

THESIS

MACHINE LEARNING-BASED FUSION STUDIES OF RAINFALL ESTIMATION FROM SPACEBORNE AND  
GROUND-BASED RADARS

Submitted by

Haiming Tan

Department of Computer Science

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Spring 2019

Master's Committee:

Advisor: Charles W. Anderson

Co-advisor: Chandrasekar V. Chandra

Indrajit Ray

Jose L. Chavez

Copyright by Haiming Tan 2019

All Rights Reserved

## ABSTRACT

### MACHINE LEARNING-BASED FUSION STUDIES OF RAINFALL ESTIMATION FROM SPACEBORNE AND GROUND-BASED RADARS

Precipitation measurement by satellite radar plays a significant role in researching the water circle and forecasting extreme weather event. Tropical Rainfall Measuring Mission (TRMM) Precipitation Radar (PR) has capability of providing a high-resolution vertical profile of precipitation over the tropics regions. Its successor, Global Precipitation Measurement (GPM) Dual-frequency Precipitation Radar (DPR), can provide detailed information on the microphysical properties of precipitation particles, quantify particle size distribution and quantitatively measure light rain and falling snow.

This thesis presents a novel Machine Learning system for ground-based and space borne radar rainfall estimation. The system first trains ground radar data for rainfall estimation using rainfall measurements from gauges and subsequently uses the ground radar based rainfall estimates to train spaceborne radar data in order to get space based rainfall product. Therein, data alignment between spaceborne and ground radar is conducted using the methodology proposed by Bolen and Chandrasekar (2013), which can minimize the effects of potential geometric distortion of spaceborne radar observations.

For demonstration purposes, rainfall measurements from three rain gauge networks near Melbourne, Florida, are used for training and validation purposes. These three gauge networks, which are located in Kennedy Space Center (KSC), South Florida Water Management District (SFL), and St. Johns Water Management District (STJ), include 33, 46, and 99 rain gauge stations, respectively. Collocated ground radar observations from the National Weather Service (NWS) Weather Surveillance

Radar – 1988 Doppler (WSR-88D) in Melbourne (i.e., KMLB radar) are trained with the gauge measurements. The trained model is then used to derive KMLB radar based rainfall product, which is used to train both TRMM PR and GPM DPR data collected from coincident overpasses events. The machine learning based rainfall product is compared against the standard satellite products, which shows great potential of the machine learning concept in satellite radar rainfall estimation. Also, the local rain maps generated by machine learning system at KMLB area are demonstrate the application potential.

TABLE OF CONTENTS

**ABSTRACT ..... II**

**1 INTRODUCTION ..... 1**

1.1 INTRODUCTION ..... 1

1.2 LITERATURE REVIEW ..... 3

    1.2.1 *Rainfall estimation based on radar measurements* ..... 3

1.3 RESEARCH QUESTION AND SPECIFIC AIMS ..... 5

**2 BACKGROUND ..... 6**

2.1 RADAR OBSERVATION AND RAINFALL ESTIMATION ..... 6

2.2 MACHINE LEARNING AND NEURAL NETWORKS ..... 7

    2.2.1 *Machine learning basics* ..... 7

    2.2.2 *Deep neural networks* ..... 8

2.3 SPACEBORNE AND GROUND-BASED RADAR ALIGNMENT METHODOLOGY ..... 14

    2.3.1 *Alignment methodology* ..... 15

**3 OBSERVATION AND MEASUREMENT PLATFORM ..... 17**

3.1 NEXT GENERATION WEATHER RADAR (NEXRAD) ..... 17

    3.1.1 *Introduction* ..... 17

    3.1.2 *NEXRAD system overview* ..... 18

    3.1.3 *Operation* ..... 19

    3.1.4 *Ground validation rain gauge* ..... 20

3.2 TROPICAL RAINFALL MEASURING MISSION (TRMM) ..... 21

    3.2.1 *TRMM instruments* ..... 21

    3.2.2 *TRMM measurement levels* ..... 23

3.2.3	<i>TRMM precipitation radar (PR) overview</i> .....	27
3.2.4	<i>TRMM-PR rain rate estimation</i> .....	29
3.3	GLOBAL PRECIPITATION MEASUREMENT (GPM).....	32
3.3.1	<i>GPM instrument</i> .....	32
3.3.2	<i>GPM_DPR overview</i> .....	33
3.3.3	<i>GPM DPR level data</i> .....	35
3.3.4	<i>GPM rainfall estimation</i> .....	37
<b>4</b>	<b>DEVELOPMENT OF DEEP NEURAL NETWORK MODEL FOR GROUND RADAR RAINFALL ESTIMATION.....</b>	<b>39</b>
4.1	NEURAL NETWORK IN RAINFALL ESTIMATION .....	39
4.2	NEURAL NETWORK ARCHITECTURE .....	40
4.3	TRAINING AND TESTING DATASET.....	41
4.4	NEURAL NETWORK IMPLEMENTATION, TRAINING, AND ESTIMATING.....	43
4.5	STUDY OF THE INFLUENCE OF SAMPLE SIZE.....	46
4.6	NEURAL NETWORK EVALUATION.....	47
4.7	STUDY OF FEATURE IMPORTANCE.....	51
<b>5</b>	<b>DEVELOPMENT OF THE DEEP NEURAL NETWORK MODEL FOR TRMM RAINFALL ESTIMATION.....</b>	<b>52</b>
5.1	PROBLEMS IN SPACEBORNE RADAR RAINFALL ESTIMATION.....	52
5.2	TWO-STAGE NEURAL NETWORK ARCHITECTURE .....	53
5.3	TRAINING AND VALIDATING DATASET.....	54
5.4	TWO-STAGE NEURAL NETWORK IMPLEMENTATION .....	55
5.4.1	<i>Ground radar validation neural network design</i> .....	55
5.4.2	<i>TRMM PR and ground radar data alignment design</i> .....	55
5.4.3	<i>The space radar neural network design</i> .....	57
5.4.4	<i>System validation</i> .....	59
5.6	LOCAL RAINFALL MAP GENERATION .....	65

<b>6</b>	<b>DEVELOPMENT OF THE DEEP NEURAL NETWORK MODEL FOR GPM RAINFALL ESTIMATION .....</b>	<b>72</b>
6.1	THE TWO-STAGE HYBRID SYSTEM ON GPM.....	72
6.2	THE DIFFERENCE BETWEEN GPM OBSERVATIONS AND TRMM OBSERVATIONS.....	73
6.3	GPM NEURAL NETWORK IMPLEMENTATION .....	75
6.4	SYSTEM VALIDATION.....	75
6.5	CASE STUDY .....	79
<b>7</b>	<b>SUMMARY AND FUTURE PLAN .....</b>	<b>81</b>
7.1	SUMMARY .....	81
7.2	FUTURE PLAN.....	82
7.2.1	<i>Performance Improvement</i> .....	83
7.2.2	<i>Performance Validation</i> .....	83
	<b>REFERENCES.....</b>	<b>84</b>

# 1 INTRODUCTION

## 1.1 Introduction

Rainfall estimation based on radar measurement has been a topic of study for five decades. In principle, rainfall on the ground can be represented by four-dimensional radar observations. However, the relation between rain rate and radar observation is difficult to express in a simple form. The empirical relations (Z-R) are not sufficient to capture the space-time variability of precipitation microphysics in terms of raindrop size distribution (DSD) and drop shapes. Because of this uncertainty, the empirical Z-R relation needs to be adaptively adjusted based on validation. Machine learning, a nonparametric method that can estimate ground rainfall directly from radar observations, has been demonstrated in prior research to yield rainfall rate estimations from radar measurements (Rongrui and Chandrasekar, 1997, Hongping et al., 2001, Stefano and Isabella, 2000). The artificial neural network (ANN) is a nonparametric machine learning method that can extract the hidden relationships between radar measurements and rainfall rate. The model of the relationship is created by a data-driven approach derived directly from a dataset consisting of radar observation (features) and rain gauge measurements (labels). The usefulness of the neural network-based rainfall estimation is subject to many factors such as the representativeness and sufficiency of the training dataset, the generalization capability of the network to new radar, seasonal changes, regional changes, and so on.

Satellite radar precipitation plays an important role in advancing our understanding of the earth's water and energy cycle and of how to forecast extreme weather events. The Tropical Rainfall Measuring Mission (TRMM) is an observing platform targeted to survey the earth's condition and climate change. It is a joint mission between the National Aeronautics and Space Administration (NASA) of the United States and the Japanese Aerospace Exploration Agency (JAXA). TRMM Precipitation Radar (PR) is the first observation platform capable of providing a high-resolution vertical profile of precipitation over the



tropics. The Global Precipitation Measurement (GPM) Core Observatory is the successor of TRMM and is expected to provide the next generation of global precipitation products through advanced observations from the GPM Microwave Imager and Dual-frequency Precipitation Radar (DPR). Compared to its predecessor, a key advancement of GPM is an extended capability to observe light rain, solid precipitation and the microphysical properties of precipitation particles. Through overlapping simultaneous measurements on Ka and Ku bands, the DPR is able to quantify precipitation particle distribution and quantitatively measure light rain and falling snow, which account for a significant part of precipitation, particularly in the middle and high latitudes. However, fundamental challenges exist in analyzing comparisons between the spaceborne radar and ground gauges. First, the spaceborne radar has a large horizontal resolution (usually several kilometers), which is much coarser than rain gauge spatial resolution which is depend on the rain gauge density. Second, available data pairs for comparison are scarce in single weather events because of the limited coincident overpasses. Unlike a rain gauge, ground radar has a similar resolution as spaceborne radar and can measure rainfall over a large spatial area to obtain more coincident samples to compare with ground rain gauges.

In machine learning, there is a sub-category called the deep learning which have deep graph architecture and many layers (Goodfellow et al. 2016). Unlike the conventional machine learning techniques requiring expertise in certain domains to extract meaningful features manually, deep learning is a representational learning method that allows the algorithm to automatically discover the useful features from raw data for object detection or classification by itself (LeCuu et al. 2015). Currently, deep learning techniques have been widely used in many aspects of society: computer vision, speech recognition, nature language processing. One of the typical deep learning algorithms is a deep neural network such as a multilayer perceptron (MLP) with more than two hidden layers. The additional layers can generate the abstract complex features from the simple features in the previous layer. The idea of deep learning has a long and rich history. Recently, it has become more useful because of

increasingly available training dataset and computation resources. This new technique has not been widely used in radar rainfall estimation. The traditional neural network based rainfall estimation algorithm usually has shallow structures and limited node numbers which may limit its performance.

In this proposal, a novel hybrid machine learning-based system on the Google TensorFlow platform consisting of two deep neural networks (DNNs) is investigated in order to improve rainfall estimation by building a relation between spaceborne radar observation and rain gauge measurement by using ground radar to bridge the gap between the spaceborne radar and the rain gauge. The first DNN model is trained from gauge measurements to ground radar rainfall estimations. The second DNN is trained from ground radar rainfall estimation to spaceborne radar rainfall estimation. Using the two DNN models, the entire system can generate a rainfall product by linking spaceborne radar observations to ground rain gauge measurements via ground radar observations.

## 1.2 Literature review

Rainfall estimation based on radar measurements has been studied for several decades. Generally, this research problem is addressed using parametric algorithms such as a Z-R relation to estimate rainfall from a radar reflectivity factor. This approach is also used in both TRMM-PR and GPM-DPR to estimate rain rate, where different Z-R relations are used based on the rainfall and the remote sensing instrument type. In recent years, a machine learning method including neural networks has been introduced to address this problem by taking into account the three-dimensional (3-D) structure of precipitation. Many approaches were tried with the goal of improving rainfall estimation using either ground radar or spaceborne measurements. These approaches are presented in this section.

### 1.2.1 Rainfall estimation based on radar measurements

A neural network technique for rainfall estimation based on ground radar observations was first introduced by Xiao and Chandrasekar (1997). This neural network is a multilayer perceptron with four input features and three hidden layers. This 3-layer MLP model was trained, validated, and tested using

data sets from several storm events collected in different seasons. One disadvantage of this MLP is the length of time it takes to train the network; thus, it is not best suited for long-term applications over months and years. However, this problem has been solved by the exponential growth of computation sources such as GPGPU from NVidia. Alternative attempts to use an adaptive radial basis function (RBF) neural network for radar rainfall estimation from horizontal reflectivity profile came later (Liu et al. 2001). Although the network was adaptively trained and tested on relatively large data sets, that adaptive RBF technique used a 9-point input vector sampled on a horizontal plane, increasing the complexity of such a neural network. This issue was investigated later (Li et al. 2003) and it was shown that reducing the size of the input vector will reduce the complexity of the designed network without having a significant effect on the estimation. Another attempt to improve the adaptive technique mentioned above was done by Li and Chandrasekar (2002). The improvement was done through a classification network, where a limit on the rainfall estimate was performed in order to not overestimate the rain rate. All above rainfall estimation algorithm are based on ground radar. The spaceborne radar rainfall estimation based on neural network has been developed by (Chandrasekar and Amin, 2008). The two stage hybrid system was created to estimate rainfall from space radar by using relation between ground radar and rain gauge as a bridge.

Other groups have done similar work. Stefano and Isabella (2000) estimated rainfall using neural networks in which they tried different NN architectures in order to get good rainfall estimates. They explored the capabilities of three artificial neural networks in order to identify and reproduce the functional relationship between Z and R. The three networks that were used were multilayer perceptron (MLP), Bayesian network, and radial basis function network. While these all got good results relative to Z-R relations, a flaw in their work was that all networks were not adaptive; each time they had new data they would need to retrain the network from the beginning. The other drawback is the all the neural network have one hidden layer and few hidden unit in this layer because of the limitation in computer

resources in year 2000. The first flaw was solved by Hongping et al. (2001), where training the neural network adaptively was begun. Another attempt to improve rainfall estimation using the neural network technique was done by Reinhard et al. (2007). This group added the height of precipitation as another input to the neural network, and they used the MLP network to do the estimation. This addition (the height of precipitation) was studied by Li et al. (2003), where it was shown that radar measurements up to 4km in height are enough to give good rainfall estimation using neural networks.

### 1.3 Research Question and Specific Aims

The main goal of this research is to address the following question: How much can we improve radar rainfall estimation by using deep neural networks to build relationship from ground/spaceborne radar observations and rain gauges measurements?

**Specific aim I:** Develop and implement a deep neural network for rainfall estimation from ground radar

**Specific aim II:** Develop and implement a two-stage hybrid system with two deep neural networks for rainfall estimation from the spaceborne radar

**Specific aim III:** generate local rainfall map from spaceborne radar observations based on hybrid system created in Specific aim II.

## 2 BACKGROUND

### 2.1 Radar observation and rainfall estimation

Rain rate is defined as the amount of raindrops that would fall during a short period of time if the rainfall intensity remained constant during this time. Normally, rain rate is expressed in terms of length per unit time (e.g., millimeters per hour, or inches per hour). The measurement of raindrop size distributions (DSD) established the basic relationship between radar reflectivity factor  $Z$  and rain rate  $R$ . One famous DSD model is an exponential expression with one parameter presented by Marshal and Palmer in 1948:

$$N(D) = N_0 e^{-\Lambda D} \quad 2.1$$

where  $D$  (in  $mm$ ) is the diameter of the raindrops,  $N(D)$  with unit ( $m^{-3}mm^{-1}$ ) is the raindrop size density (the number of raindrops per unit volume per unit size) between the  $D$  to  $(D + dD)$ ,  $N_0$  is the value of  $N(D)$  where  $D = 0$ , and  $\Lambda$  is the slope parameters given by the function of rain rate  $R$  (in  $mm\ hr^{-1}$ ). The value  $\Lambda = 41R^{-0.21}$  found by Marshal and Palmer is good over a wide range of rain rates, but other values were reported by (Battan 1973).

In conditions where the drop size diameter is much less than the wavelength, the radar reflectivity factor  $Z$  can be represented (following Atlas and Ulbrich, 1977) as:

$$Z = \int_0^{\infty} D^6 N(D) dD \quad 2.2$$

Using the Marshal and Palmer DSD,  $Z$  can be written as follows:

$$Z = \int_0^{\infty} D^6 N_0 e^{-\Lambda D} dD = 6! N_0 \Lambda^{-7} \quad 2.3$$

Considering equations (2.2) and (2.3), the relationship between radar reflectivity  $Z$  and rainfall rate  $R$  could be achieved in the power law  $Z$ - $R$  relation:

$$Z = aR^b$$

2.4

Where  $a$  and  $b$  are coefficients determined by the least squares fitting with large samples of observation of  $Z$  and  $R$  in different regions. For various data and models,  $a$  ranges from 127 to 505, while  $b$  ranges from 1.41 to 2.39. A very common pair of coefficients of  $a$  and  $b$  is  $a=200$  and  $b=1.6$  (Skolnik 2007).

## 2.2 Machine learning and neural networks

### 2.2.1 Machine learning basics

The definition of machine learning is a type of algorithm that has the ability to learn a mapping function from input to output from the historical dataset (See Fig 2.1).

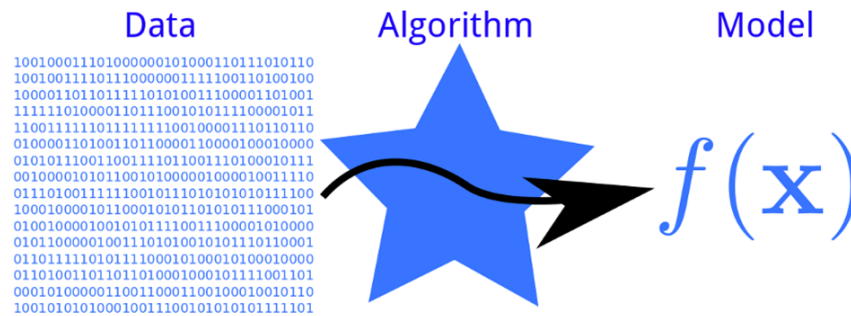


Fig 2.1. General concept of machine learning.

The keyword “learning” has been defined by Mitchell (1997) as “A computer program is said to learn from experience  $E$  with respect to some class of tasks  $T$  and performance measure  $P$ , if its performance at tasks in  $T$ , as measured by  $P$ , improves with experience  $E$ .” The task  $T$ , experience  $E$ , and performance  $P$  vary widely by type of problem. Task  $T$  is described as the way the machine learning system should process an example. The example refers a collection of features that have been quantitatively measured from some object or event that we want the machine learning system to process. The common tasks encountered in both academic and industry research could be classification, regression, transcription, and translation. The performance measure  $P$  is a quantitative measure of performance that evaluates the ability of certain machine learning algorithms. The choice of the performance measure is based on the specific task  $T$ . For example, a measure of accuracy is often used to estimate the performance in classification problems, while mean squared error is often used to

evaluate performance in regression problems. Experience  $E$  can categorize the machine learning algorithms into two types: supervised learning and unsupervised learning. In most cases, experience  $E$  will be learned from an existing dataset. The supervised learning algorithm can gain experience from a dataset containing both features and labels, then make a prediction or forecast based on that experience. An unsupervised learning algorithm will experience a dataset including features only, then understand the property of dataset structures.

