 20, 2017, May 15, 2018 and August 28, 2018, with varying durations of 4-6 hours.

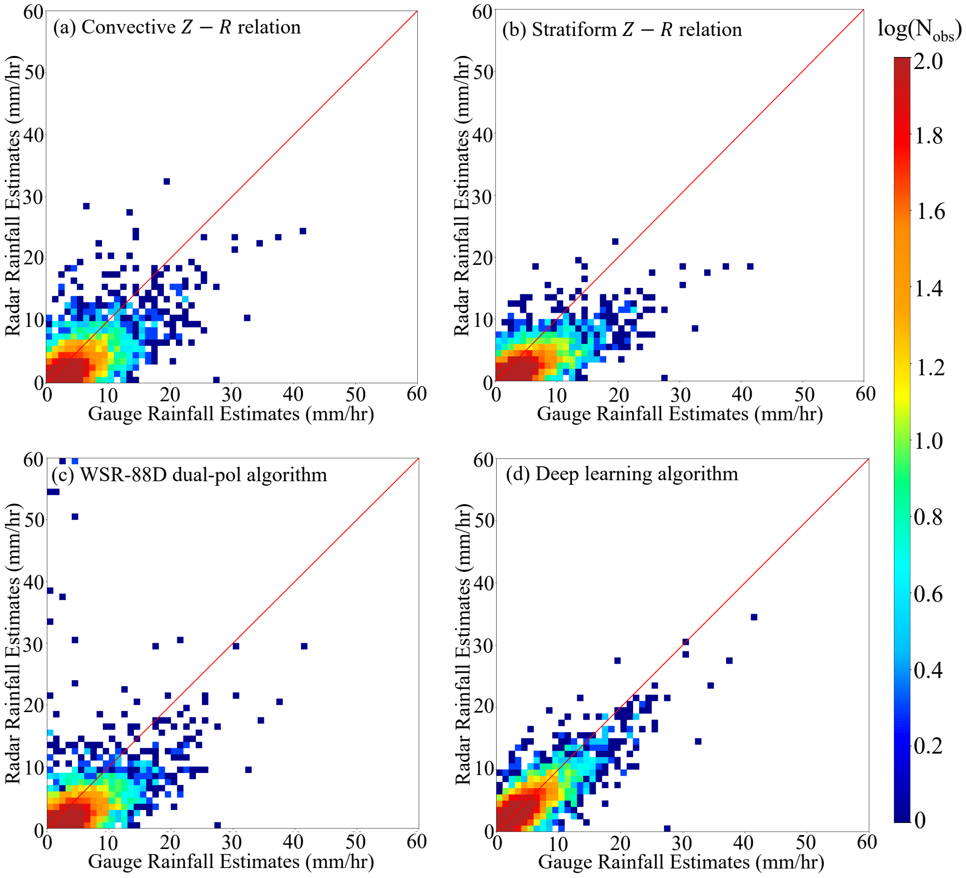

Figure 4.6: 2D histograms of radar rainfall estimates vs. rain gauge observations based on independent test data from 2017 to 2018: (a) WSR-88D convective $R(Z_h)$ relation; (b) WSR-88D stratiform $R(Z_h)$ relation; (c) WSR-88D dual-pol algorithm;(d) deep learning model proposed in this study. The quantitative evaluation results are detailed in Fig. 4.7.

$$Z_h = 300R^{1.4} \tag{4.1a}$$

$$Z_h = 200R^{1.6} \tag{4.1b}$$

$$R = 1.42 \times 10^{-2} \times Z_h^{0.770} \times \left(10^{Z_{dr}/10}\right)^{-1.67} \tag{4.1c}$$

Figure 4.6 shows the scatter plots of precipitation for the testing dataset in the study area where subgraphs (a)-(d) present the radar rainfall estimates compared with gauge rainfall estimates by using convenctive $R(Z_h)$ relation, stratiform $R(Z_h)$ relation, WSR-88D dual-pol relation and deep learning model, respectively. The horizontal coordinate is the true value of ground surface measured by the rain gauge and the vertical coordinate is three operational radar rainfall relations and deep learning model. The colorbar represents the number of samples within the precipitation interval. Figure 4.6(a)-(c) illustrate that the three radar rainfall algorithms are overestimate the precipitation where the true value around 0-15 mm. However, the deep learning algorithm model can effectively limit the overestimation of precipitation. In addition, when the true value is higher than 20 mm, the three radar rainfall algorithms is underestimate, which are basically below the diagonal line. However, Figure 4.6(d) illustrates that the deep learning model proposed in this paper can significantly improve the underestimation of heavy precipitation estimations. Therefore, the deep learning based model can effectively estimate the precipitation generally in the study domain quantitatively.

Then the following scoring metrics which including mean absolute error (MAE), root mean square error (RMSE), correlation coefficient (CORR), normalized standard error (NSE) and bias ratio (BR) are calculated for quantitative assessment of the QPE algorithms with ground rain gauge measurements.

$$\mathrm{MAE} = \frac{\sum_{n=1}^{n} |R_n - G_n|}{N} \tag{4.2a}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{n=1}^{N} |R_n - G_n|^2}{N}} \tag{4.2b}$$

$$\mathrm{CORR} = \frac{\sum_{n=1}^{N} \left(R_n - \overline{R_n}\right)\left(G_n - \overline{G_n}\right)}{\sqrt{\sum_{n=1}^{N} \left(R_n - \overline{R_n}\right)^2}\sqrt{\sum_{n=1}^{N} \left(G_n - \overline{G_n}\right)^2}} \tag{4.2c}$$

$$\mathrm{NSE} = \frac{\sum_{n=1}^{N} |R_n - G_n|}{\sum_{n=1}^{N} G_n} \tag{4.2d}$$

$$\mathrm{BR} = \frac{\sum_{n=1}^{N} R_n}{\sum_{n=1}^{N} G_n} \tag{4.2e}$$

where R_n and G_n stand for the radar rainfall estimate and gauge rainfall record (mm/hr) at sample time n; $\overline{R_n}$ and $\overline{G_n}$ denote the mean of radar and gauge rainfall estimates; N stands for the total number of samples.