Almost all machine learning algorithms can be separated into four types: a specification of a dataset, a loss/cost/error function to describe the performance of the system, an optimization procedure to minimize the loss function, and a model to map the input to the output. Typical machine learning algorithms contain linear regression, logistic regression, decision tree, random forest, support vector machine, and neural network.

### 2.2.2 Deep neural networks

From the first deep learning book written by Goodfellow et al (2016), a deep neural network, also called a feedforward network or multi-layer perceptron (MLPs), is one type of machine learning algorithm inspired by neuroscience. This technique has been widely applied in academia and industry such as computer vision, machine translation, neural language processing, and pattern recognition. The goal of deep neural networks is to find a mapping function  $f$  to approximate the task function  $f^*$ . In supervised learning, the task function is always given by

$$\hat{y} = f^*(x) \tag{2.5}$$

where the  $x$  is the data input,  $\hat{y}$  is the target from the historical dataset. And neural network defines a mapping from data input  $x$  to system output  $y$  by following the expression:

$$y = f(x; \theta) \tag{2.5}$$

where  $\theta$  is the parameter the system should learn to achieve the best approximation.

The information in the model flows from the input to the output. There are no feedback connections from the output of the model to itself. This machine learning algorithm is called a neural network because it represents the composition of many different functions. The acyclic graph model depicts the connection between different functions ( $f_1, f_2 \dots f_n$ ) to form a chain structure ( $f(x) = f_n(\dots f_2(f_1(x; \theta_1); \theta_2) \dots; \theta_n)$ ). The chain structures are called layers in neural networks. The length of the chain gives the depth of the model. The last layer of a neural network is called the output layer, which generates the output  $y$  for each training sample  $x$ . The other layers are called hidden layers because they are not directly specified by the training data and the learning algorithm must determine how to use these hidden layers to produce the desired output. Each hidden layer can be represented as a fixed length of vector. Each element of the vector is called a node, which can be interpreted as playing a role analogous to a neuron in the human neuron system. Every node in the layer receives input from many other nodes in the previous layer and computes its own value as output to the next layer.

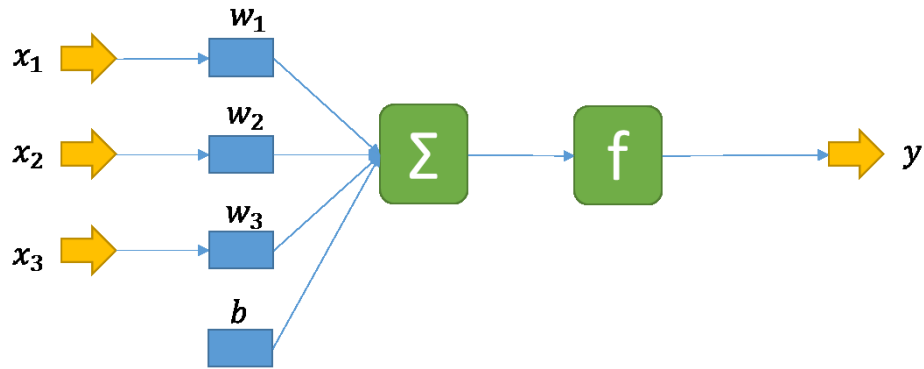
In the training process, the algorithm drives  $f(x)$  to match  $f^*(x)$  by minimizing the loss function between  $y$  and  $\hat{y}$  through the gradient descent algorithm. Commonly, to choose the loss function, the classification task will use cross entropy and the regression task can use the mean square error. The gradient descent technique uses the first order of derivative or partial derivative to minimize the target function  $l(\theta)$  by continuously reducing  $l(\theta)$  in moving  $\theta$  in small steps with the opposite sign of the derivative. In the neural network, the parameter  $\theta$  is always defined as weight  $w$  and bias  $b$ . The mapping function in a single layer could be written as a combination of the linear matrix multiplication and the non-linear activation function:

$$f(x; \theta) = f(x; w, b) = g(x * w + b) \quad 2.6$$

where function  $g$  is the activation function to bring the non-linearity between the input and output.

Figs 2.2, 2.3, and 2.4 show the forward calculation path in a node, a layer, and a fully connected neural network.

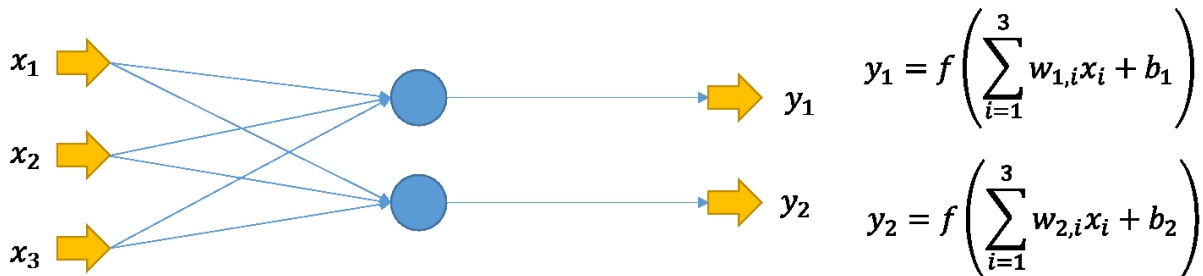




$$y = g \left( \sum_{i=1}^3 w_i x_i + b \right)$$

$x_1, x_2, x_3$  are the input variables;  $w_1, w_2, w_3$  are the weights, which can be changed in training process;  $b$  is the bias, which can also be changed in training process;  $g$  is the activation function to bring the non-linearity.

Fig 2.2. Single node diagram of feed-forward calculation in MLP.



$$y_j = f \left( \sum_{i=1}^3 w_{j,i} x_i + b_j \right) \quad j = 1, 2, 3 \dots$$

$$\mathbf{y} = f(\mathbf{w}\mathbf{x} + \mathbf{b})$$

$\mathbf{x}$  is the input vector and  $\mathbf{y}$  is the output vector;  $\mathbf{w}$  is the weight matrix in the single layer and  $\mathbf{b}$  is the bias vector;  $f$  is the activation function for bringing non-linearity

Fig 2.3. Single layer diagram of feed-forward calculation in MLP.

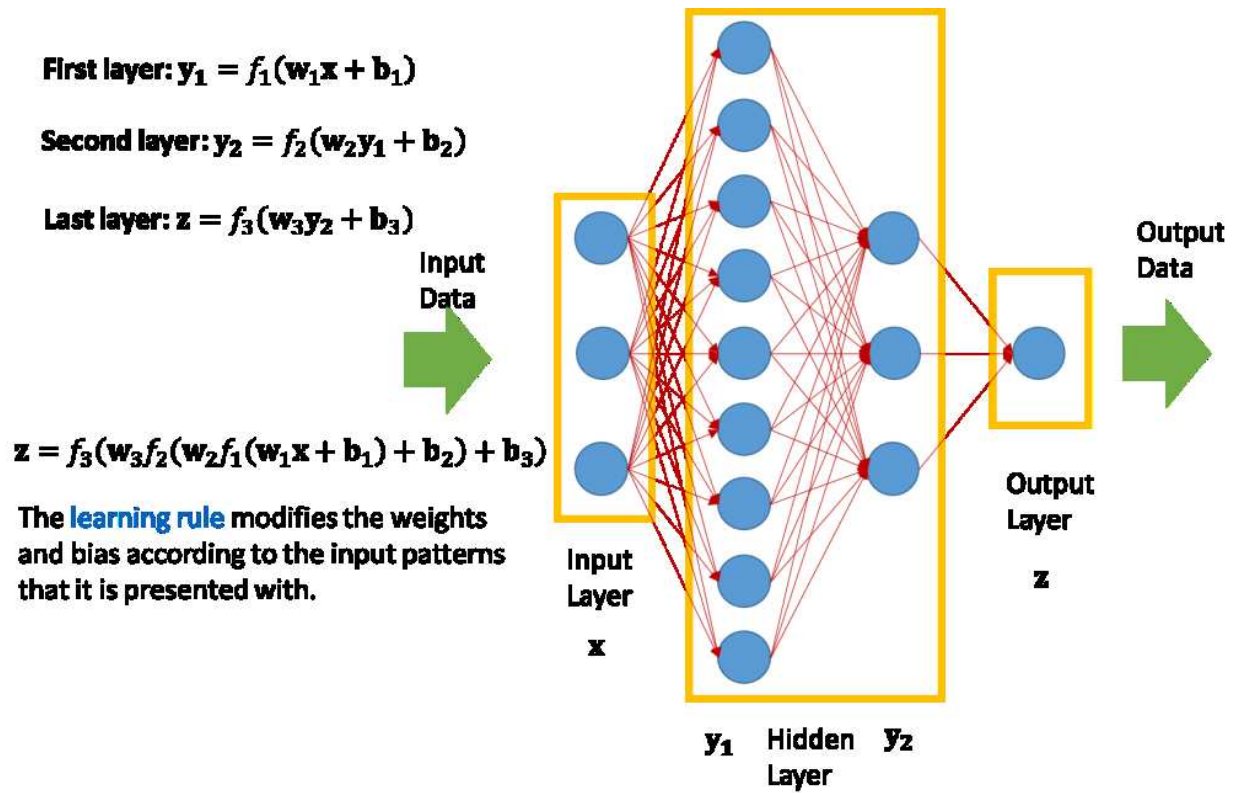


Fig 2.4. Full connected neural network diagram of feed-forward calculation in MLP.

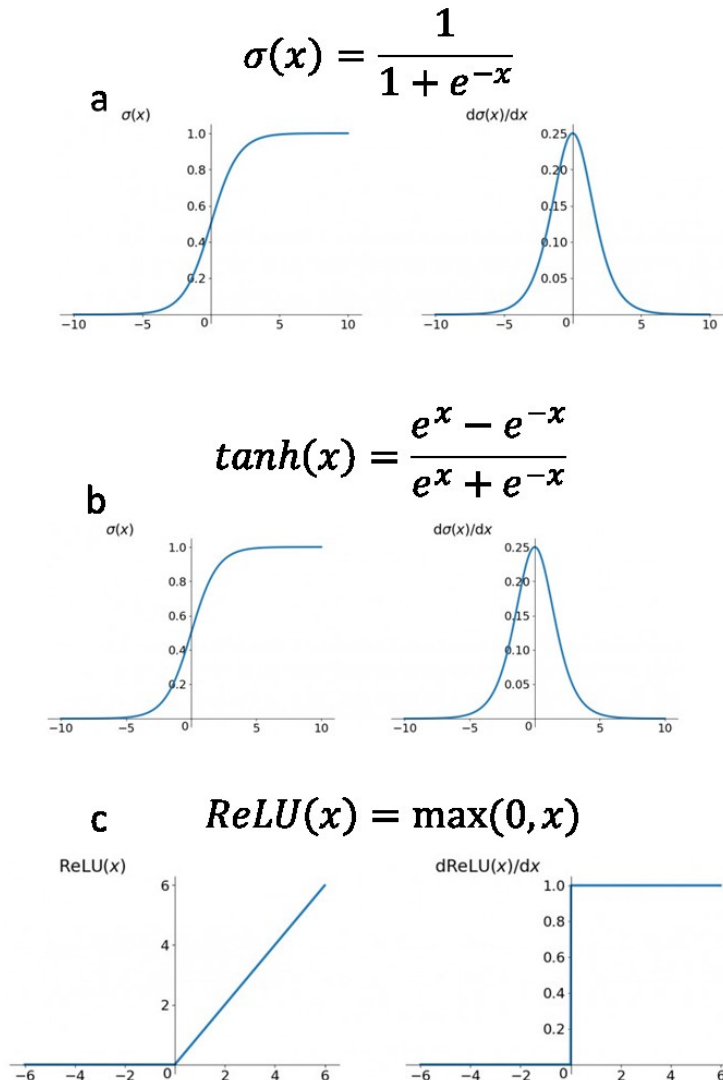


Fig. 2.5. Three non-linear activation functions a) sigmoid b) hyperbolic tangent c) rectifier. Three types of non-linear functions are normally used as activation function  $g$  in MLP: the

sigmoid function, hyperbolic tangent function, and rectifier function (ReLU). (Fig. 2.5). The non-linear activation function brings enough capacity to capture complex patterns of input. The sigmoid function was the most frequently used activation function in the early stage of neural network development. It calculate the derivative everywhere smoothly, but it caused problems in 1) gradient vanishing in training process, 2) non-zero-centered output, and 3) the exponential computing time cost. The hyperbolic tangent function solves the non-zero-centered output problem. However, in this system gradient vanishing still existed and the computing source requirement was also expensive. The final solution

came from using the rectifier function. The gradient will be vanish at  $x > 0$  and the computing speed is very fast. Most deep learning algorithms choose ReLU as the activation function for the default setting.

The initial information providing the input  $x$  propagates up to the nodes at each hidden layer and finally produce the output  $y$ . The entire process is called forward propagation. During training, forward propagation can continue computing the lost function. The back-propagation algorithm (Runmelhart et al., 1986a) allows the error information from the loss function to flow backwards through the network from output layer to input layer and gradients (partial derivatives) are computed based on the chain rule of calculus in the backward direction for all nodes in the neural network. These gradients are generally used by the gradient-based algorithm to update the parameters in the model for learning. In summary, the entire training process can be defined in four steps, as seen in Fig 2.6:

**Forward calculation.** Calculate the output  $y_i$  and  $z$  at each layer from the input  $x$

**Backward propagation:** Calculate the error  $E$  by using a different cost function from the ouput  $z$  and lable  $z^*$

**Backward propagation:** Calculate the gradient  $\frac{\partial E}{\partial w_{i,j}}$  using error  $E$  ,  $y_i$  and  $z$

**Gradient descent:** Calculate the updated weight  $w_{i,j}$  using the gradient  $\frac{\partial E}{\partial w_{i,j}}$

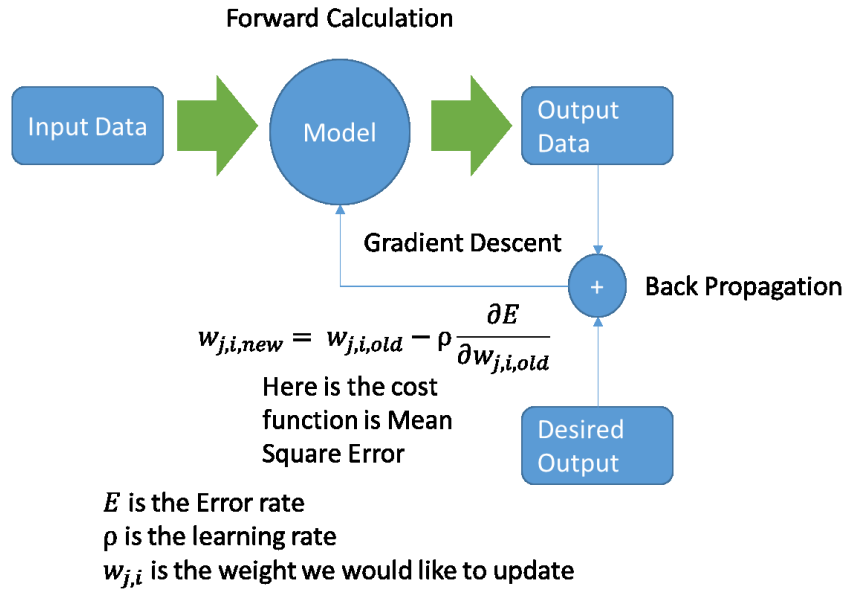


Fig 2.6. The entire training process in neural networks.

The computation in each layer is parallel. The training process of the neural network could be scaled on a general purpose graphic processing unit (GPGPU) or distributed system. Training efficiency is greatly improved by the parallel computing units.

### 2.3 Spaceborne and ground-based radar alignment methodology

To compare and quantify analysis of the measurement between space- and ground-based radar, the data points should be matched to each other in time and space. The difficulty of direct inter-comparisons is caused by the differences in viewing aspects between two observation systems, propagation paths, frequencies, resolution volume sizes, and time synchronization mismatches. And the geometric distortions in satellite observation induced by the spacecraft's movements and attitude perturbations make this problem even a more challenging task. A method developed by Bolen and Chandrasekar (2003) aligns measurements between these two systems by matching the variable resolution volume between the two systems and minimizing the effects of potential geometric distortion in space radar observations related to ground radar measurements. The application of this method has been applied to the TRMM PR and presents a decrease in error in the point-to-point inter-comparison between spaceborne and ground-based radar systems.

### 2.3.1 Alignment methodology

The alignment method of Bolen and Chandrasekar has three steps. First, resample both SR and GR data to the grid: 0.5x0.5km in the horizontal and 0.25km in the vertical. To minimize the time synchronization mismatch error, ground and space-based data were collected in similar time intervals with 2-3 minutes difference. The region of interest in the horizontal plane is defined as a small square window (normally 50km by 50km) in order to minimize the nonlinear spatial geometric distortion effect. Both ground- and space-based data are resampled to a similar Cartesian coordinate system (it could be satellite-centered or ground radar-centered) by using a nonspherical earth model (WGS-84 model).

Second, find a matched volume for gridded SR and GR data. Common reference point pairs are found between ground and space radar datasets. The procedure begins by finding the  $(x, y)$  coordinates of each SR beam location, derived from satellite ephemeris and PR pointing data in a horizontal plane at a nominal altitude. A three-dimensional set of reflectivity points is averaged in a linear scale at each of the beam locations in both horizontal and vertical directions in order to match the resolution volumes. The horizontal and vertical limits are taken as the maximum extent of either the SR or GR resolution at SR beam location. The average value of reflectivity computed in the volume is taken as the satellite measured reflectivity  $Z_M(\text{SR})$  at that  $(x, y)$  location. Next, the points that comprise the volume are translated in  $x, y,$  and  $z$  in the ground radar reflectivity dataset starting from the SR beam location  $(x, y)$ . The computed value is taken as the ground radar measured reflectivity  $Z_M(\text{GR})$  at location  $(x, y)$ .

Finally, do an alignment between the matched volumes. A ground-space radar point-pair is established by minimizing a cost function when a shift in  $(x, y, z)$  is found that minimizes a cost function based on the volume-averaged space and ground radar reflectivities,  $Z_M(\text{SR})$  and  $Z_M(\text{GR})$ , and the Euclidean distance of the volume shift. The cost function (CF) is normalized based on the expected error in reflectivity measurements and PR beam location measurements, and is given by

$$CF = \left( \frac{\sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}}{\sigma_d} \right) + \left| \frac{Z_m(PR) - Z_m(GR) + bias}{\sigma_m} \right| \quad 2.7$$

where  $\Delta x$ ,  $\Delta y$ , and  $\Delta z$  are the incremental volume shift distance,  $\sigma_d$  is the error related to satellite geolocation and beam-pointing error, and  $\sigma_m$  is the reflectivity measurement error. The bias is from the calibration between the two systems.