Figure 4.7 exhibits that the deep learning algorithm has the better overall estimation performance compared with WSR-88D convective $R(Z_h)$ relation, WSR-88D stratiform $R(Z_h)$ relation and WSR-88D dual-pol algorithm. The smaller NSE, RMSE and MAE values is, the closer the estimated value of quantitative precipitation is to the observed value of the rain gauge. The higher CORR and BR means the more accurate radar estimation performance to the ground surface precipitation. In general, the deep learning model has the best performance evaluation scores among the four radar rainfall algorithms based on the independent testing dataset.

To visually analyze the rainfall estimates of each algorithm, this study performed a qualitative analysis of the estimation results for four individual precipitation cases in the study area in 2017 and 2018. Figure 4.84.94.104.11 show the plots of the cases which contains a more extensive range of heavy precipitation. All the four figures illustrate that three operational radar precipitation relations are underestimate the precipitation for most areas, especially where has heavy rain, which is consistent with the scatter plot analysis in Figure 4.6. The closeness and behavior of rainfall indicated by the plots demonstrate that there is a good potential this nonparametric technique can

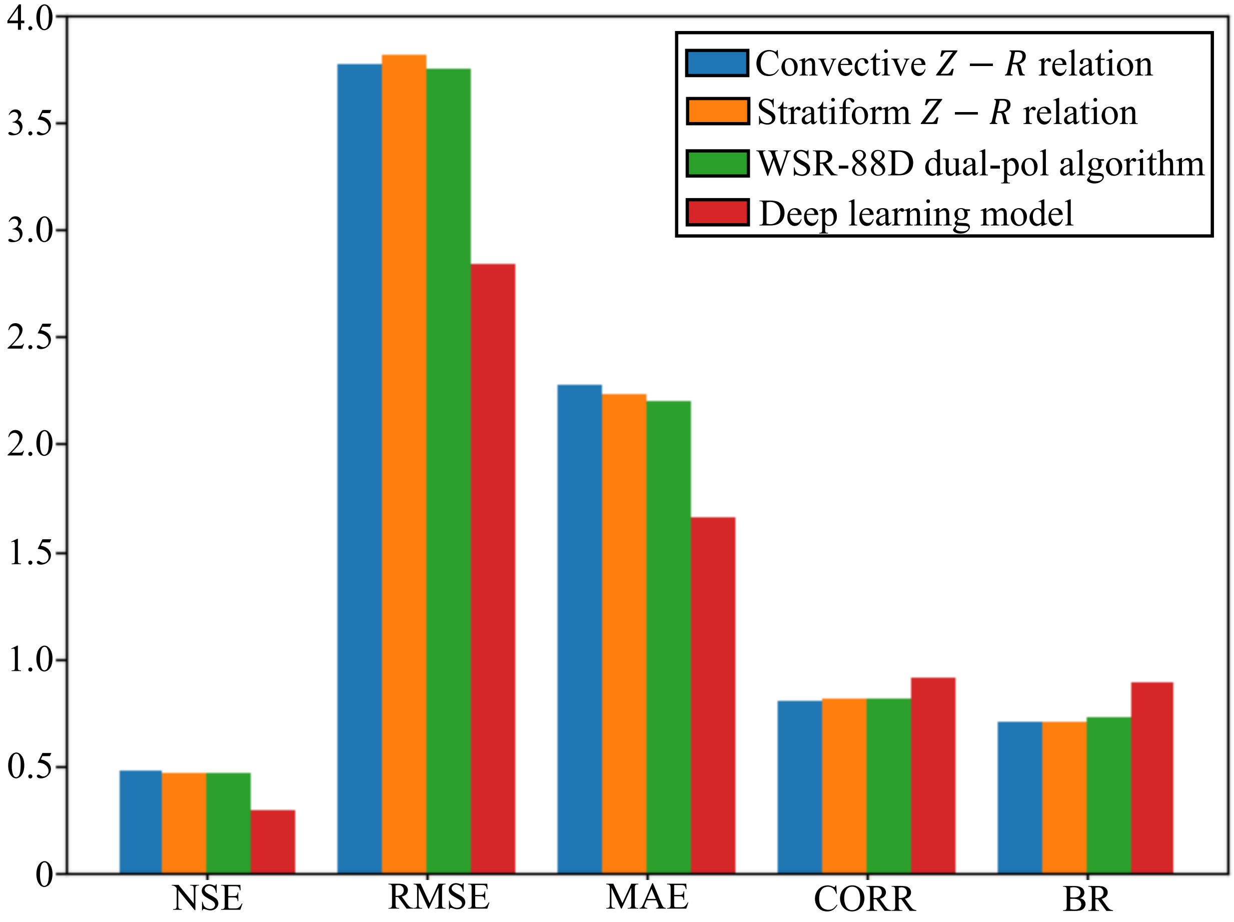

Figure 4.7: Performance evaluation scores of various radar rainfall algorithms based on independent test data from 2017 to 2018.

be applied to radar QPE. However, the different between deep learning algorithm estimation and two operational MRMS products are not significant. Based on this, this study take a quantitative analysis of this case.