Once the set of point pairs has been found, a polynomial fit is used to determine the spatial alignment of the space radar image to the ground radar image. A polynomial of order  $N$  can be used to relate the coordinates of the contour of the distorted spaceborne image [ $x_c(SR)$ ,  $y_c(SR)$ ] to the corresponding ground radar contour coordinates [ $x_c(GR)$ ,  $y_c(GR)$ ] via

$$x(GR) = \sum_{i=0}^N \sum_{j=0}^{N-i} \alpha_{i,j} x(PR)^i y(PR)^j \quad 2.8$$

$$y(GR) = \sum_{i=0}^N \sum_{j=0}^{N-i} \beta_{i,j} x(PR)^i y(PR)^j \quad 2.9$$

The coefficients alpha and beta are determined and applied, based on the SR dataset, to map the space radar data to the ground radar data. Normally, a quadratic polynomial fitting is enough for this approximation.

## 3 OBSERVATION AND MEASUREMENT PLATFORM

### 3.1 Next Generation Weather Radar (NEXRAD)

#### 3.1.1 Introduction

Next generation weather radar, or NEXRAD, is a weather Doppler radar network operating across the United States that can provide information to ensure public safety from weather forecasting to military operations and emergency management. This worldwide network contains 160 high-resolution S-band Doppler weather radars covering the entire United States and selected international locations as shown in Fig. 3.1.

The formal name of the radar in the NEXRAD network is WSR-88D, which stands for Weather Surveillance Radar established in 1988 (D stands for Doppler). The radar network is operated by the National Weather Service (NWS), an agency of the National Oceanic and Atmospheric Administration (NOAA) of the United States Department of Commerce, the Federal Aviation Administration (FAA) within the Department of Transportation, and the U.S Air Force within the Department of Defense, and is directly supported by a radar operation center (ROC) in Norman, Oklahoma. The responsibility of ROC is to provide centralized meteorological, software, maintenance, and engineering support to keep the radars running smoothly and improve radar technology and capabilities to maintain peak performance in all radars.



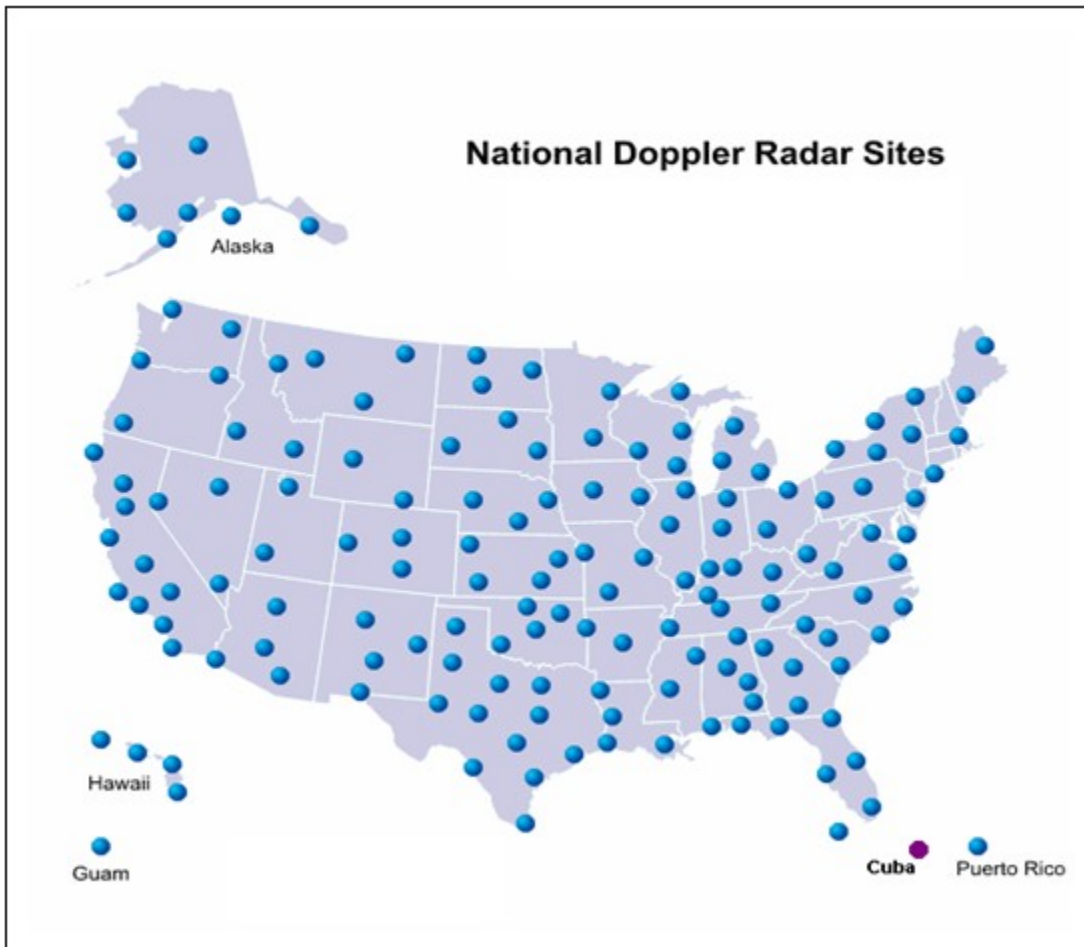


Figure 3.1. Locations of NEXRAD radars in the United States ([https://radar.weather.gov/index\\_lite.htm](https://radar.weather.gov/index_lite.htm)).  
 3.1.2 NEXRAD system overview

The NEXRAD system is able to generate three basic meteorological radar quantity measurements: radar reflectivity, radial velocity, and spectrum width. Meteorological products can be created from numerous analyses of these three basic quantities measurements.

NEXRAD has two major functional components: Radar data acquisition (RDA) and the radar products generator (RPG). (<https://www.roc.noaa.gov/WSR88D/Engineering/NEXRADTechInfo.aspx>). RDA is the information-gathering component of the system and produces the base moments (reflectivity, velocity, and spectrum width) and dual pol variables (differential reflectivity, correlation coefficient, and differential phase) representing the radar characteristics of the atmosphere. There are four primary components in RDA, including the antenna, the transmitter, the receiver, and the signal

processor. The RDA component generates the base monuments data by sampling the atmosphere, then transferring the measured information to the next component (RPG), where this information is temporarily stored to be used by the algorithms resident at the RPG.

The RPG takes the base monuments data and produces user-requested meteorological and hydrological products. The RPG uses multiple analysis programs (algorithms) to convert the base data into different meteorological and hydrological products. These products are stored and distributed by the PRG so that they can be easily accessed by the users. Table 3.1 shows some of the technical characteristics of NEXRAD system.

### 3.1.3 Operation

The antenna of the NEXRAD system continuously scans the atmosphere in two different modes based on the status of the atmosphere (precipitation/no precipitation). In the “no precipitation” or “clear air” mode, the scan of antenna completes 360° in the azimuth in ten minutes at five elevation angles (0.5° to 4.5°). In the “precipitation mode,” the scan of the antenna completes 360° in the azimuth in five minutes at 14 elevation angles (0.5° to 19.5°). Figure 3.2 shows the beam width and height above the earth surface as a function of the range for both precipitation and no precipitation modes.

Table 3.1. NEXRAD radar characteristics.

Type	S-band, center-fed, parabolic dish
Antenna size	8.54 m diameter; circular
Beam width	0.88 (at 3.0GHz)-0.96 deg (at 2.7 GHz)
Gain at 2850 MHz	45.8 dB (at 2.85 GHz)
Wave length	10.0 - 11.1 cm
Peak power	500 kW
Pulse width	1.57 - 4.5 ms
Polarization	Dual, Horizontal, Vertical
Range	460 km

Frequency	2.7-3.0 GHz
Sensitivity	10 dBZ
Normal Scan	+0.5 to +19.5
Range increment	250 m
Azimuth increment	1 deg

Source: ( <https://www.roc.noaa.gov/WSR88D/Engineering/NEXRADTechInfo.aspx>)

### 3.1.4 Ground validation rain gauge

A rain gauge is a type of instrument designed by meteorologists to gather and measure the amount of rain liquid over a set period of time. Millimeter is a common unit of rainfall measurement, and the unit of rain rate is millimeters per hour. A wide rain gauge network is for radar calibration and Z-R relation adjustments for the NEXRAD system. The gauges used in the NEXRAD rain gauge networks are commonly tipping bucket gauges. The rain gauge has limitations based on its characteristics. Collecting rain data during a strong storm event is nearly impossible and unreliable. Further, rain gauge measurements can only represent a local area.

The rain gauge dataset used in this research is gathered from precipitation measurement missions (PMM) ground validation (GV) programs, which provide ground-based validation for the PMM satellite observations (including TRMM and GPM). Three rain gauge networks at the Melbourne, Florida, NEXRAD site (KMLB 28.113° N, 80.654° W) were considered in this study. They were: South Florida Water Management District (SFL), St. Johns Water Management District (STJ), and Kennedy Space Center (KSC). Within a 100 km radius around the KMLB site, the three networks have 46, 99, and 33 rain gauges, respectively.

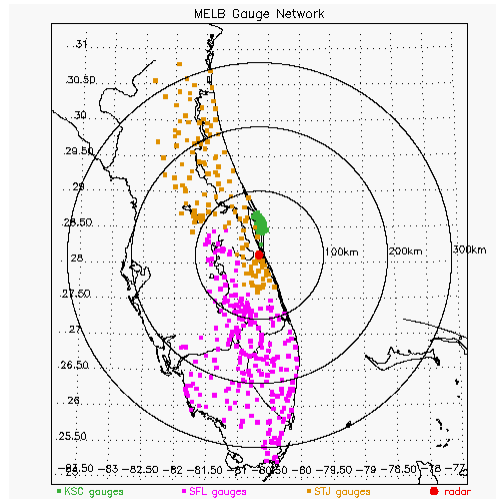


Figure 3.2. The three gauge networks of KLMB  
[\[http://trmmfc.gsfc.nasa.gov/trmm\\_gv/data/gv\\_maps/MELB\\_map.gif\]](http://trmmfc.gsfc.nasa.gov/trmm_gv/data/gv_maps/MELB_map.gif)

### 3.2 Tropical Rainfall Measuring Mission (TRMM)

The Tropical Rainfall Measuring Mission (TRMM) is a joint mission between the National Aeronautics and Space Administration (NASA) and the Japanese Aerospace Exploration Agency (JAXA) for weather study and climate research. The TRMM satellite was launched by an H-II rocket from Tanegashima Space Center on November 28, 1997, and ended on April 15, 2015. During its 17-year long operation, TRMM delivered data for monitoring and studying global tropical rainfall and lightning. The mainly TRMM observations were rain structure, rain rate, and distribution in both tropical and subtropical regions, observations that played a very important role in understanding the global climate and monitoring the environmental variation.

#### 3.2.1 TRMM instruments

As shown in Figure 3.3, the TRMM observatory contained five scientific instruments: the precipitation radar (PR), the TRMM microwave imager (TMI), the visible and infrared scanner (VIRS), the clouds and earth's radiant energy system (CERES), and the lightning imaging sensor (LIS). The PR, TMI, and VIRS are rainfall measurement instruments for generating tropical and subtropical rainfall measurements, rain profiles, and brightness temperature. The other two are for the earth observation

system program: CERES measured the Earth's radiation budget and LIS investigated the global distribution of lightning. Some of the characteristics of these five instruments are summarized in Table 3.2.

The precipitation radar was the first spaceborne quantitative rain radar instrument. Its operating frequency is 13.8GHz with a range resolution of 150m. Its main goal is to provide 3-D rainfall maps as well as to quantitative estimates of rainfall over land and ocean.

The VIRS is a cross-track scanning radiometer with five channels of different frequencies or wavelengths: 0.63, 1.6, 3.75, 10.8, and 12  $\mu\text{m}$ . The VIRS provides high-resolution observations about cloud coverage, cloud type, and cloud top temperature.

The TMI is a multichannel dual-polarized passive microwave radiometer with five different operating frequencies: 10.65, 19.35, 21.3, 37.0, 85.5GHz. All these channels operate at dual polarization (vertical and horizontal polarization) except the 21.3GHz channel, which operates at a single (vertical) polarization. The TMI provides reliable measurements of cloud liquid water, cloud ice, rain intensity, and rainfall types (stratiform, convective) over the ocean.

The LIS is an optical sensor operating at 0.7774  $\mu\text{m}$  that observes the distribution and variability of lightning over the earth. The LIS data can be combined with data from PR, TMI, and VIRS to study the relationship between lightning and rainfall.

The CERES is a scanning radiometer with operating wavelength ranges from 0.3 to 50  $\mu\text{m}$ . It measures the emitted and reflected radiative energy from the surface of the earth and from the atmosphere. The goal of CERES is to reduce uncertainties in predicting long-term changes in the climate.

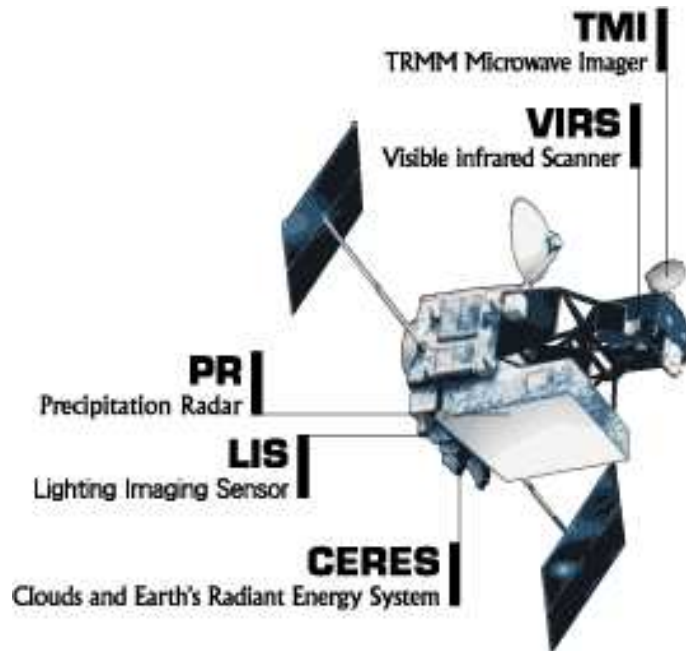


Figure 3.3. The deployed TRMM spacecraft.  
 [https://pmm.nasa.gov/sites/default/files/imageGallery/trmm\_sat.gif]

Table 3.2. TRMM instrument characteristics.

	PR	TMI	VIRS	LIS	CERES
Frequency/ Wavelength	13.8GHz	10.65, 19.35, 21.3, 37.0, 85.5 GHz	0.63, 1.6, 3.75, 10.8, and 12 $\mu\text{m}$	0.7774 $\mu\text{m}$	0.3 to 50 $\mu\text{m}$
Ground Resolution	5.0 km	11 km with 8 km at 85.5 GHz	2.5 km	4km	10km
Swath Width	260 km	880 km	830 km	600km	$\pm 82$ deg

Source: (TRMM Manual February 2006).

### 3.2.2 TRMM measurement levels

The TRMM measurements are processed by NASA and distributed to the user in the format of four levels from 0 to 3; each level has sublevels. In this section we summarize the products in the different levels related to the TRMM-PR instrument. Table 3.3 shows the definition of the TRMM products related to those instruments, and Figure 3.4 shows how these products are related.

#### LEVEL-0:

The unprocessed data from the TRMM-PR instrument with time ordered, quality checked, and redundancy eliminated.

LEVEL-1:

The PR Level-1 has two data products: 1B21 and 1C21. They are generated from the Level 0 data by processing with different algorithms.

1B21 is concerned with the calibrated received power. It has other information about geolocation, noise power, minimum echo flag (rain/no rain flag), storm height, etc. The received power is presented in three arrays: normal sample, surface oversample, and rain oversample. Each array is a 3-dimensional array whose dimensions are scan, ray, and range bin. In the case of a normal sample the power has 250m vertical spacing; otherwise it is 125m.

1C21 is related to radar reflectivity (Z Factor). It carries information about radar reflectivity without rain attenuation correction. The 1C21 product has the same format as 1B21. The radar reflectivity factor is calculated from the measured power in 1B21 without rain attenuation (TRMM-PR Manual, version 7).

Table 3.3. TRMM products.

Product No.	Name	Products
1B21	PR calibration Rain/No rain	Total received power, Noise Level Clutter contamination flag.
1C21	PR reflectivities	Profiled $Z_m$ (Radar reflectivity factors without rain attenuation correction)
2A21	Surface scattering coefficient $\sigma^0$	Path integrated attenuation (PIA) of $\sigma^0$ (in case of rain) and its reliability. Data based on $\sigma^0$ (ocean/land, in case of no rain)
2A23	PR qualitative	Detection of bright band, Bright band height, strength width, Rain type classification, Detection of shallow isolated

		rain. Output of rain/no rain, height of storm top.
2A25	PR profile	Range profiles of attenuation-corrected radar reflectivity factors, rainfall rate. Estimated near surface, and surface rainfall rates, and average rainfall rates between the two predefined altitude (2, 4 km)
3A25	Space-time average of radar products	Space-time averages of accumulations of 1C21, 2A21, 2A23 and 2A25.
3A26	Estimation of space-time rain rate statistics	Rain rate statistics over 5 degree by 5 degree 1 month space-time regions using a multiple thresholding technique.

Source: (TRMM-PR Manual, version 7).

#### LEVEL-2:

PR Level-2 has three products: 2A21, 2A23, and 2A25. They are generated from the corresponding Level 1 data.

The 2A21 product is mainly responsible for calculating the radar surface scattering coefficient. In the cases where rain exists, it computes the path integrated attenuation (PIA) with the surface as a target using the surface reference technique (SRT).

The 2A23 product is referred to as PR qualitative. It produces the rain flag. In the cases where rain is detected, it will detect the existence of the bright band and determine its height width and height. From there, it can classify the rain profile to be stratiform or convective.

The 2A25 product is known as PR profile. It produces a vertical profile of the rain rate estimate as well as the corrected version of the reflectivity. For ground validation purposes, the rain rate near the surface and the reflectivity near the surface are also given. In order to correct for attenuation, this algorithm uses a hybrid method of both the surface reference technique (SRT) and the Hitschfeld-Bordan method (HB) (Iguchi et al., 2000).



LEVEL-3:

The PR Level-3 has two main products: 3A25 and 3A26. They are generated from both the Level 1 and Level 2 data.

The 3A25 product is the space-time average of radar products. It contains monthly statistics of PR Level-1 and Level-2 products (rainfall, reflectivity, path-integrated attenuation, storm height, and bright band height). These statistics are performed using two resolutions:  $5^{\circ} \times 5^{\circ}$  and  $0.5^{\circ} \times 0.5^{\circ}$  latitude/longitude at 5 layers (2km, 4km, 6km, 10km, and 15km heights). The statistics include probability of occurrence, means and standard deviations, and histograms.