Analyzing Figure 4.12, it becomes evident that the deep learning algorithm exhibits superior performance across multiple evaluation metrics such as MAE, RMSE, CORR, NME, and BR. This finding highlights the effectiveness of the deep learning approach in accurately estimating rainfall.

Furthermore, the quantitative analysis demonstrates the advantages of the nonparametric scheme over conventional methods that rely on "fixed" relationships between radar parametrics and rainfall. The nonparametric scheme allows for adaptive relationships that can better capture the variability

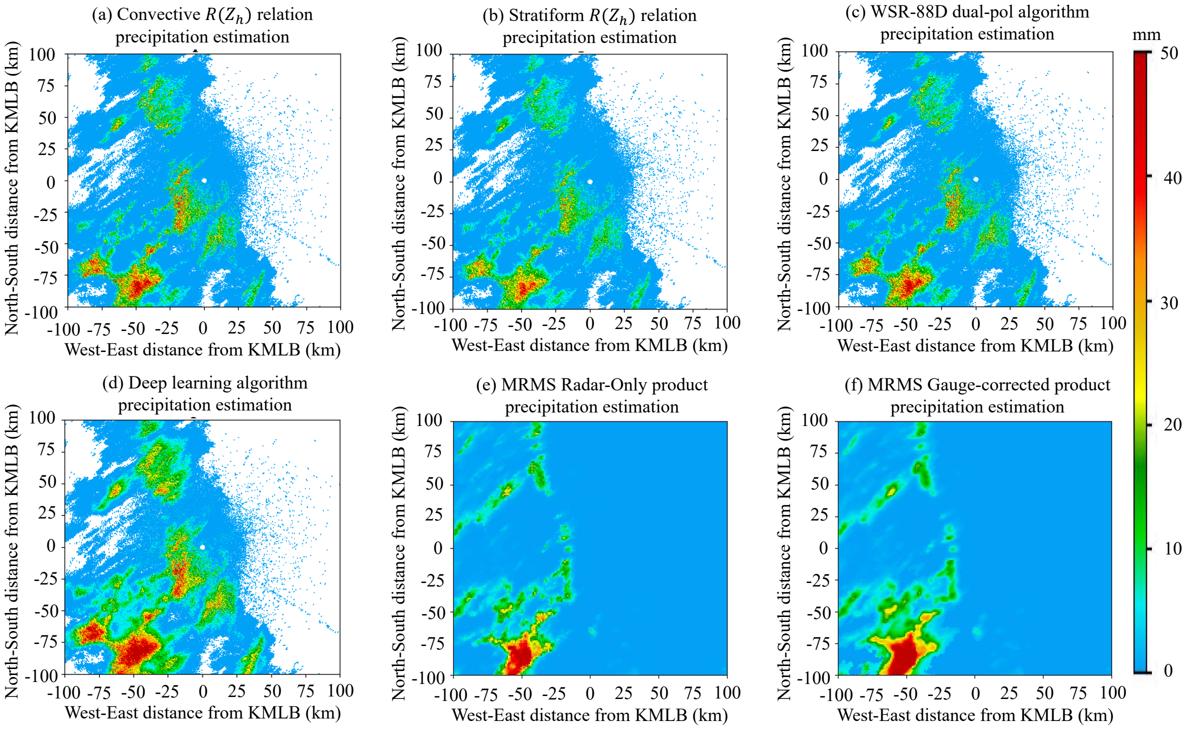

Figure 4.8: Rainfall accumulations from 2000UTC to 2200UTC, May 1, 2017 based on different radar rainfall algorithms: (a) WSR-88D convective $R(Z_h)$ relation; (b) WSR-88D stratiform $R(Z_h)$ relation; (c) WSR-88D dual-pol algorithm; (d) deep learning model proposed in this study; (e) operational product from MRMS (radar-only); (f) operational product from MRMS (gauge-corrected).

of precipitation. In contrast, traditional approaches may fail to adequately represent the intricate and dynamic nature of precipitation patterns. The flexibility and adaptability offered by the non-parametric scheme enhance the accuracy and reliability of rainfall estimation.

4.5 Summary and Conclusion

Based on this study, it is concluded that deep learning can allow for adaptive mapping between polarimetric weather radar observations and ground surface rain rate, which may not be achieved through conventional “fixed” parametric relations. The training, validation, and test datasets based on multi-year operational radar observations and rain gauge measurements from SFL have quantitatively demonstrated the better performance of deep learning for precipitation estimation. Future study will focus on operational implementation of the proposed deep learning system in generat-

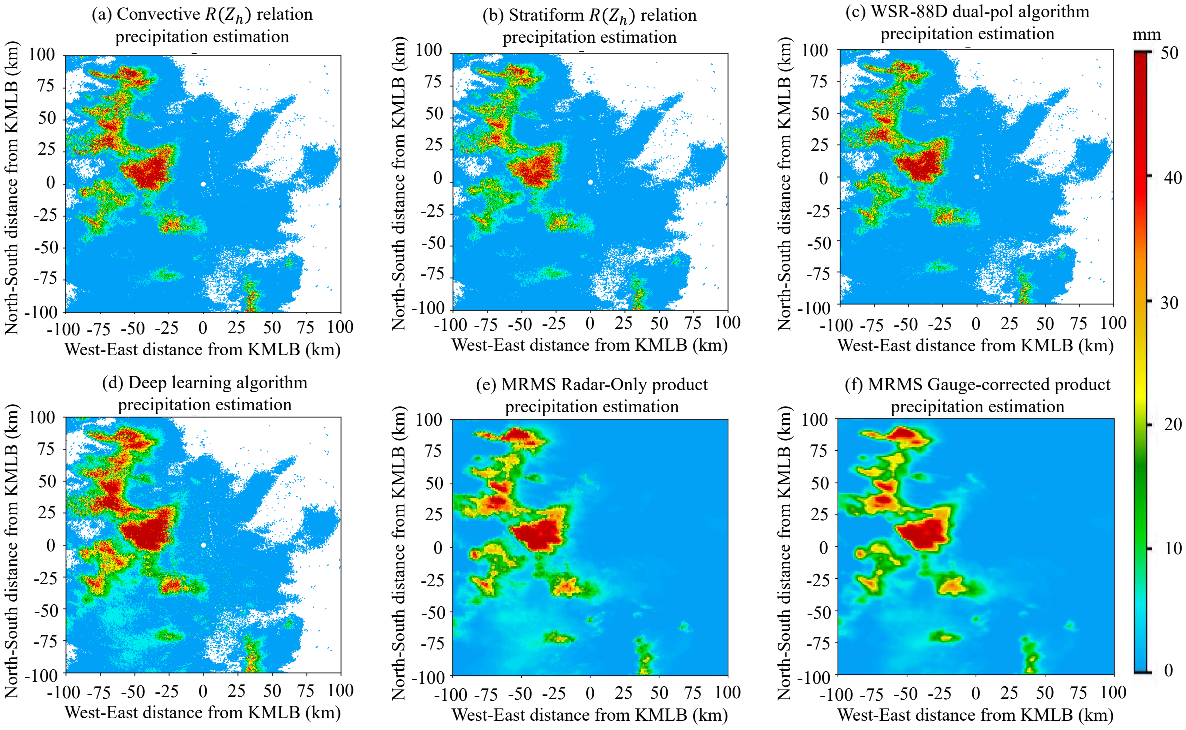

Figure 4.9: Rainfall accumulations from 2100UTC to 2300UTC, July 20, 2017 based on different radar rainfall algorithms: (a) WSR-88D convective $R(Z_h)$ relation; (b) WSR-88D stratiform $R(Z_h)$ relation; (c) WSR-88D dual-pol algorithm; (d) deep learning model proposed in this study; (e) operational product from MRMS (radar-only); (f) operational product from MRMS (gauge-corrected).

ing real-time rainfall products as well as incorporating nowcasting models for producing short-term predictions of rainfall.

Through qualitative and quantitative analysis, the validity and reliability of the precipitation estimation model proposed in this paper are verified. From the qualitative analysis, the performance of deep learning models is better than that of traditional algorithms, and the quantitative precipitation estimation model proposed in this paper has the best performance in heavy precipitation events. In addition, from the perspective of the three evaluation criteria of quantitative analysis, the quantitative precipitation estimation model proposed in this paper is superior to other quantitative precipitation estimation algorithms, and the good estimation results can also be seen in the scatter plot, which can significantly improve the performance of heavy rainfall events. underestimation problem. The performance in the precipitation case proves the effectiveness and feasibility of using