3A26: Estimation of space-time rain rate statistics over a  $5^{\circ} \times 5^{\circ}$  grid on a monthly basis at 3 layers (2km, 4km, and 6km heights). The statistics are the same as in 3A25 above (probability of occurrence, means and standard deviations, histograms and correlation coefficients).

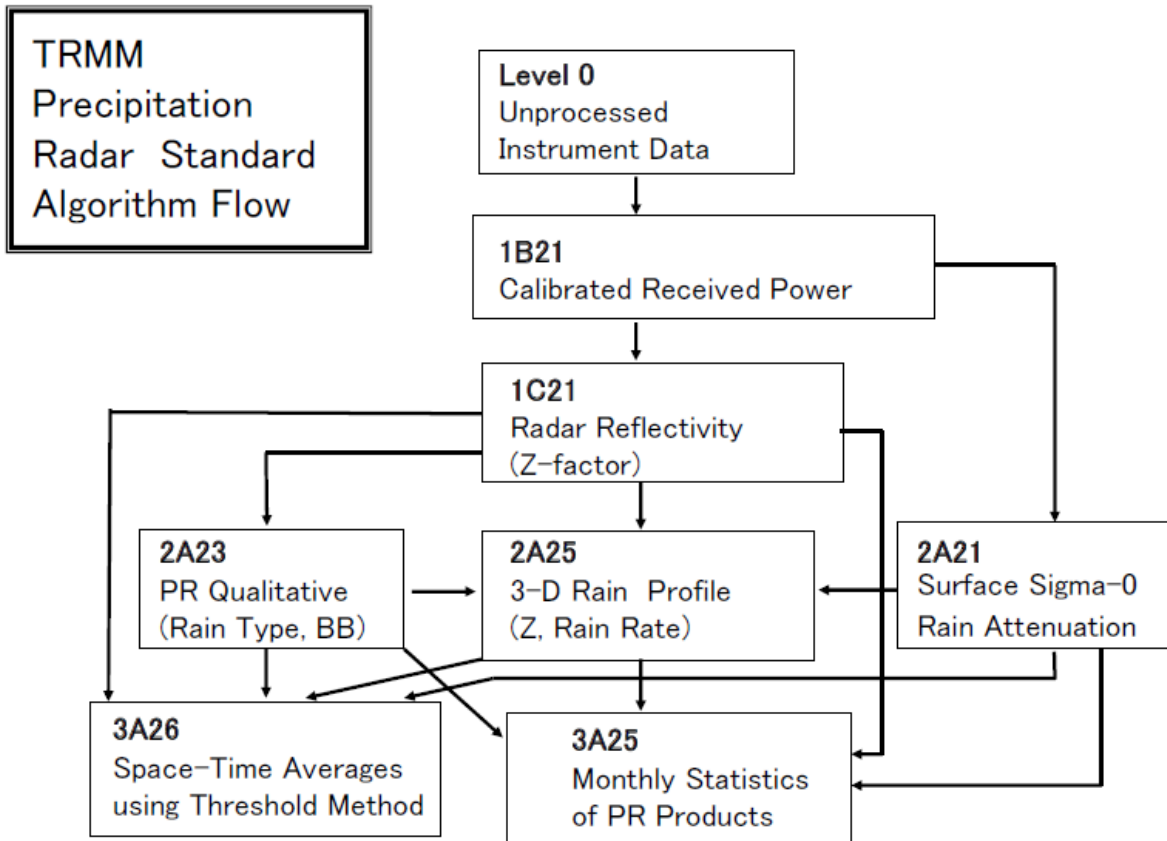


Figure 3.4. TRMM-PR algorithm flow diagram (TRMM-PR Manual, Version 7).

### 3.2.3 TRMM precipitation radar (PR) overview.

The TRMM PR is the first spaceborne rain radar with the ability to observe vertical distributions of rain directly. TRMM PR works at 13.8GHz and provides quantitative rainfall estimation over land as well as ocean. The PR major design and performance parameters are shown in Table 3.4 (TRMM Data users handbook), and its observation concept is shown in Figure 3.5. The PR antenna beam scans in the cross-track direction over  $\pm 17^\circ$  to give a 220km swath. The antenna beam width of the PR is  $0.71^\circ$  and there are 49 angle bins within the scanning angle of  $\pm 17^\circ$ . The horizontal resolution is around 4.3km at nadir and 5km at the scan edge when TRMM is in a 350km orbit. The range resolution of PR is 250m. The minimum detectable Z is improved from 23.3dBZ to 20.8dBZ by increasing transmit power and decreasing noise figure.

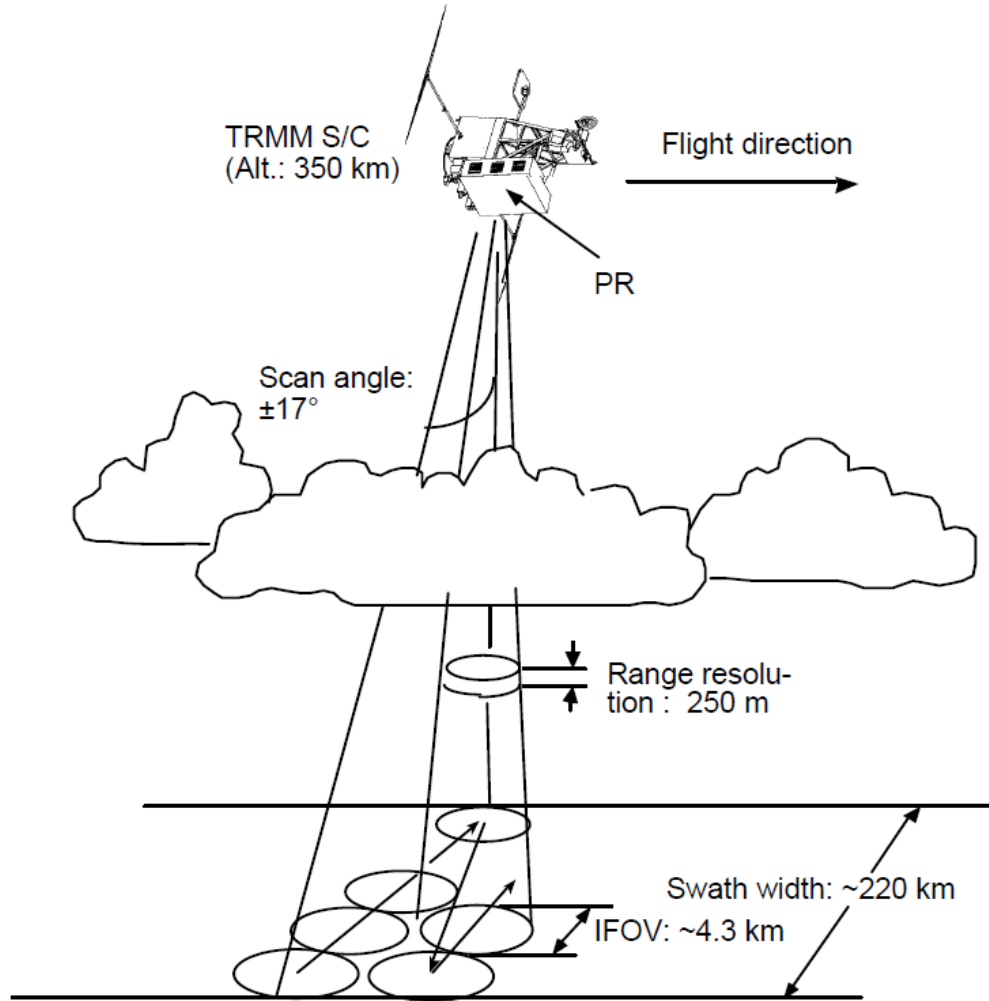


Figure 3.5. The observation concept of the PR (adopted from TRMM PR Manual V7).

Table 3.4. Major parameters of TRMM-PR.

Frequency Swath Width	13.8 GHz About 215 km
Observable Range	Over 15 km
Range Resolution	250 km
Horizontal Resolution	4.3 km (nadir)
Weight	465 kg
Power	250 watts
Beam Width	$0.71^\circ \times 0.71^\circ$
Aperture	2.0 m x 2.0 m
Scan Angle	$\pm 17^\circ$
Gain	About 47.4 dB

Peak Power	Over 700 W
------------	------------

Source: TRMM Data User Handbook, 2006

### 3.2.4 TRMM-PR rain rate estimation

The TRMM-PR 2A25 provides a vertical profile of the rainfall estimation by using a simple Ze-R relation from the vertical profile of radar reflectivity Ze. The Ze is the measured radar reflectivity factor (Zm) with attenuation correction. In this research, we are only concerned with the rainfall estimate near the surface. Therefore, only measurements of the rain vertical profile at the lowest altitude will be considered.

There are two DSD model assumptions in the TRMM-PR product: one is for stratiform rain and the other is for convective rain. Two models were made from a collection of Ze–R relations measured near the ocean from widely distributed locations around the world. Typical Z-R relations with parameters were found to be (Iguichi et al. 2000).

Stratiform:

$$Z = 300R^{1.38}$$

Convective:

$$Z = 185R^{1.43}$$

Once the DSD model has been determined, the parameters in the k-Ze and Ze-R relations can be calculated for rain and snow at different temperatures and mixing ratios, where these parameters are functions of the rain type, existence of the bright band, the heights of the 0°C isotherm, and the storm top.

In the case of stratiform rain with bright band, the stratiform DSD model with a vertical profile model is used to calculate the k-Ze and Ze-R coefficients at 5 levels, as shown in Figure 3.6. The five

levels include three for snow-water mixture (A, B, C) and two for water drops (D, E).

Levels A, B, and C are defined such that the snow-water mixtures have fractional water contents of 1.1%, 1.7%, and 17%, respectively. Level D was defined where water drops have a temperatures of 0°C, while level E was defined at 20°C. Level C is the bright band peak, level B was taken to be 500m above level C, while level D was taken to be 500m below level C, and level A was considered as the storm top.

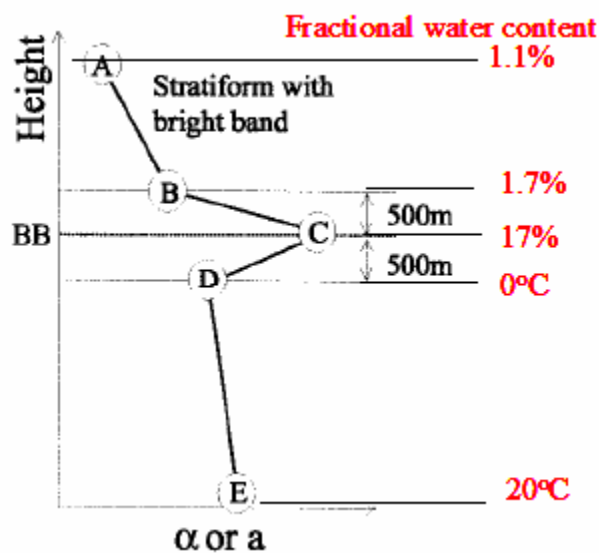


Figure 3.6. Schematic presentation of the profiles for stratiform profile with bright band detected (adopted from Iguchi et al., 2000).

Other values of the k-Ze and Ze-R coefficients at intermediate levels are calculated by linear interpolation from temperature by assuming a lapse rate of 5°C/Km.

In the cases of stratiform rain without bright band detected or convective rain type, the DSD model shown in Figure 3.7 is used to calculate the k-Ze and Ze-R parameters. A similar DSD model is used when the hydrometeors are assumed to be at 0°C in 750m on either side of the of 0°C isotherm.

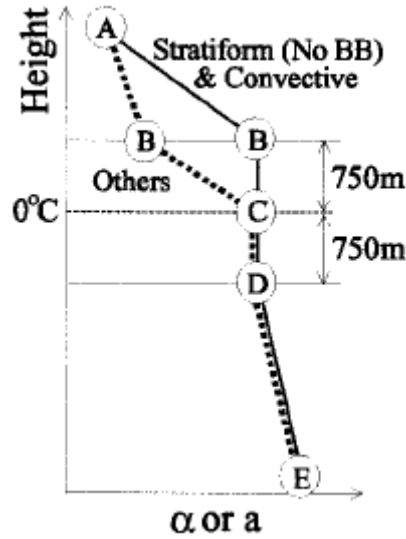


Fig. 3.7. Schematic presentation of the profiles for stratiform profile with bright band detected (adopted from Iguchi et al., 2000).

e

Table 3.5. Initial k-Z and Z-R relationship parameters

Parameter		A	B	C	D (0°C water)	20°C water
Stratiform	$\alpha$	0.000 086 1	0.000 108 4	0.000 414 2	0.000 282 2	0.000 285 1
	$\beta$	0.792 30	0.792 30	0.792 30	0.792 30	0.792 30
	$a$	0.013 98	0.012 63	0.004 521	0.020 10	0.022 82
	$b$	0.7729	0.7644	0.7288	0.6917	0.6727
	$a''$	250.8	304.6	1649.3	283.9	275.7
	$b''$	1.294	1.308	1.372	1.446	1.487
Convective	$\alpha$	0.000 127 3	0.000 410 9	0.000 410 9	0.000 410 9	0.000 417 2
	$\beta$	0.7713	0.7713	0.7713	0.7713	0.7713
	$a$	0.020 27	0.034 84	0.034 84	0.034 84	0.040 24
	$b$	0.7556	0.6619	0.6619	0.6619	0.6434
	$a''$	174.1	159.5	159.5	159.5	147.5
	$b''$	1.323	1.511	1.511	1.511	1.554
Others	$\alpha$	0.000 127 3	0.000 159 8	0.000 410 9	0.000 410 9	0.000 417 2
	$\beta$	0.7713	0.7713	0.7713	0.7713	0.7713
	$a$	0.020 27	0.018 71	0.034 84	0.034 84	0.040 24
	$b$	0.7556	0.7458	0.6619	0.6619	0.6434
	$a''$	174.1	207.4	159.5	159.5	147.5
	$b''$	1.323	1.341	1.511	1.511	1.554

Source: Iguchi et al., 2000

### 3.3 Global precipitation measurement (GPM)

Following the success of the TRMM, the next mission was the global precipitation measurement (GPM) for measuring global precipitation more accurately and frequently for elucidating changes and water circulation. This is also a joint mission between the Japan Aerospace Exploration Agency (JAXA) and the US National Aeronautics and Space Administration (NASA). The GPM core satellite was launched by H-IIA from the Tanegashima Space Center on February 28, 2014. The GPM project achieved rainfall measurements every 3 hours by cooperation between one core satellite with the DPR and the GMI instrument and other sub-satellites with a microwave imager. Figure 3.8 provides an overview of the GPM Project.

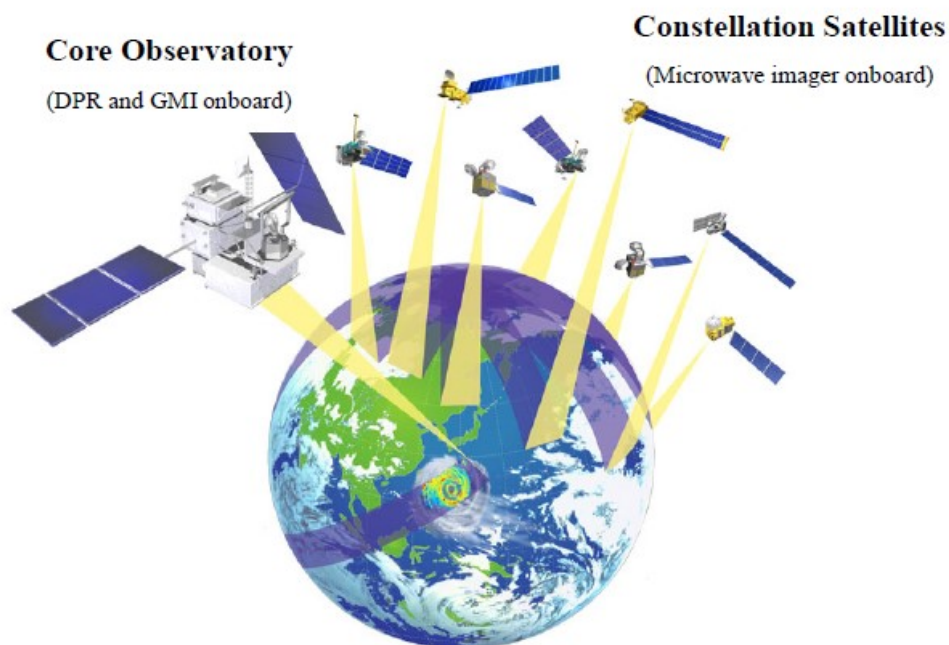


Fig. 3.8: Overview of GPM project. ([global.jaxa.jp/projects/sat/gpm/](http://global.jaxa.jp/projects/sat/gpm/))

#### 3.3.1 GPM instrument

The GPM core observatory carries two scientific instruments: the dual-frequency precipitation radar (DPR) and a GPM microwave imager (GMI). The DPR performs measurements in two different frequencies that can provide additional information about the meteorological target. Its operating frequency is 13.8GHz (Ka band) with 250m range resolution and 35.55GHz (Ku band) with 250m/500m

range resolution. Its main goal is to measure 3-D rainfall structure and to observe snow from space for the first time. The GMI is a multichannel dual-polarized passive microwave radiometer with 13 microwave channels in 7 different operating frequencies ranging from 10GHz to 183GHz. Comparing the TRMM TMI, GMIs have 9 channels similar to the TMI and 4 high-frequency, millimeter-wave channels in 166GHz and 183GHz. The GMI provides reliable measurements of the cloud liquid water, cloud ice, rain intensity, and rainfall types (stratiform, convective) over the ocean.