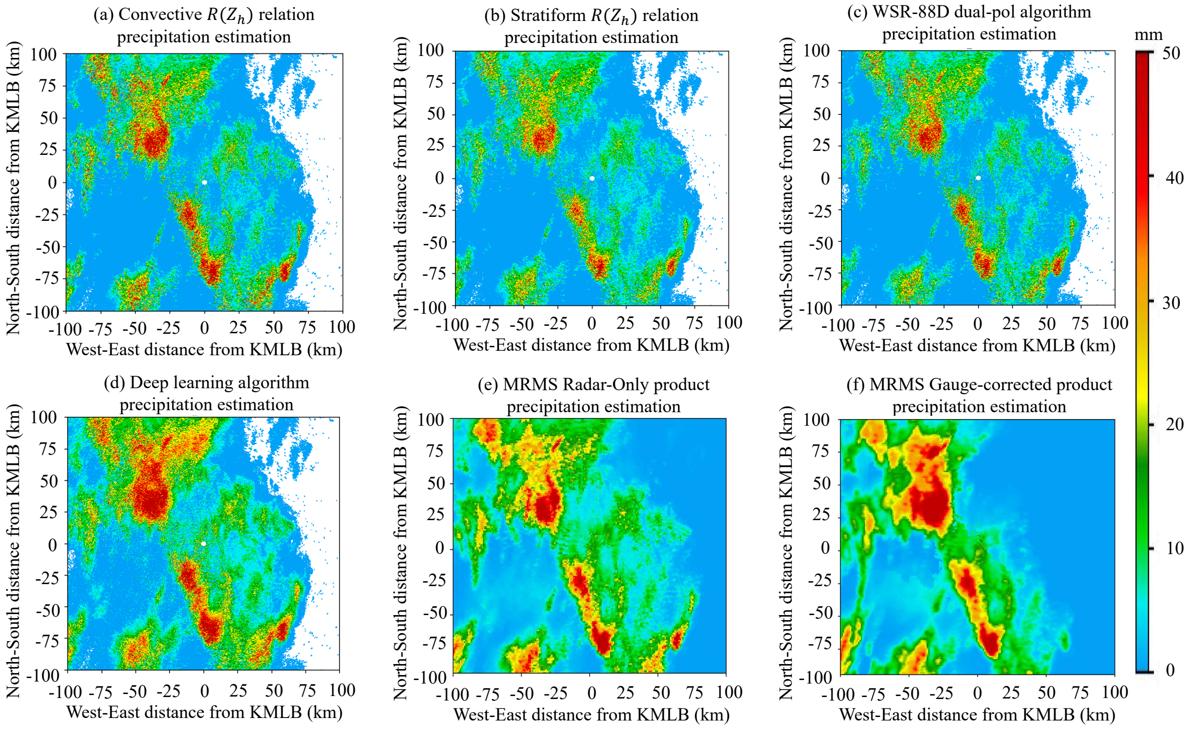

Figure 4.10: Rainfall accumulations from 1900UTC to 2300UTC, May 15, 2018 based on different radar rainfall algorithms: (a) WSR-88D convective $R(Z_h)$ relation; (b) WSR-88D stratiform $R(Z_h)$ relation; (c) WSR-88D dual-pol algorithm; (d) deep learning model proposed in this study; (e) operational product from MRMS (radar-only); (f) operational product from MRMS (gauge-corrected).

the deep learning algorithm based on dual-polarization radar data, and the dual-polarization radar observation data with spatial information can significantly improve the accuracy and stability of quantitative precipitation estimation.

This deep learning algorithm can produce comparative estimates with the operational dual-polarization radar product if applied to other regions without changing the model, which is very appealing especially in the areas such as ocean and remote regions where no gauge or radar available. With more and more gauges and ground radars being deployed in the world and many of them become operational, this hybrid system can be trained at different locations with different atmosphere properties to further increase the dynamic adjustment and generality.

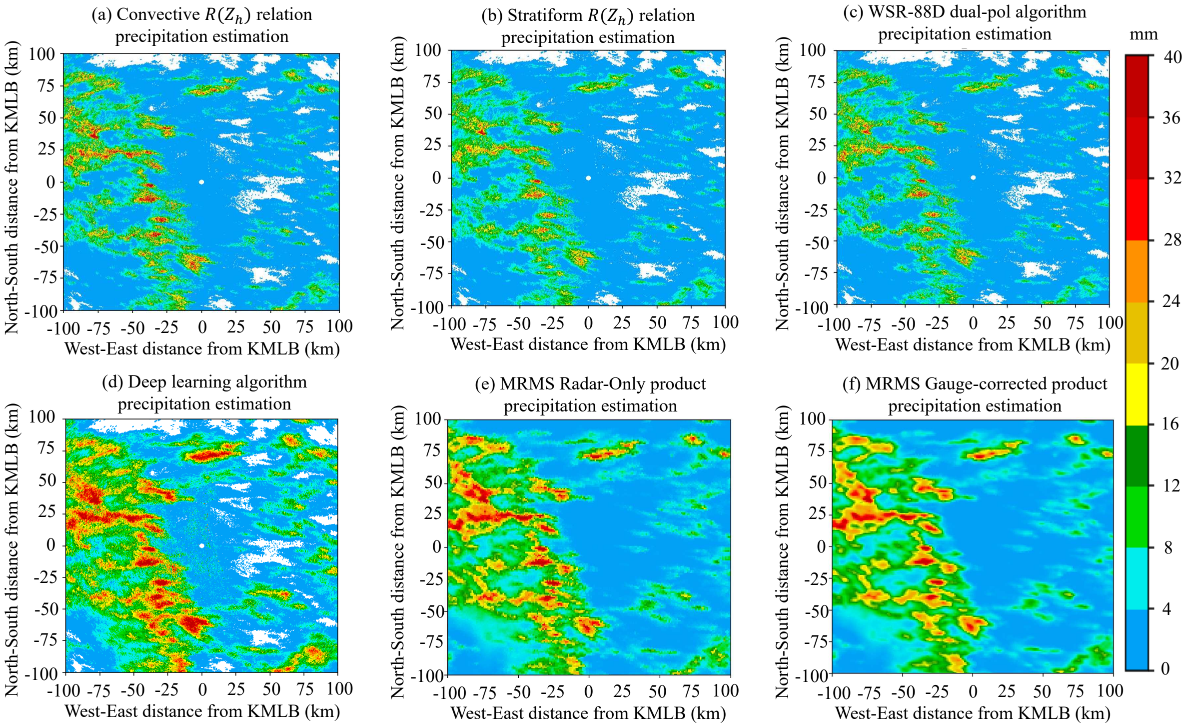

Figure 4.11: Rainfall accumulations from 1600UTC to 2200UTC, August 28, 2018 based on different radar rainfall algorithms: (a) WSR-88D convective $R(Z_h)$ relation; (b) WSR-88D stratiform $R(Z_h)$ relation; (c) WSR-88D dual-pol algorithm; (d) deep learning model proposed in this study; (e) operational product from MRMS (radar-only); (f) operational product from MRMS (gauge-corrected).

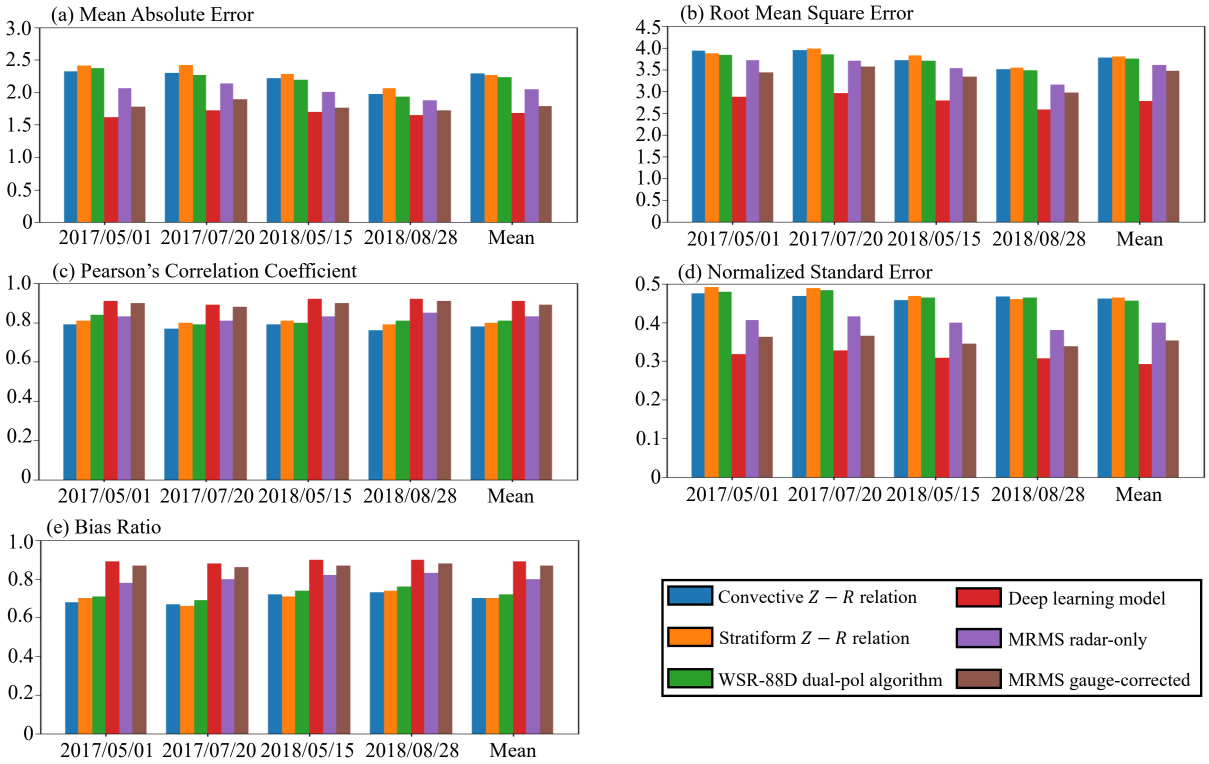

Figure 4.12: Quantitative evaluation results of radar rainfall estimates during the four selected precipitation events, and the mean results combining estimates from the four events: (a) mean absolute error (MAE); (b) root mean square error (RMSE); (c) correlation coefficient; (d) normalized standard error (NSE); and (e) bias ratio (BR).

Chapter 5

Summary and Conclusions

5.1 Summary

Precipitation has important significance for social production activities. Accurate and timely estimation of precipitation intensity can guide transportation, industrial and agricultural production, water conservancy and hydropower, flood and disaster prevention and warning engineering. Starting from practical problems, this paper first discusses the improvement of accuracy of quantitative precipitation estimation by lowering the radar observation elevation angle. Then, for the task of quantitative precipitation estimation based on polarimetric Doppler radar data, a deep learning precipitation estimation model based on depth-wise separable convolution and Transformer structure is proposed. By comparing with the results of quantitative estimation using traditional methods and MRMS, and based on the evaluation criteria proposed in previous studies, quantitative and qualitative analysis of QPE results is conducted, and the experimental results show the effectiveness of the proposed precipitation estimation model. The main research work of this paper is as follows: (1): The study establishes polarimetric radar rainfall relations based on local disdrometer data to estimate surface rainfall from radar observations. Rainfall estimates using the KMUX radar observations at both 0° and 0.5° scan elevation angles are compared with rainfall measurements obtained from surface rain gauges. Results show that the lower scan elevation angle provides more accurate radar-based rainfall estimates. (2): Data sample spatiotemporal coordinate matching, handling of non-numeric data in radar data, and normalization of input data to the model are carried out to address the problem of different spatiotemporal resolutions of radar and rain gauge data. Meanwhile, a deep learning model combining depth-wise separable convolution and Transformer structure is proposed for quantitative precipitation estimation based on polarimetric radar. The quantitative precipitation estimation model is divided into three parts: feature extraction, feature fusion, and regression calculation. Depth-wise separable convolution is used to extract local spatial information of precipitation reflection in radar data. Transformer-Encoder is used for local feature information fusion, and the feature vector is finally obtained through regression calculation by fully connected layers to obtain precipitation intensity. Three input channels are used for reflectivity factor, differential reflectivity factor, and Specific Differential Phase because these three variables have different physical meanings, value ranges, and dimensions. (3): The experimental results show that data-driven deep learning models can significantly improve the accuracy of radar quantitative precipitation estimation compared with the Z-R relationship method. Moreover, in the deep learning model, the use of radar data within a certain range as input results in the most accurate model estimation. By analyzing the performance of precipitation cases in 2017 and 2018, it is shown that the spatial information of polarimetric radar data can improve the accuracy and stability of radar quantitative precipitation estimation and alleviate the problem of underestimating heavy rainfall. Therefore, the effectiveness of the deep learning-based polarimetric radar quantitative precipitation estimation model proposed in this study is verified.

5.2 Future Work

This study first demonstrated that reducing the minimum elevation angle of a radar improves the accuracy of quantitative precipitation estimation, and then used a deep learning precipitation model to invert precipitation from low-elevation angle data obtained from a dual-polarization Doppler radar. The accuracy of the deep learning algorithm was verified using data from 2016 to 2021. The effectiveness of the model was demonstrated through four precipitation cases in 2017 and 2018. In future work, deeper research can be conducted in the following areas:

(1) In future research, it is important to investigate how incorporating the vertical structure of radar observations and topographical information in complex terrain, can further improve radar-based quantitative precipitation estimation (QPE). Additionally, conducting similar studies with WSR-88D stations in different geophysical regions and seasons will help to fully understand the potential benefits of the upgraded scan strategy for radar-based QPE. It should be noted that radar QPE uncertainty is influenced by various physiographic factors.