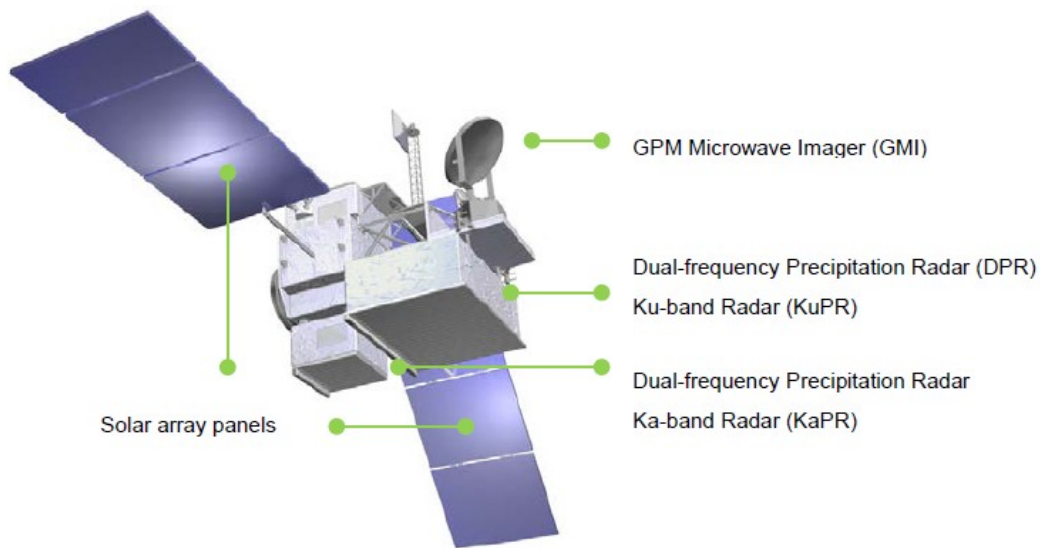


Fig.3.9. Instruments on the GPM core observatory. (GPM Data Utilization Handbook, 2017)

### 3.3.2 GPM\_DPR overview

The DPR's major characteristics are shown in Table 3.6 and its observation geometry is shown in Figure 3.10. The Ku-band radar have approximately the same specifications as the TRMM PR, with some improvements. Since the Ku-band PR can detect moderate to heavier rainfall but cannot observed weak rainfall and snowfall, the Ka-band PR is specifically targeted on those cases by providing more sensitive observations. By combining the Ku- and Ka-band PR together, the DPR will enable accurate measurement of a range of precipitation from heavy rainfall in the tropics to weak snowfall in high latitudes areas.

Table 3.6. GPM DPR characteristics.

Main characteristics	Ku band	Ka band



Radar Type	Active Phased Array Radar	
Antenna	Slotted Waveguide Antenna	
Frequency	13.6GHz	35.55GHz
Peak Transmit Power	>1000W	>140W
Range Resolution	250m	250m/500m
Pulse Width	1.67ms	1.67ms/3.34ms
Horizontal Resolution	5.2km (at nadir)	
Beam Width	0.71 degrees +/- 0.02 digress (at nadir)	
Pulse Repetition Frequency	Variable	
Swath Width	245 km	125km
Scan Interval	0.7sec	
Observable Range	19km to surface	
Minimum Detectable Rain Rate	0.5mm.hr	0.2mm/hr
Receiver Power Accuracy	+/- 1dB	
Beam-matching Error	<1000m	
Design Life	3 years and 2 months	
Data Rate	<109kpbs	<81kpbs
Mass	<472kg	<336kg
Power Consumption	<446W	<344W
Dimensions	2.5m x 2.4m x 0.6m	1.4m x 1.2m x 0.8m

Source: (GPM Data Utilization Handbook, 2017)

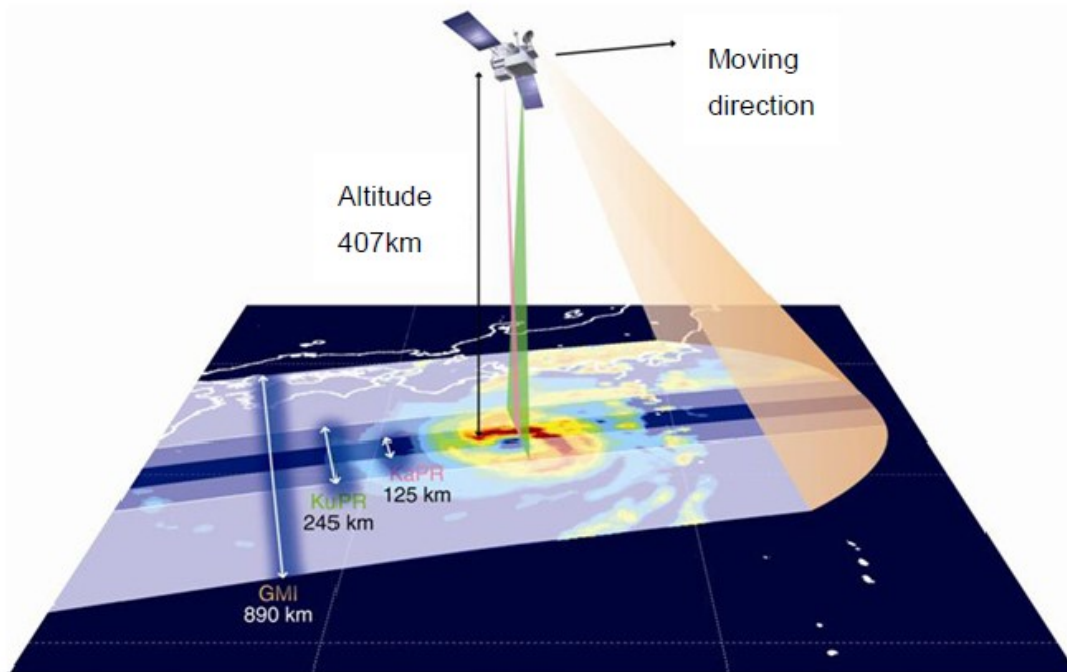


Fig. 3.10. GPM observation geometry. (GPM Data Utilization Handbook, 2017)

The Ku-band PR antenna beam scans in the cross-track direction over  $\pm 17^\circ$  to give a 245km swath, while the Ka-band PR has a narrower swath width of about 120km formed by a  $\pm 8^\circ$  scan. The antenna beam width is  $0.71^\circ$  for both Ku and Ka bands. The horizontal resolution is around 5km at a 400km orbit altitude. The range resolution is 250m in both KuPR and KaPR, but the KaPR has another range resolution that equals 500m. The minimum detectable reflectivity in KuPR and KaPR is 18 dBZ and 12 dBZ, which means the minimum detectable rain rate is 0.5mm/hr and 0.2mm/hr, respectively.

### 3.3.3 GPM DPR level data

GPM provides standard products and near-real-time products with different levels. The standard products are verified by JAXA and NASA so that they can be used in academic research, while the near-real-time products are used for weather forecasting by meteorological organizations. In this research, we are more interested in standard products that give historical rainfall representation. Table 3.7 shows the definition of the GPM DPR standard products, and Figure 3.11 shows the diagram of these products.

Table 3.7: GPM DPR product and combined product.

Group	Product
DPR product	KuPR Level 1
	KaPR Level 1
	KuPR Level 2
	KaPR Level 2
	DPR level 2
	DPR level 3
DPR/GMI combined product	DPR/GMI level 2
	DPR/GMI level 3

Source: (GPM Data Utilization Handbook, 2017)

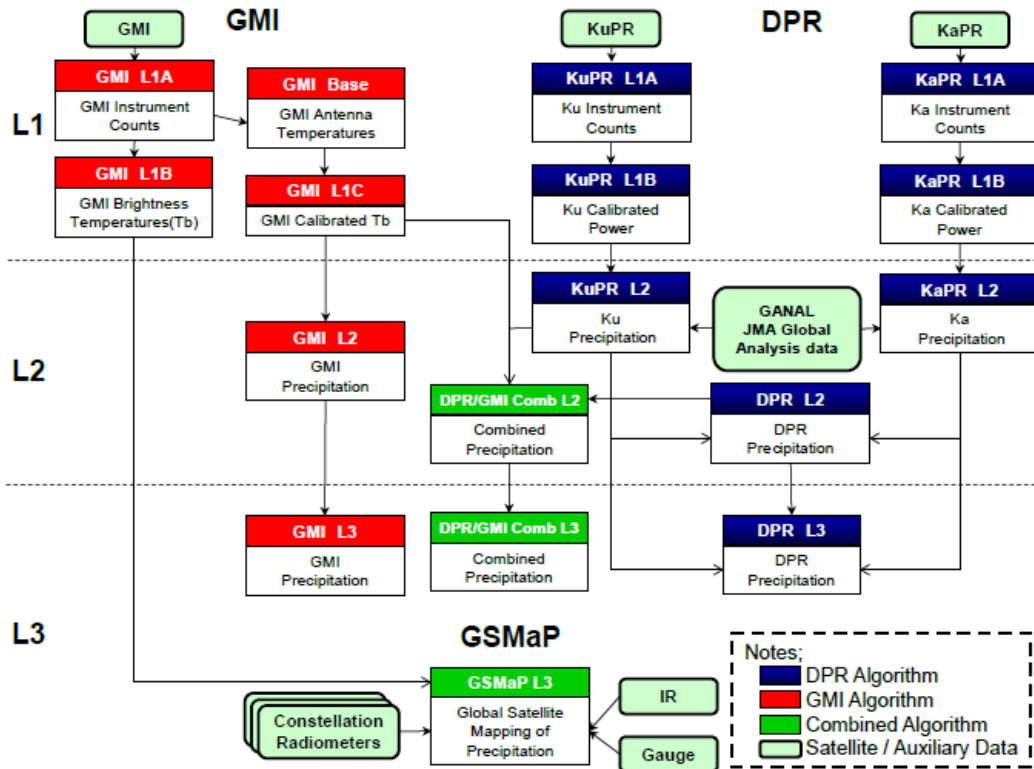


Fig. 3.11. GPM DPR products algorithm diagram. (GPM Data Utilization Handbook, 2017)

The DPR level 1 products have four sub-types: the Ku-level 1A product, the Ku-level 1B product, the Ka-level 1A product, and the Ka-level 1B product. The 1A product is generated by the 1A algorithm,

which converts the satellite ancillary data to the instrument counts with data-cutting and a telemetries-checking process. The 1B algorithm produces the 1B product from the 1A product by system noise removal, radiometric correction, and geometrical calculation. The 1B product is the calibrated power.

The DPR level 2 algorithm estimates rainfall intensities, types, top heights, and bright band heights based on the received power profiles from KuPR and KaPR. This algorithm inherits its information from the TRMM PR standard algorithm, adding the information from the dual-frequency observation. Three level 2 products (KuPR, KaPR and DPR) are generated from this algorithm. The Ku level 2 product only contains rainfall estimation from Ku observations in the 245km swath width, while the Ka level 2 product only uses Ka observations in a 125km-wide swath. The DPR level 2 product includes both Ku and Ka observations and covers the 250km-wide swath using extrapolation techniques.

The DPR level 3 algorithm conducts spatial and temporal statistical processing to create a 3-D grid rainfall map from the level 2 products. The level 3 daily rainfall average products have two spatial resolutions, 0.1 degree in text format and 0.25 degree in HDF format. The level 3 month average rainfall product is only in 0.25 degree grids resolution.

#### 3.3.4 GPM rainfall estimation

The GPM DPR has three type of rainfall estimations based on two frequencies: Ku band, Ka band. The Ku-only and Ka-only rainfall estimation is retrieved by the single-frequency algorithm. This single-frequency algorithm 2.2 was mentioned in the first section of chapter 2. In this algorithm, the raindrop size density  $N(D)$  is considered to only be involved with a single parameter, the diameter  $D$ . Thus, the reflectivity  $Z$  and raindrop diameter  $D$  would be in one-to-one correspondence in this situation. Once the reflectivity  $Z$  is obtained from the measured radar power, the rainfall rain  $R$  could be calculated based on this one-to-one relation. The TRMM PR also uses this principle to estimate the rain rate with single frequency radar.

However, the  $N(D)$  cannot be simply characterized by a single diameter by nature. A rainfall estimation based on single-frequency radar measurements would lead to many errors and biases. The dual-band algorithm consider both the Ku and Ka bands to overcome the drawback of a single-frequency algorithm. A two-parameter rainfall retrieval method is developed to represent the variation of  $N(D)$  with two different type of frequencies. If one wavelength is smaller than or comparable to the average raindrop size in radar with a dual wavelength, the corresponding backscattering cross-section could be derived from Rayleigh scattering and the reflectivity at this wavelength will be different from the reflectivity at other wavelengths. This situation enables us to use two parameters to characterize the DSD model and improve the estimation accuracy. This is the principle for GPM DPR rainfall estimation (GPM/DPR Level-2 Algorithm Theoretical Basis Document).

## 4 DEVELOPMENT OF DEEP NEURAL NETWORK MODEL FOR GROUND RADAR RAINFALL ESTIMATION

### 4.1 Neural Network in Rainfall Estimation

As mentioned in Chapter 2, the neural network can create non-linearity mapping from three-dimensional radar observations to ground rain gauge measurements by creating a group of interconnected artificial neurons with a simple activate function. This is a non-parametric method. However, the word 'non-parametric' does not indicate that the network doesn't have any parameters. Instead, the number and nature of the parameters are flexible and not fixed in advance. This method does not depend on any assumptions regarding data distribution. As a system, the neural network can be trained on particular inputs and specific outputs by updating the weights and biases in the neural network to represent the relationship between the input and the output.

Unlike a traditional parametric reflectivity-rainfall approach, a neural networks-based technique will derive the relationship between the radar reflectivity and rain gauge measurement directly. The non-linearity activation function in the neural network can increase the capacity to capture this complicated Z-R relation. The input of the neural networks can be used only with reflectivity Z or the combination between reflectivity Z, differential reflectivity Zdr, and differential phase Kdp, while the output could be the rain rate R from measurement of the rain gauge. After appropriate training and cross-validating, the system could provide higher accuracy in rainfall estimation and more generalization to adapt to the variety in locations and seasons.

Neural network techniques are widely used in radar system for snow estimation, rainfall estimation, rain detection, and rain type classification. Xiao and Chandrasekar (1996) first introduced the three layers of MLP (multiplayer perceptron) to apply to snowfall estimation. Rainfall estimation and

detection algorithm based on neural MLP were introduced by a similar group later on (Xiao and Chandrasekar 1997, 1998). An alternative precipitation estimation model was demonstrated by Liu and Chandrasekar (2001, 2003) using an adaptive radial basis function (RBF) neural network. The application to rain type classification was proposed in Liu and Chandrasekar 2003 using an unsupervised learning method named self-organizing maps.

#### 4.2 Neural Network Architecture

In this research, we built a deep neural network system for ground radar rainfall estimation by linking ground radar (GR) observations and rain gauge (RG) measurements. This system has the ability to convert radar reflectivity to a radar rainfall product by taking into account the historical dataset of GR observations and RG measurements in a similar region. The brief system diagram is shown in Fig-4.1

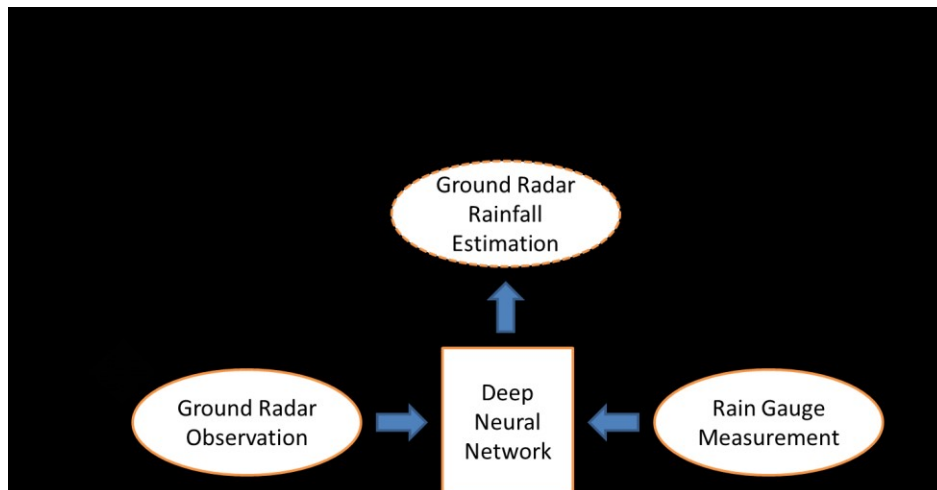


Fig. 4.1. The diagram of neural network based GR rainfall estimation system.

The neural network we chose for the ground radar training is the MLP including one input layer, one output layer, and one or more hidden layers in the center. The number of hidden layers and nodes in each layer are determined by the grid search method in the cross-validation process. The rectifier function was selected as the activation function between each layer to provide non-linearity for mimicking the target function between input and output data.

In this research, the output value is rain rate, which is a continuous value. Thus, the whole problem could be defined as a regression. Commonly, the loss function used in regression problem is

the mean squared error (MSE). We wish to minimize this loss function in the neural network by the gradient descent optimizer. The entire architecture of the MLP can be seen in Fig 4.2.

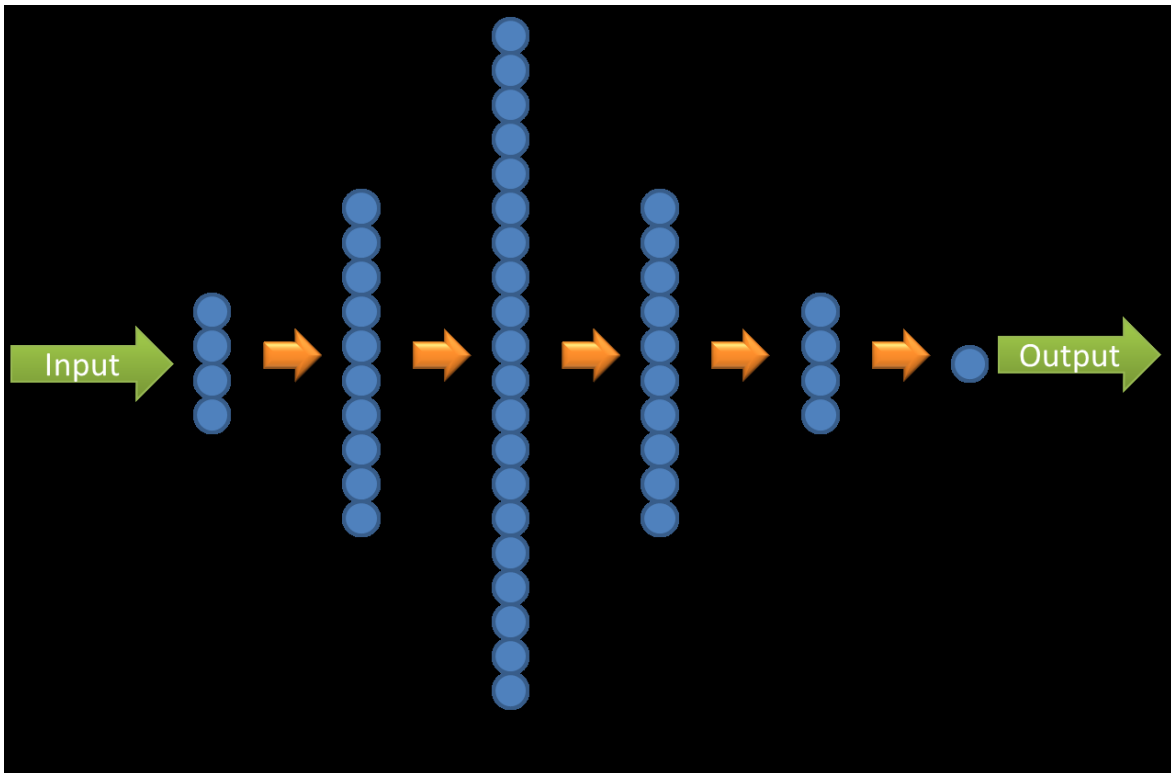


Fig. 4.2. The MLP architecture we used to establish the relationship between radar observations and rain gauge measurements.