(2) Precipitation is a continuous spatial and temporal process. The precipitation estimation model proposed in this paper, based on separable convolution and Transformer structures, extracts and integrates only spatial position information from the input data, and performs quantitative precipitation estimation using the spatial information of radar observation data. However, this model cannot extract information about precipitation events in the time dimension from input continuous radar observation time-series data. In subsequent research, neural network models that can handle time-series data can be used, while modeling the relationship between precipitation in time and space dimensions.

(3) The observation parameters obtained from a dual-polarization Doppler radar are strongly correlated with the physical structure information of precipitation particles, such as their type and phase, and can better reflect the characteristics of precipitation droplet spectra. Therefore, using more observation parameters from a dual-polarization Doppler radar as input features in the estimation model can improve the accuracy of quantitative precipitation estimation. This is also the main way to improve the traditional Z-R relationship in the field of radar quantitative precipitation estimation research. The dual-polarization radar observation parameters used in this study are reflectivity factor, differential reflectivity, and differential phase shift. Although this can reduce uncertainty caused by differences in droplet spectra, it is possible to consider adding other radar observation parameters in future research to further improve the quality of QPE products.

(4) In recent years, researchers have not only focused on the accuracy of models in different tasks but also on the floating-point computational power requirements of deep learning models and the time consumed by the model inference process. More reasonable and scientific model structure design and hyperparameter selection not only play an important role in improving the accuracy of precipitation estimation models but also affect the practical application and deployment of the model. In future research, while increasing the depth of precipitation estimation models, the computational complexity of the model should be reduced through better structure design.

(5) The problem of imbalanced data distribution in the dataset affects the performance of deep learning models. The deep learning model proposed in this paper cannot fundamentally solve

the problem of the small number of strong precipitation cases, and may still underestimate the precipitation to some extent. In subsequent research, the dataset can be expanded, and the number of strong precipitation cases can be increased to improve the underestimation phenomenon.

5.3 Multisensor Precipitation Estimation

In recent years, there has been an increasing interest in using multisensor precipitation estimation to improve the accuracy and reliability of precipitation measurements. The approach combines data from multiple sources, such as radar, satellite, and rain gauges, to generate more comprehensive and accurate precipitation estimates. While there have been many studies on the use of individual sensors for precipitation estimation, relatively few have explored the potential benefits of integrating multiple sensors.

In future work, it is important to investigate the effectiveness of multisensor precipitation estimation in different geographical regions and under various weather conditions. One area of focus should be on developing new algorithms and techniques for integrating data from different sensors, as well as for dealing with data quality issues and uncertainties. Another area of research could be on optimizing the design and placement of the sensors to maximize their effectiveness and coverage.

Ultimately, the goal of using multisensor precipitation estimation is to improve our understanding of precipitation patterns and processes, which is critical for a wide range of applications, such as weather forecasting, flood prediction, and water resources management. By advancing our knowledge and capabilities in this area, we can better prepare for and respond to extreme weather events, and ultimately improve the resilience and sustainability of our communities and ecosystems.

Bibliography

[1] K. Yilmaz, T. Hogue, K. Hsu, S. Sorooshian, H. Gupta, and T. Wagener. Intercomparison of rain gauge, radar, and satellite-based precipitation estimates with emphasis on hydrologic forecasting. *Journal of Hydrometeorology*, pages 497–517, 2005.

[2] C. Kidd and V. Levizzani. Status of satellite precipitation retrievals. *Hydrol. Earth Syst. Sci.*, 15(4):1109–1116, 2011.

[3] T. Seliga and V. Bringi. Potential use of radar differential reflectivity measurements at orthogonal polarizations for measuring precipitation. *Journal of Applied Meteorology and Climatology.*, 15(1):69–76, 1976.

[4] P. Smith. Equivalent radar reflectivity factors for snow and ice particles. *Journal of Applied Meteorology and Climatology*, 23(8):1258–1260, 1984.

[5] Richard A. Fulton, Jay P. Breidenbach, Dong-Jun Seo, Dennis A. Miller, and Timothy O'Bannon. The WSR-88D Rainfall Algorithm. *Weather and Forecasting*, 13(2):377 – 395, 1998.

[6] V Chandrasekar, Haonan Chen, and Brenda Philips. Principles of high-resolution radar network for hazard mitigation and disaster management in an urban environment. *Journal of the Meteorological Society of Japan. Ser. II*, 96:119–139, 2018.

[7] Robert Cifelli, V. Chandrasekar, Haonan Chen, and Lynn E. Johnson. High resolution radar quantitative precipitation estimation in the san francisco bay area: Rainfall monitoring for the urban environment. *Journal of the Meteorological Society of Japan. Ser. II*, 96A:141–155, 2018.

[8] Rob Cifelli, V Chandrasekar, Liv Herdman, David D Turner, Allen B White, TI Alcott, M Anderson, P Barnard, SK Biswas, M Boucher, et al. Advanced quantitative precipitation infor-

mation: Improving monitoring and forecasts of precipitation, streamflow, and coastal flooding in the san francisco bay area. *Bulletin of the American Meteorological Society*, 2022.