### 4.3 Training and Testing Dataset

Fig. 4.1 shows that the input feature is the GR reflectivity vertical profile and the input label is the RG measurement. Both radar and rain gauge data for this study were collected during storm events during 2005, 2009, and 2014 over Melbourne, Florida. This dataset includes a large number of convective and stratiform rainfall cases. The GR data, which was obtained from the Melbourne NEXRAD (KMLB) radar, was first mapped to a constant altitude plan position indicator (CAPPI) at multiple vertical levels with a spatial resolution of 1km x 1km, as shown in Fig. 4.3. The lowest level of the CAPPI scans is 1km and the highest level is 4km. The vertical resolution of the CAPPI scan was chosen to be 1km. Around KMLB radar, direct rainfall measurements are from the gauge stations at three districts in 100km: 33 at the Kennedy Space Center (KSC), 46 at the South Florida Water Management District (SFL),



and 99 at the St. Johns Water Management District (STJ). Only data within 100km was considered in this research because the vertical extent of the GR beam is about 1.8km at 100km, and going farther would cause coarse sampling for ground radar measurements.

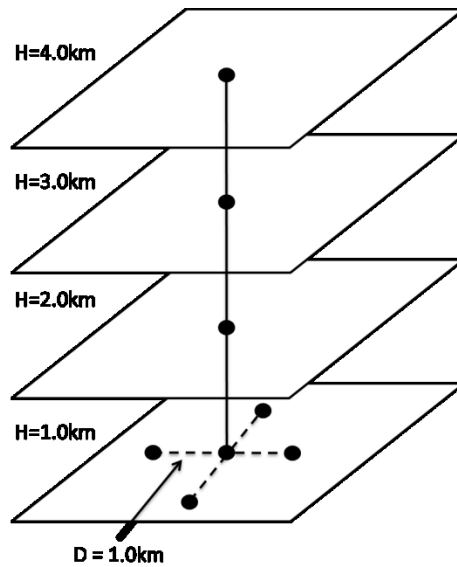


Fig. 4.3. GR data in CAPPI format.

In order to train and test the MLP, the GR and RG datasets were matched in time and space to form feature-label pairs. Only the location with RG was considered as a valid place to collect the GR vertical profile. In the time-matching process, the GR time stamp was retrieved from the GR data file name and the RG time stamp was obtained from rainfall measurements with a 1 min interval. The minimum time resolution in NEXRAD is 5 minutes. A 5 min rainfall rate was generated corresponding to a single GR vertical file by averaging the RG rainfall in that duration. Finally, the pairs between GR reflectivity profile and 5 min rain rate were created by the process shown in Fig. 4.4

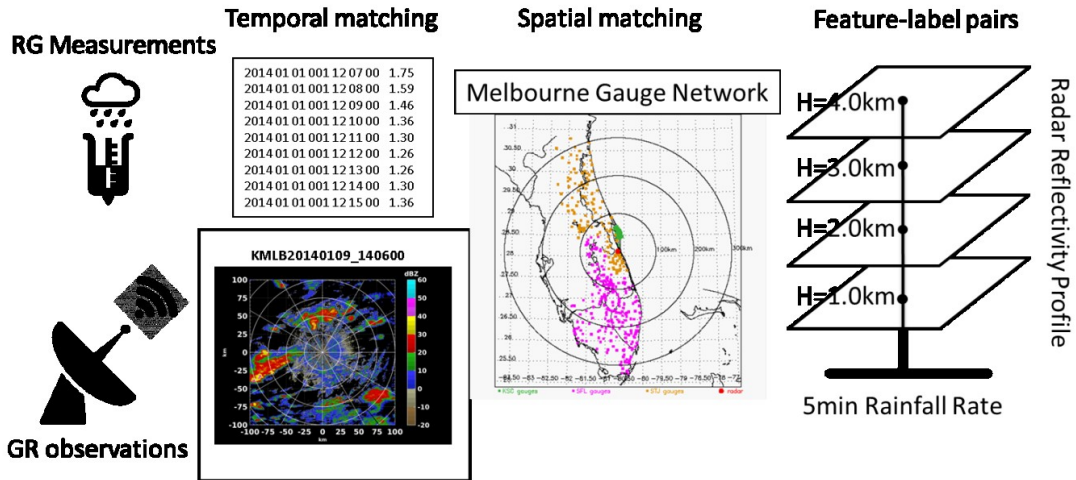


Fig 4.4. The matching process between the GR observations and RG measurements.

#### 4.4 Neural Network Implementation, Training, and Estimating

We have implemented an MLP-based machine learning model for rainfall estimation using ground radar reflectivity measurements as shown in Fig. 4.5. This model can switch between two different stages: the training stage, to learn the internal relation between the feature-label pairs; and the estimation stage, to calculate the output depending on the learned relations.

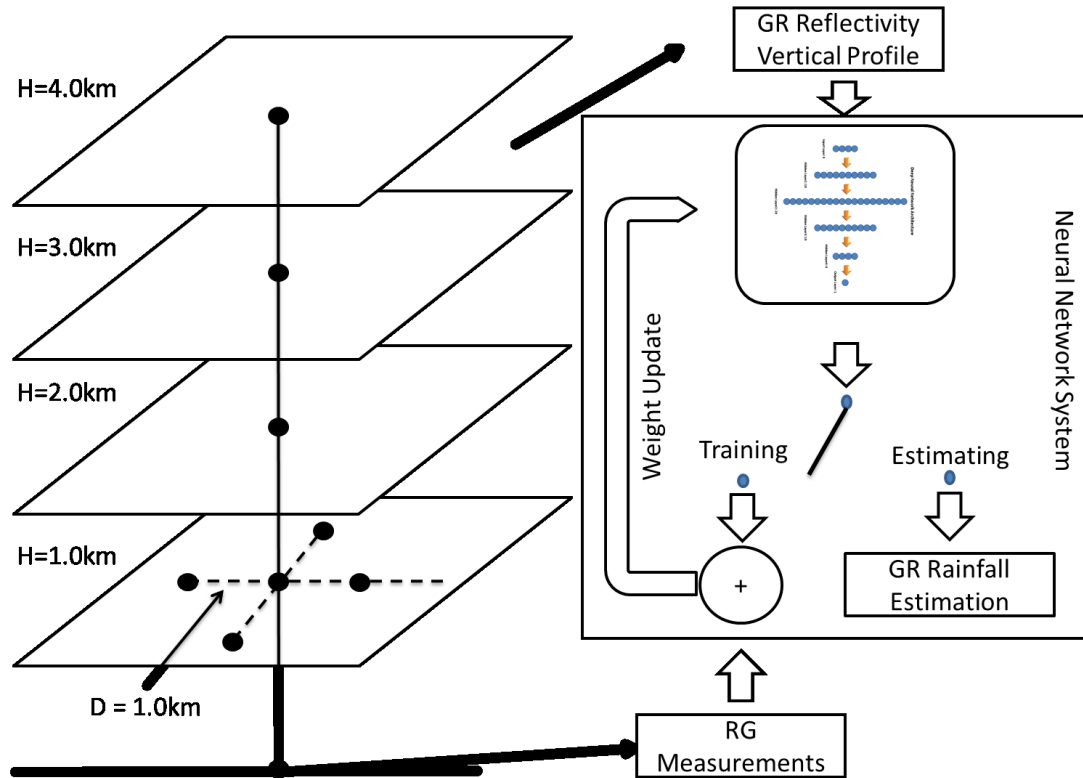


Fig. 4.5. The detailed implementation of the neural network system.

At the training stage, data pairs of GR vertical profiles and RG measurements feed into the neural network as the input features and labels. The estimated rainfall rates are computed as an output of the neural network from input features in the feedforward calculation. Comparing the output (estimated rain rate) and input label (rain rate from gauge), the loss/cost function is derived for optimization of the existing model. In this case, we used the mean squared error as the loss function. Finally, the model parameters, normally called weights and biases, are modified by a back-propagation process in order to minimize the loss function, which can lead to the accuracy estimation. Fig. 4.6 shows the entire process of training the ground radar neural network (GRNN).

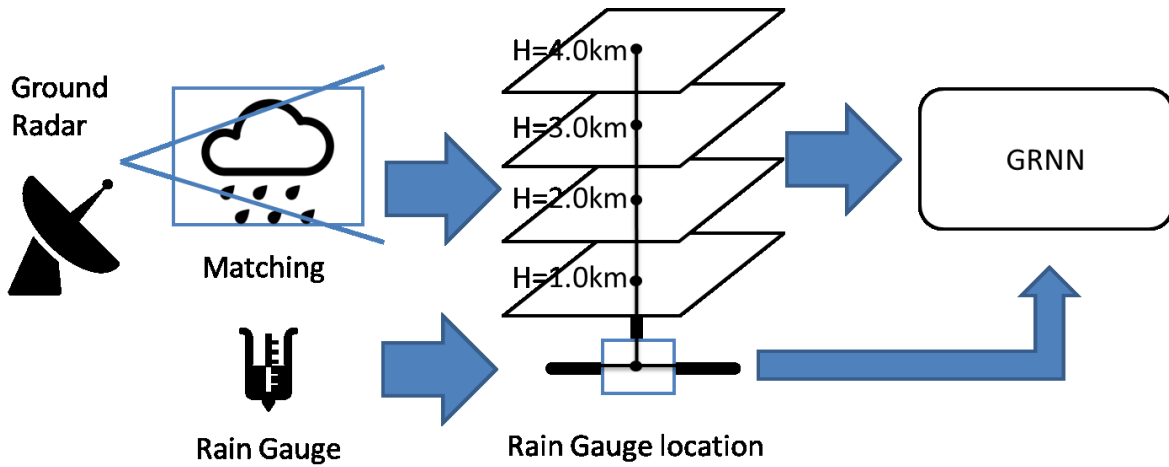


Fig. 4.6 The training process of the neural network system.

At the estimation stage, the input feature is not only limited to the ground radar vertical profile at the gauge location but also works for any location within a valid range. Then, the forward calculation can tell the rainfall estimation at the earth's surface under the GR profile. See the details in Fig. 4.7.

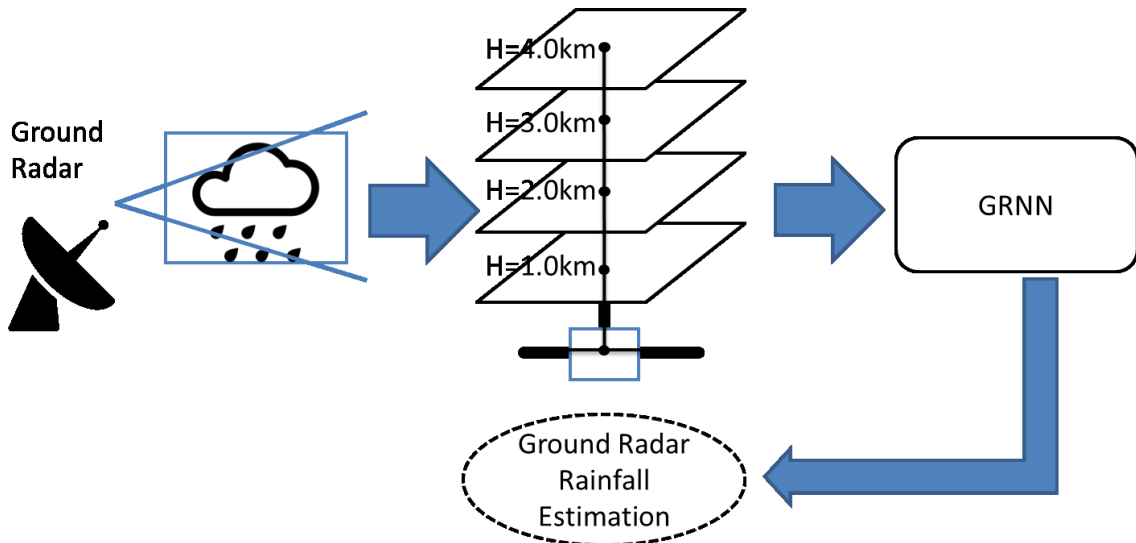


Fig. 4.7. The estimating process of the neural network system.

#### 4.5 Study of the Influence of Sample Size

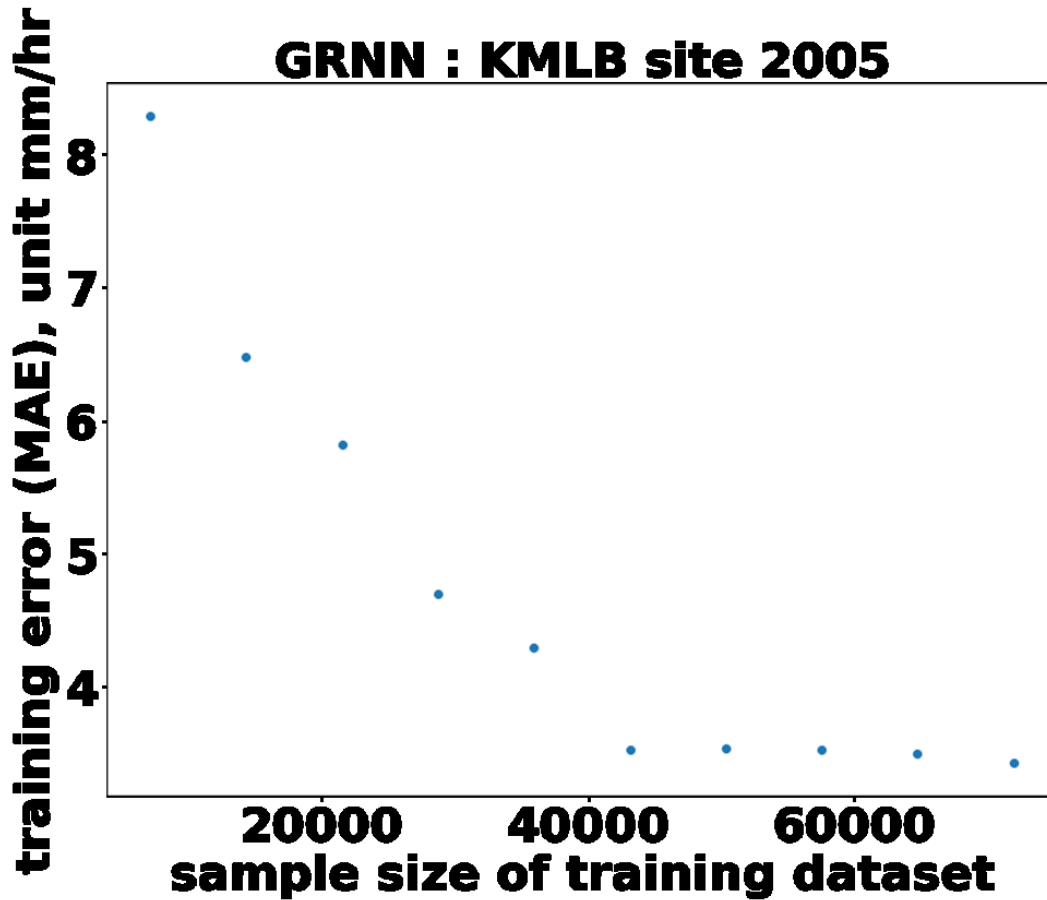


Fig-4.8. The training errors of different training dataset sample sizes.

In order to evaluate the performance of the GR neuron network, the entire dataset has been randomly separated in a ratio of 9:1: 90% is the training dataset and 10% is the testing dataset. The training dataset was used to teach the neuron network to find the best internal parameters while the testing dataset validated the final performance of the well-trained neuron network.

The minimum size of the training dataset is shown in Fig-4.8. We can see a continuous decrease in the training error when the sample size increases from 0 to 40000 and remains when the sample size increases from 40000 to 70000. It indicates that our neural network architecture requires at least 45000 samples to obtain the fewest training errors.

## 4.6 Neural Network Evaluation

Once the network is well trained via the training dataset, the testing dataset is used to estimate model performance. In the evaluation process, the rain rate estimation is calculated from the data from the testing dataset and validated against the rain gauge measurements. The following evaluation matrices 4.1~4.4 were used to do the performance evaluation:

$$Bias = \frac{1}{N} \sum_{i=1}^N (RF_N(i) - RF_g(i)) \quad 4.1$$

$$NSE = \frac{\frac{1}{N} \sum_{i=1}^N |RF_N(i) - RF_g(i)|}{\overline{RF_g}} \quad 4.2$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (RF_N(i) - RF_g(i))^2} \quad 4.3$$

$$Corr = \frac{\frac{1}{N} \sum_{i=1}^N (RF_N(i) - \overline{RF_N})(RF_g(i) - \overline{RF_g})}{\sigma(RF_N) \cdot \sigma(RF_g)} \quad 4.4$$

where  $RF_N$  represents the estimated rainfall from radar observation and  $RF_g$  is the rainfall measurement from the actual rain gauge, respectively;  $N$  is the data sample size; and  $NSE$  and  $RMSE$  are normalized standard error and root mean squared error.

The performance of the network was calculated using the evaluation criteria mentioned above. The network performance was also compared with the simple Z-R relation algorithm used in NEXRAD radars. Figs. 4.9, 4.10, 4.11, and Tables 4.2, 4.4, and 4.6 demonstrate that the neural network model has the low bias,  $NSE$ ,  $RMSE$ , and high correlation when compared with the simple Z-R relation, which means the machine learning algorithm yields a better performance than the Z-R algorithm in rainfall estimation.

Table 4.1 Dataset information from year 2005.

Site	KMLB
Year	2005
Data coverage duration	1~12 months
Total number of samples	79888
Training dataset	71899
Testing dataset	7989
Training and Test ratio	9:1

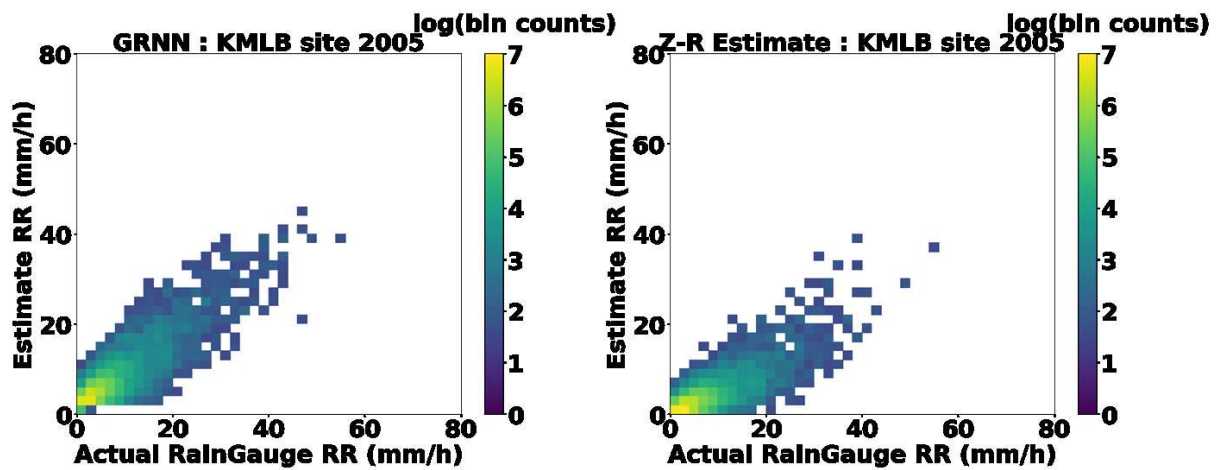


Fig. 4.9. The scatter plot density for NN and Z-R approaches in year 2005.

Table 4.2 Evaluation metrics on year 2005.

	NN	Z-R relation
RainGauge RR (mm/h)	9.27	
Estimate RR (mm/h)	9.45	5.64
Bias (mm/h)	0.18	-3.62
NSE	0.38	0.92
RMS	5.19	13.56
Corr.	0.87	0.83

Table 4.3 Dataset information on year 2009.

Site	KMLB
Year	2009
Data coverage duration	1~12 months
Total number of samples	111995
Training dataset	100795
Testing dataset	11200
Training and Test ratio	9:1

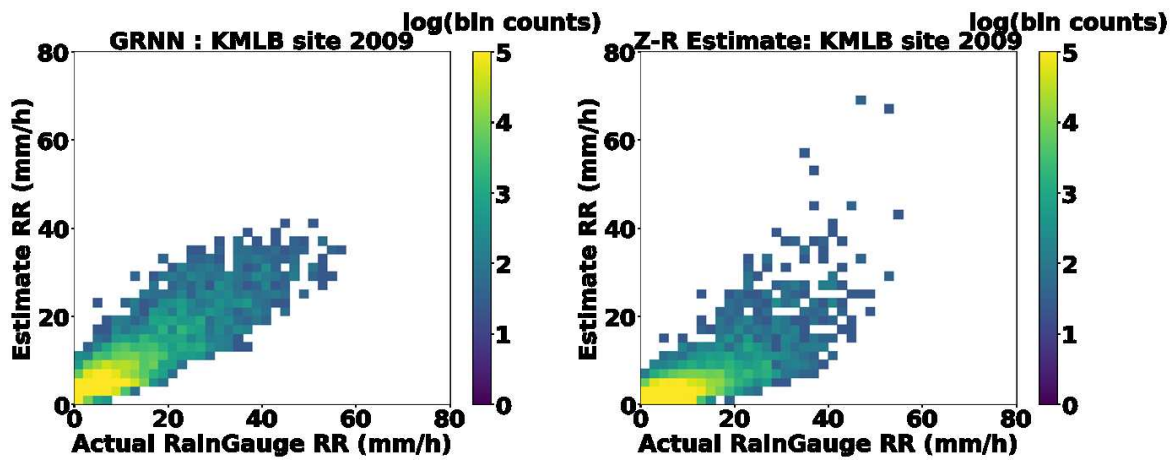


Fig. 4.10. The scatter plot density for the NN and Z-R approaches in year 2009.

Table 4.4. Evaluation metrics for year 2009.

	NN	Z-R relation
RainGauge RR (mm/h)	10.49	
Estimate RR (mm/h)	9.98	5.41
Bias (mm/h)	-0.51	-5.08
NSE	0.41	0.93
RMS	6.25	15.33
Corr.	0.81	0.73



Table 4.5 Dataset information for year 2014.

Site	KMLB
Year	2014
Data coverage duration	1~12 months
Total number of samples	88648
Training dataset	79783
Testing dataset	8865
Training and Test ratio	9:1

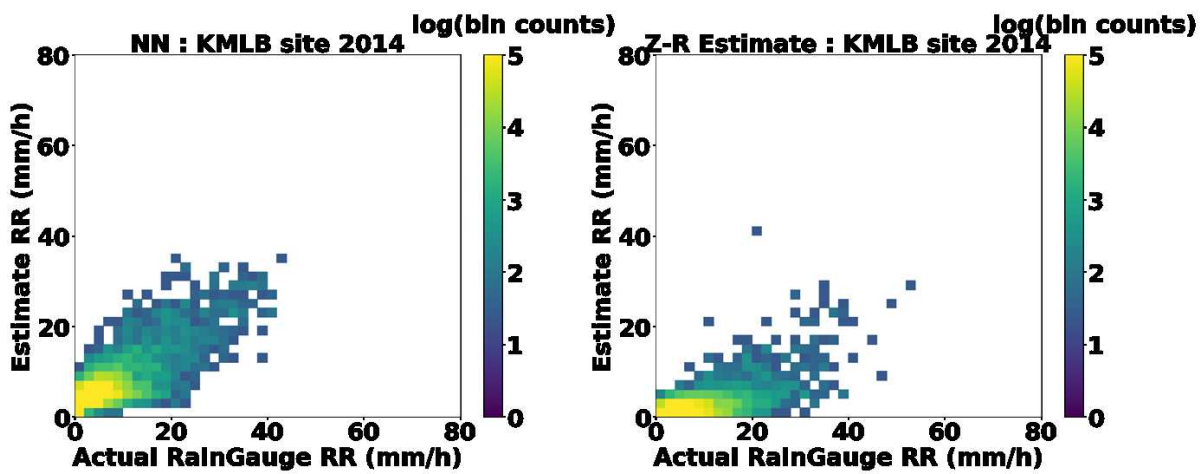


Fig. 4.11. The scatter plot density for the NN and Z-R approaches in year 2014.

Table 4.6 Evaluation metrics on year 2014.

	NN	Z-R relation
RainGauge RR (mm/h)	6.06	
Estimate RR (mm/h)	6.28	2.35
Bias (mm/h)	0.22	-3.71
NSE	0.48	1.05
RMS	4.73	10.81
Corr.	0.84	0.74

#### 4.7 Study of Feature Importance

After evaluating the performance of the neuron network, an analysis on the weights of the first layer in the neuron network provided information about feature importance. The importance of the feature is defined in equation 4.5:

$$feature\ importance = \frac{\sum_1^N |W_i|}{N} \quad 4.5$$

Where  $N$  the number of nodes in the input is layer and  $W_i$  is the weight of the  $i$ th node of the input layer.

In this design, the input features are the reflectivities at the different altitudes. From Fig-4.12, we can see that the reflectivity at 1km has the highest value in feature importance for 2005, 2009 and 2014. It means this features made the greatest contribution to the rainfall estimation based on the designed neuron networks.

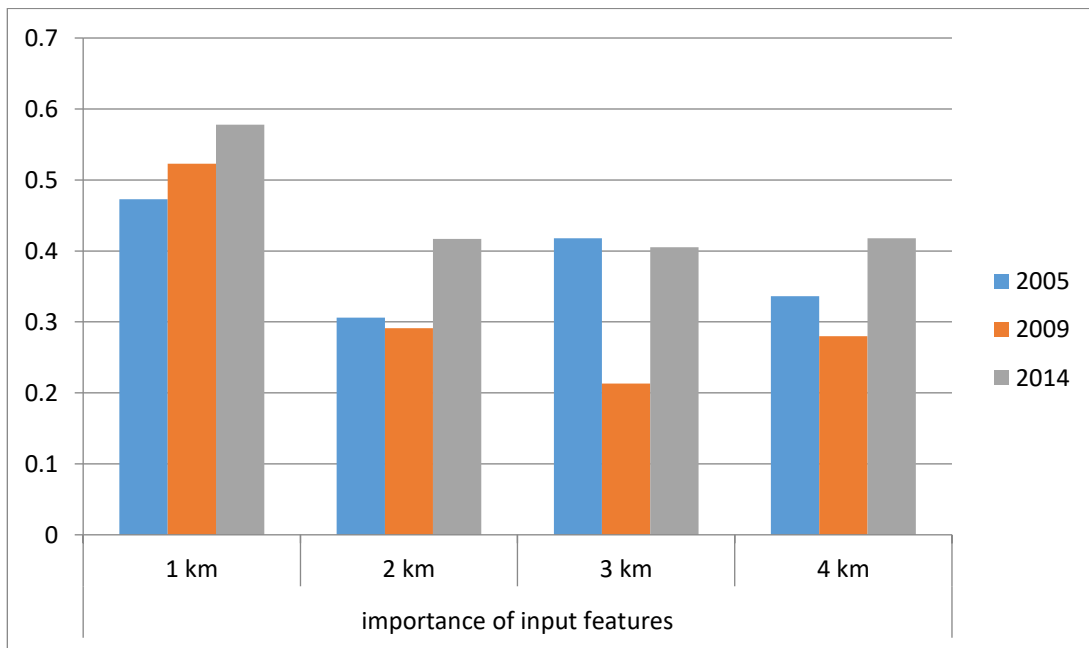


Fig. 4.12. Statistical results of the input features' importance.

## 5 DEVELOPMENT OF THE DEEP NEURAL NETWORK MODEL FOR TRMM RAINFALL ESTIMATION

### 5.1 Problems in Spaceborne Radar Rainfall Estimation

The rainfall algorithm used by TRMM PR was introduced in Chapter 3. The ideal rainfall algorithm of space radar (SR) should capture the space-time variability of precipitation microphysics. This empirical relation includes significant uncertainty and needs to be adjusted based on validation. Commonly, rain gauge networks are used for ground validation of SR observations. However, the different characterizations between SR and rain gauge networks in terms of resolution, scale, viewing aspect, and environment lead to challenges in the ground validation process. For example, TRMM PR, which is one type of spaceborne radar, has a 5km x 5km horizontal resolution, which is much coarser relative to a rain gauge network. In addition, the data pairs between the TRMM vertical profile and rain gauge measurement for comparison are scarce because of the limits of the TRMM overpass during a single weather event. Deploying a dense gauge network for TRMM PR validation is impractical due to engineering and financial issues.

On the other hand, ground radar can obtain a large amount of training pairs much more easily than can a spaceborne radar like TRMM PR. This is because both ground radar observations and rain gauge measurements share a similar location, with fine temporal resolution. Using ground radar for TRMM PR validation is a more practical approach because ground radar can bring many more data pairs for validation with rain gauge comparisons due to the wide scan area of the ground radar. Thus, the data pairs were built using ground radar observations instead of TRMM PR observations with rain gauge measurements. Amin and V. Chandrasekar (2008) have proposed a two-stage hybrid system using two radial-based function (RBF) neural networks to build a relation between the rain gauge and ground

radar, then transfer this relation to TRMM PR ground validation. This work shows great potential for validating spaceborne radar by introducing ground radar using a neural network technique.

## 5.2 Two-Stage Neural Network Architecture

However, the RBF brings a great deal of nonlinearity as activation function prevents the neural network from moving deeper. We plan to replace the RBF neural network with a deep neural network to increase its abstraction capability. In this chapter, we describe how our group has designed a similar architecture of a hybrid two-stage system with a deep neural network to build a relation between rainfall gauge measurements and ground radar observations, then transfer this relationship to TRMM PR observations for rainfall estimation and local rainfall mapping. The hybrid system first trains ground radars for rainfall estimation using rain gauge data and subsequently uses the trained ground radar neural network rainfall estimated result to train the TRMM PR-based neural network for rainfall estimation. This system provides an alternative method for estimating the rainfall from TRMM observations based on a non-parametric method. The system architecture is shown in Fig. 5.1. The first deep neural network (DNN1) is designed for ground radar rainfall estimation. It is also called ground radar neural network (GRNN). The second deep neural network (DNN2) estimate satellite radar rainfall, and it has another name satellite radar neural network (SRNN).

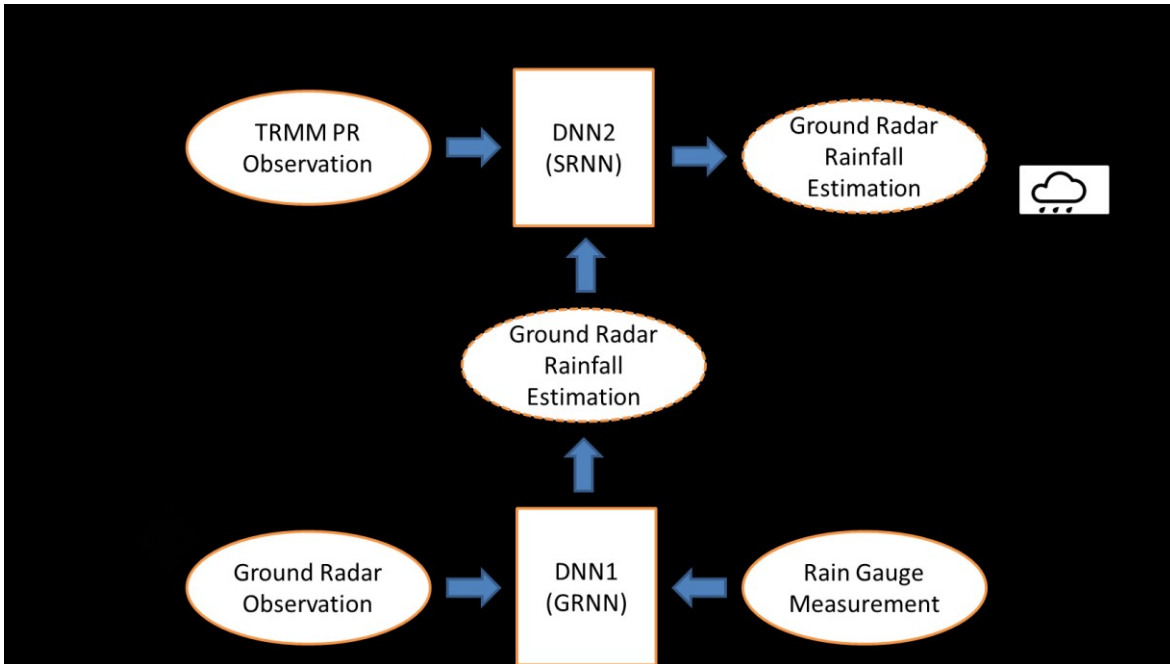


Fig. 5.1. System architecture of the two-stage hybrid system.

Chapter 4 has demonstrated that the MLP, as one type of deep neural network, can learn the relation between ground radar measurements and rain gauge data. MLP has proven its capability to learn complex relationships from high-input features to a target space. The similar architecture of MLP in ground radar estimation will be used in this chapter to convert TRMM PR observations to rainfall estimations.

### 5.3 Training and Validating Dataset

The ground radar neural networks (NN1) were trained using NEXRAD level 2 momentum data around the Melbourne, Florida, ground validation site (KMLB) from storm events in 2005, 2007, 2009, and 2014. The overpasses at this ground validation site were used as the dataset for training the satellite radar neural networks. There were 600 overpasses for TRMM in one year, while only a very small amount of them (around 30 to 40 cases) were good precipitation cases and can be considered in this study. The entire dataset containing ground and satellite radar observation and gauge measurements were used to train this two-stage neural network. The first network was trained based on ground radar data and its corresponding gauge values. This network was used to generate the rainfall estimation in

the entire overpass region as the target label for training the next stage of the neural network. The second network was trained based on satellite radar observation and rainfall estimation from the first network at the time of the overpass. Rainfall estimation of any new data was done based on the network built by previous overpass data.

## 5.4 Two-Stage Neural Network Implementation

The process of building a hybrid model for TRMM PR rainfall estimation has four main steps: ground radar validation neural network design, TRMM PR and ground radar data alignment, the space radar neural network design, and system validation. These steps are described in detail in the following subsections.

### 5.4.1 Ground radar validation neural network design

In this step, a neural network must estimate rainfall radar based on ground validation radar and the rain gauge. A deep neural network-based approach was designed and implemented, as described in chapter 4. This approach has demonstrated better performance when compared with traditional a Z-R relation and has the advantage of a large population of training samples. We used this approach as the first stage of the hybrid system to create the ground radar estimation as the training label for the neural network in the second stage. More details can be seen in chapter 4.

### 5.4.2 TRMM PR and ground radar data alignment design

For training purpose, the observation of ground radar and TRMM-PR need to match the data pairs, which are reflectivity profiles with the same space resolution and time stamp (see Fig. 5.2). The method developed by Bolen and Chandrasekar (2003) is used to align ground and space radar reflectivity profiles despite the difference of viewing angle and spatial resolution from two observation platforms. This alignment method can generate a product to compare radar reflectivity factors at

different heights by re-sampling the ground-based and spaceborne radar dataset to a common grid. The resolution of final alignment product of TRMM PR and ground radar is 4 x 4 x 0.5 km.

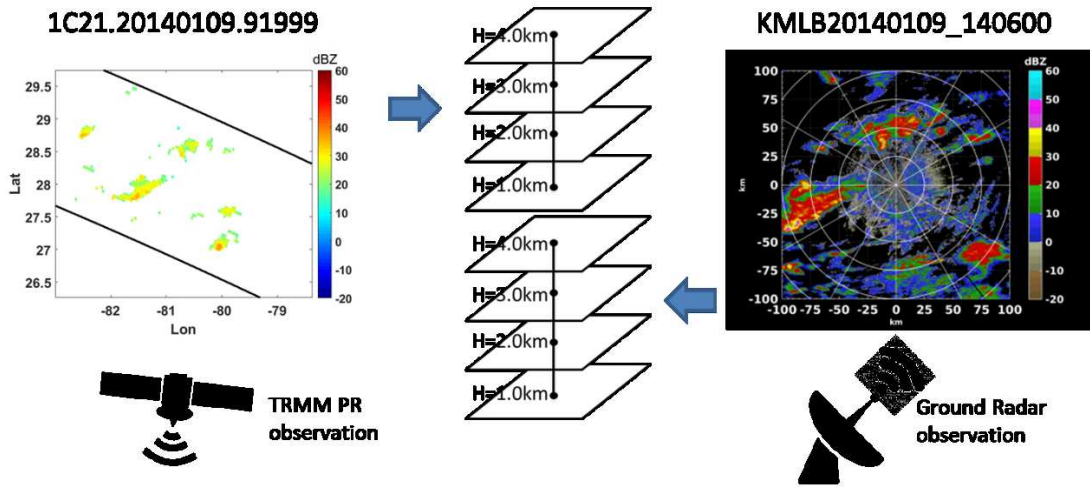


Fig. 5.2. The data pair required for neural network training.

An example of the alignment process is shown in Fig. 5.3. Data were taken from the KMLB region from 2014. We can see a good match between the two radar products. This alignment will bring around 1dB bias in radar reflectivity factor, and this bias can be automatically compensated by the second neural network.