[9] Haonan Chen, Robert Cifelli, and Allen White. Improving Operational Radar Rainfall Estimates Using Profiler Observations Over Complex Terrain in Northern California. *IEEE Transactions on Geoscience and Remote Sensing*, 58(3):1821–1832, 2020.

[10] D Willie, H Chen, V Chandrasekar, R Cifelli, C Campbell, D Reynolds, S Matrosov, and Y Zhang. Evaluation of multisensor quantitative precipitation estimation in Russian River Basin. *Journal of Hydrologic Engineering*, 22(5):E5016002, 2017.

[11] Haonan Chen, Rob Cifelli, V. Chandrasekar, and Yingzhao Ma. A flexible bayesian approach to bias correction of radar-derived precipitation estimates over complex terrain: Model design and initial verification. *Journal of Hydrometeorology*, 20(12):2367 – 2382, 2019.

[12] F. Martin Ralph, Paul J. Neiman, Gary A. Wick, Seth I. Gutman, Michael D. Dettinger, Daniel R. Cayan, and Allen B. White. Flooding on california's russian river: Role of atmospheric rivers. *Geophysical Research Letters*, 33(13), 2006.

[13] Alexander V. Ryzhkov, Scott E. Giangrande, and Terry J. Schuur. Rainfall estimation with a polarimetric prototype of wsr-88d. *Journal of Applied Meteorology*, 44(4):502 – 515, 2005.

[14] Yanting Wang and V. Chandrasekar. Quantitative precipitation estimation in the casa x-band dual-polarization radar network. *Journal of Atmospheric and Oceanic Technology*, 27(10):1665 – 1676, 2010.

[15] Haonan Chen and V. Chandrasekar. The quantitative precipitation estimation system for Dallas–Fort Worth (DFW) urban remote sensing network. *Journal of Hydrology*, 531:259–271, 2015.

[16] T.B Trafalis, M.B Richman, A. White, and B. Santosa. Data mining techniques for improved wsr-88d rainfall estimation. *Computers Industrial Engineering*, 43(4):775–786, 2002.

[17] H. Chen, V. Chandrasekar, H. Tan, and R. Cifelli. Rainfall estimation from ground radar and trmm precipitation radar using hybrid deep neural networks. *Geophysical Research Letters*, 46:10669–10678, 2019.

[18] Y.M Chiang, F.J Chang, B.J.D Jou, and F.P. Lin. Dynamic ann for precipitation estimation and forecasting from radar observations. *Journal of Hydrology*, 334(1):250–261, 2007.

[19] Yajing Wu, Yongqiang Tang, Xuebing Yang, Wensheng Zhang, and Guoping Zhang. Graph convolutional regression networks for quantitative precipitation estimation. *IEEE Geoscience and Remote Sensing Letters*, 18(7):1124–1128, 2021.

[20] Xuan Peng, Qian Li, and Jinrui Jing. Cngat: A graph neural network model for radar quantitative precipitation estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2022.

[21] N. V Chawla, K. W Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: Synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[22] D. H Hubel and T. N. Wiesel. Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, 195(1):215–243, 1968.

[23] Y LeCun and J. Boser, B. Denker. Handwritten digit recognition with a back-propagation network. *Advances in neural information processing systems*, 2, 1989.

[24] A Krizhevsky, I Sutskever, and Hinton G.E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systemsg*, 25, 2012.

[25] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, Wenyu Liu, and Bin Xiao. Deep high-resolution representation learning for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10):3349–3364, 2021.

[26] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. *International conference on machine learning*, pages 6105–6114, 2019.

[27] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1492–1500, 2017.

[28] Fisher Yu and Vladlen Koltun. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*, 2015.

[29] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. *Proceedings of the IEEE international conference on computer vision*, pages 764–773, 2017.

[30] Wenjie Luo, Yujia Li, Raquel Urtasun, and Richard Zemel. Understanding the effective receptive field in deep convolutional neural networks. *Advances in neural information processing systems*, 29, 2016.

[31] François Chollet. Xception: Deep learning with depthwise separable convolutions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017.

[32] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. Shufflenet: An extremely efficient convolutional neural network for mobile devices. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6848–6856, 2018.

[33] Ningning Ma, Xiangyu Zhang, Hai-Tao Zheng, and Jian Sun. Shufflenet v2: Practical guidelines for efficient cnn architecture design. *Proceedings of the European conference on computer vision (ECCV)*, pages 116–131, 2018.

[34] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 20(11):1254–1259, 1998.

[35] Volodymyr Mnih, Nicolas Heess, Alex Graves, et al. Recurrent models of visual attention. *Advances in neural information processing systems*, 27, 2014.

[36] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, 27, 2014.

[37] Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

[38] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[39] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[40] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[41] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

[42] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[43] Michał Daniluk, Tim Rocktäschel, Johannes Welbl, and Sebastian Riedel. Frustratingly short attention spans in neural language modeling. *arXiv preprint arXiv:1702.04521*, 2017.

[44] Richard A. Fulton, Jay P. Breidenbach, Dong-Jun Seo, Dennis A. Miller, and Timothy O'Bannon. The wsr-88d rainfall algorithm. *Weather and Forecasting*, 13(2):377 – 395, 1998.

[45] J. S. Marshall and W. Mc K. Palmer. The distribution of raindrops with size. *Journal of Atmospheric Sciences*, 5(4):165 – 166, 1948.

[46] Scott E. Giangrande and Alexander V. Ryzhkov. Estimation of rainfall based on the results of polarimetric echo classification. *Journal of Applied Meteorology and Climatology*, 47(9):2445 – 2462, 2008.