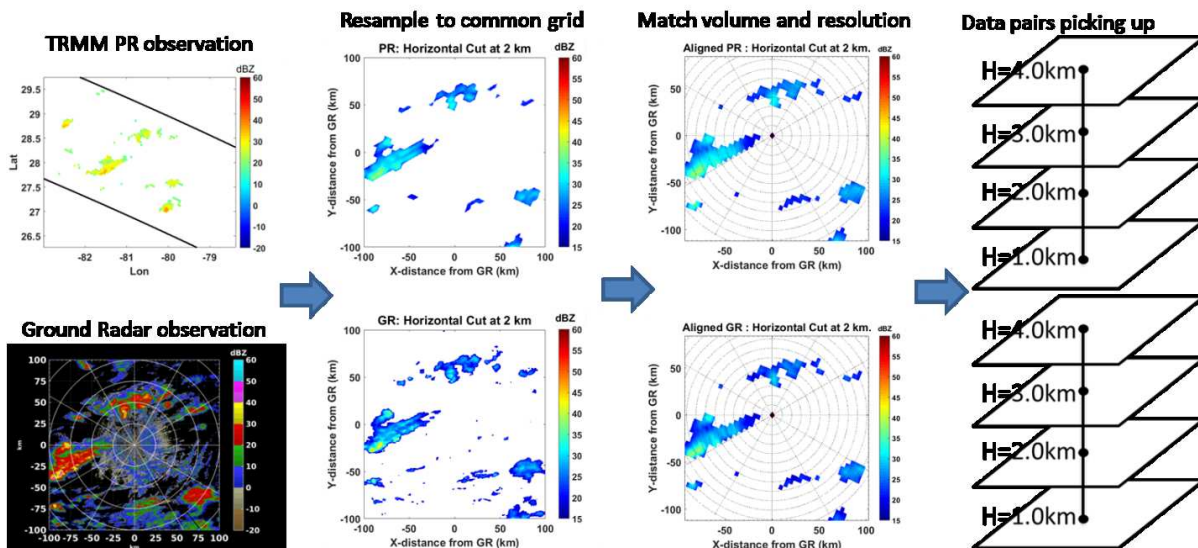


Fig. 5.3. The alignment process for generating the data pairs.

### 5.4.3 The space radar neural network design

In the training process, the input features of the second neural network are TRMM PR reflectivity vertical profiles from 1km to 4km with 1km vertical resolution. These profiles were aligned to ground radar reflectivity vertical profiles using the method described in 5.4.2. The target of the second neural network is the rainfall rate estimation generated by the first neural network with the ground radar reflectivity vertical profile aligning with the TRMM PR data. This target will be used to train the second network. The entire process of training the TRMM-PR neural network is shown in Fig. 5.4. In the estimation process, the input features are the TRMM PR profiles, which are not limited to the overpass at the KMLB region. The outputs of the second neural network are rainfall estimation of the TRMM PR at the earth's surface. See the details in Fig. 5.5.

The training and estimation of the second neural network were based on the same technique used in the ground radar neural network we discuss in chapter 4. Rainfall estimation was done using the model of the previous overpasses in 2005, 2007, 2009, and 2014.

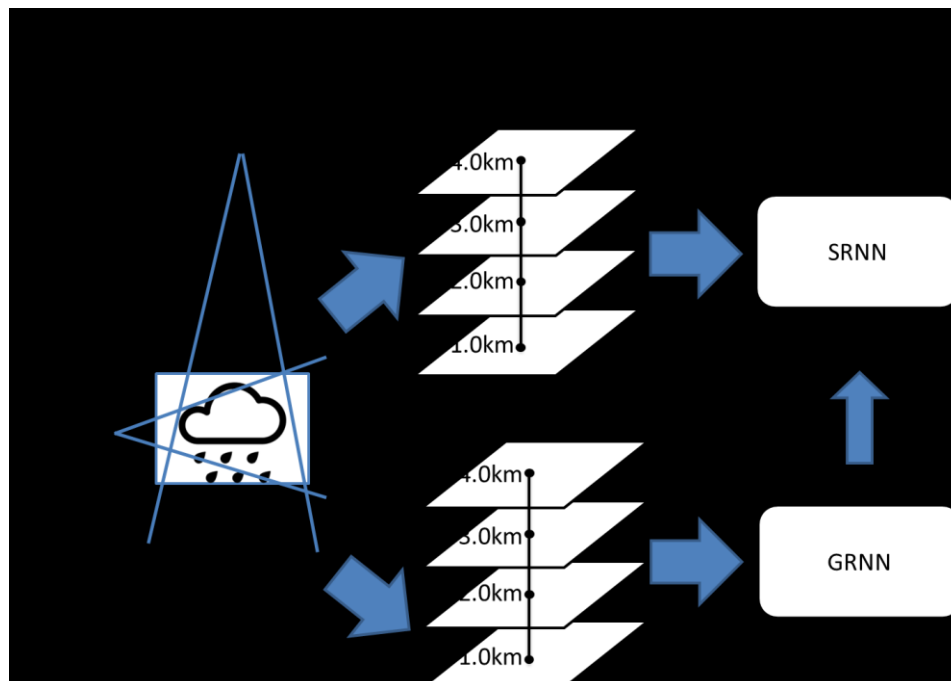


Fig. 5.4. The training process of the hybrid system.



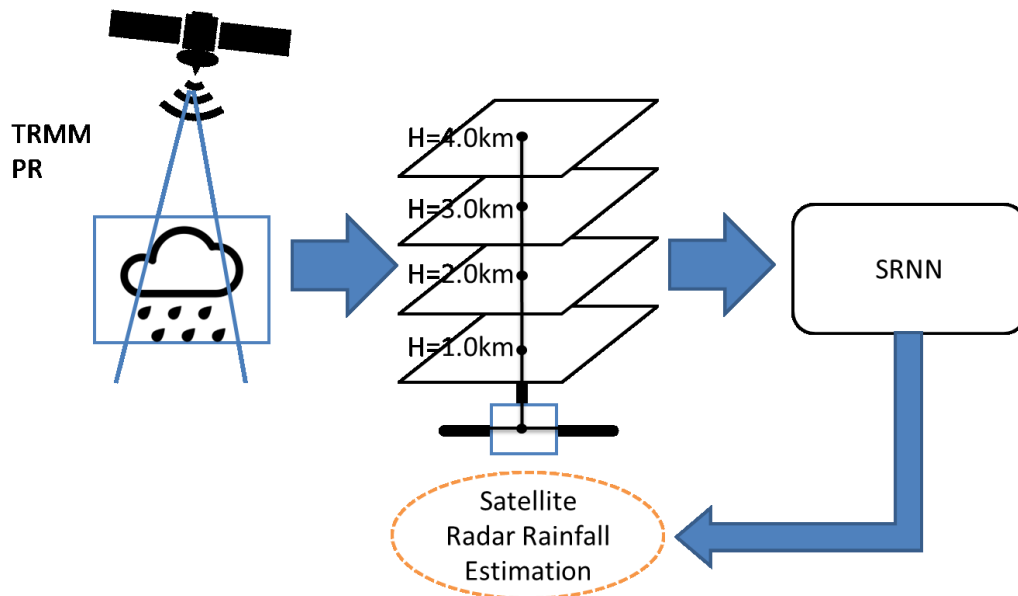


Fig. 5.5. The estimation process of the hybrid system.

The validation results shown in Fig. 5.6 demonstrate the capability of the second neural network

to capture the relation between the TRMM reflectivity profiles and the GR estimated rain rate.

Table 5.1. The sample size of the dataset for TRMM neural network training and testing.

Year	2005	2007	2009	2014
Data coverage duration	1~12 months	1~12 months	1~12 months	1~12 months
Total number of samples	7903	6018	7185	7772
Training dataset	7112	5416	6466	6994
Testing dataset	791	602	719	778
Training and Test ratio	9:1			

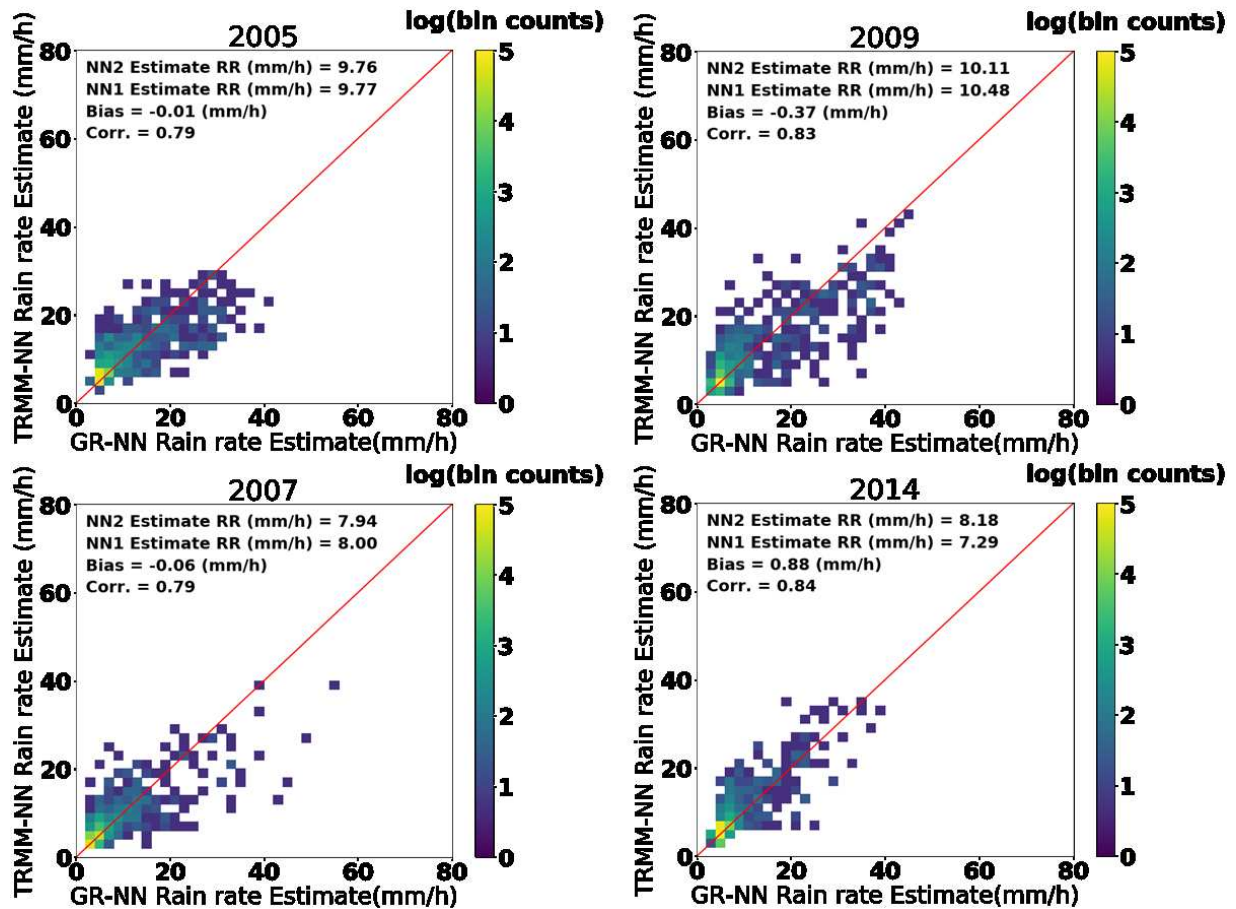


Fig. 5.6. Validation results from the TRMM neural network in 2005, 2007, 2009 and 2014.

#### 5.4.4 System validation

In order to validate the hybrid neural network system, the rain rate estimated by the first network and second network will be compared to the rain gauges. The TRMM PR rain rate product was also compared to the rain gauges in order compare the TRMM-PR rain rate product with the neural network product. The same scoring metrics used in chapter 4 were used to evaluate the performance of each network as well as the performance of TRMM-PR product in this chapter.

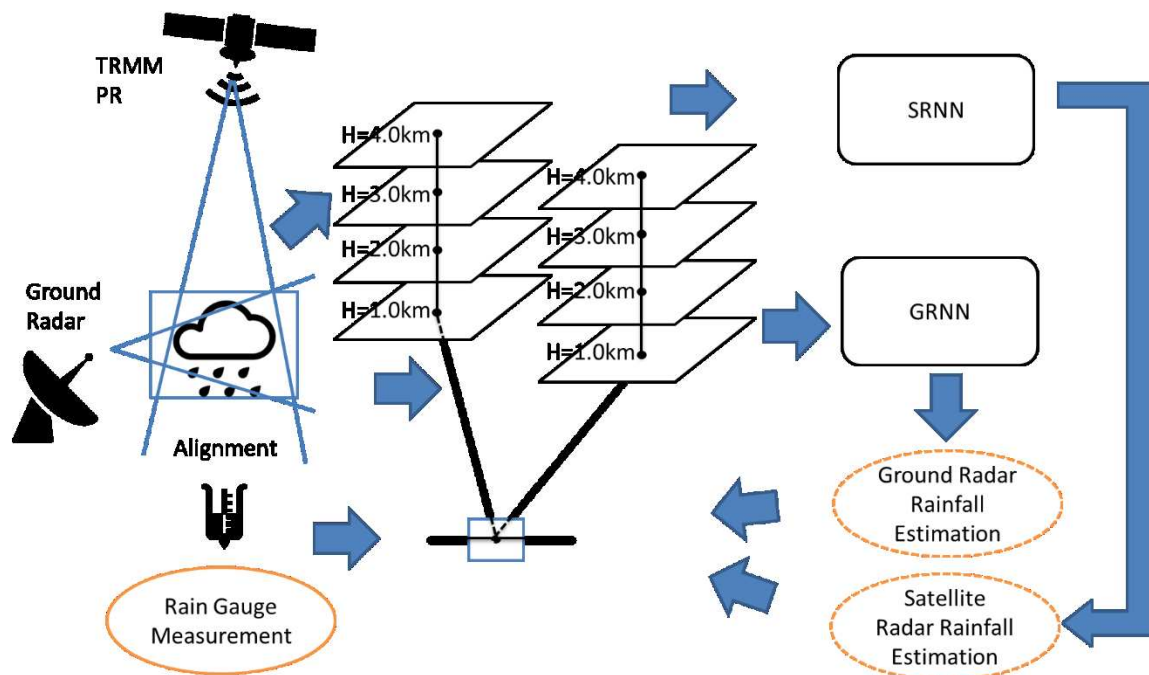


Fig 5.7. The validation process of the hybrid system.

For system validation, the TRMM PR observations, TRMM PR products, ground radar observations, and rain gauge measurements were all collected from overpass cases during 2005, 2007, 2009, and 2014 over Melbourne, Florida. All the data except the TRMM PR products are aligned by the pre-process method described in sections 4.3 and 5.3. The limited number of overpass cases directly restricts the size of the validation dataset. The TRMM PR rain rate product used for comparison was the surface rain rate derived by the parametrical method from the TRMM 2A25 products. Tables 5.2, 5.4, 5.6, and 5.8 show that the sample size of the validation dataset ranges from 100 to 200 in one year, which is at least one order smaller than the size of the GR-RG dataset and the TRMM-GR dataset. This is the reason that direct training of the TRMM NN with rain gauge measurements is impossible.

Tables 5.3, 5.5, 5.7, and 5.9 show the scores of the two-stage system using data from 2005, 2007, 2009, and 2014, respectively. As shown, the performance of the hybrid system is better than the performance of the TRMM product in terms of lower bias and higher correlations. Figs. 5.8, 5.9, 5.10,

and 5.11 show the same conclusion from using scatter plots of both neural networks against the TRMM standard product.

Table 5.2. System validation dataset information for 2005.

Site	KMLB
Year	2005
Data coverage duration	1~12 months
GR-RG dataset size	79888
TRMM-GR dataset size	7903
System validation dataset size	209

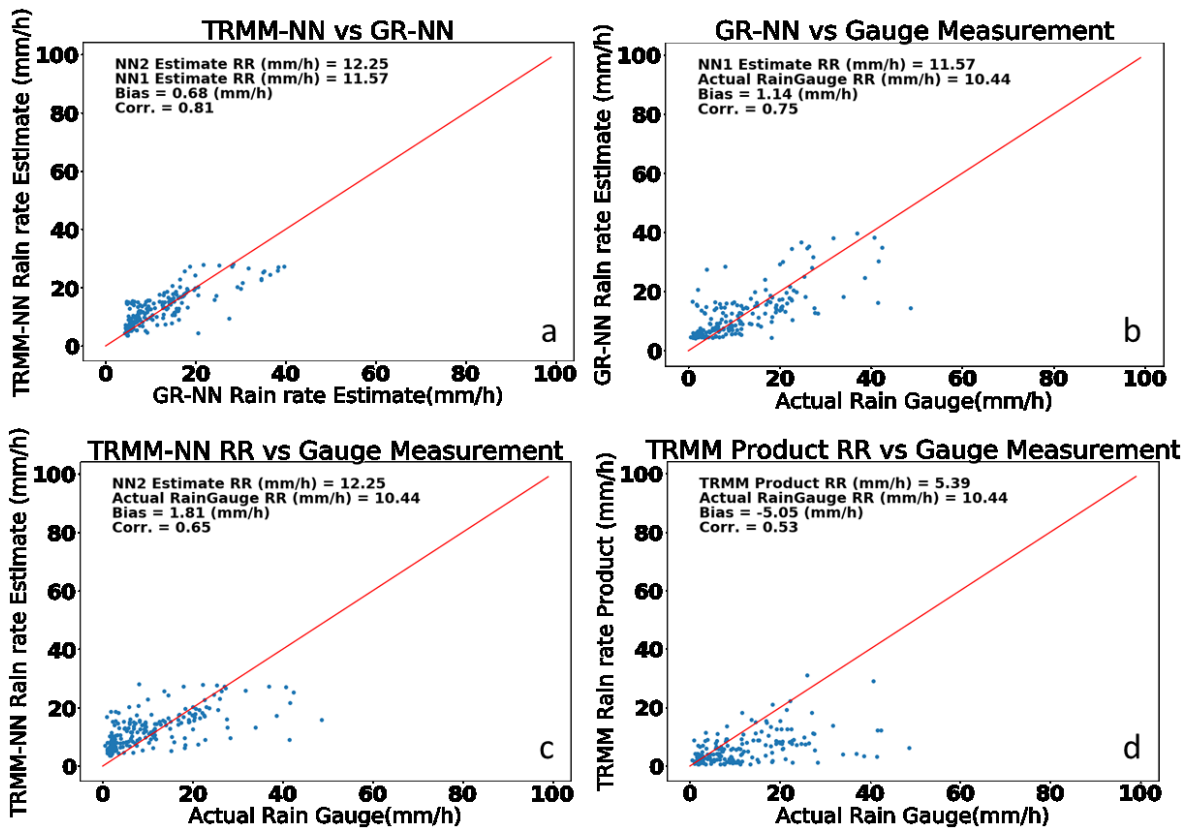


Fig. 5.8. System validation results for 2005.

Table 5.3. Performance comparison between the TRMM NN and TRMM PR products ( 2005).

	Estimate RR (mm/h)	RainGauge RR (mm/h)	Bias (mm/h)	Corr.
TRMM NN	12.25	10.44	-1.81	0.65
TRMM Product	5.39		-5.05	0.53

Table 5.4. System validation dataset information for 2007.

Site	KMLB
Year	2007
Data coverage duration	1~12 months
GR-RG dataset size	74792
TRMM-GR dataset size	5416
System validation dataset size	106

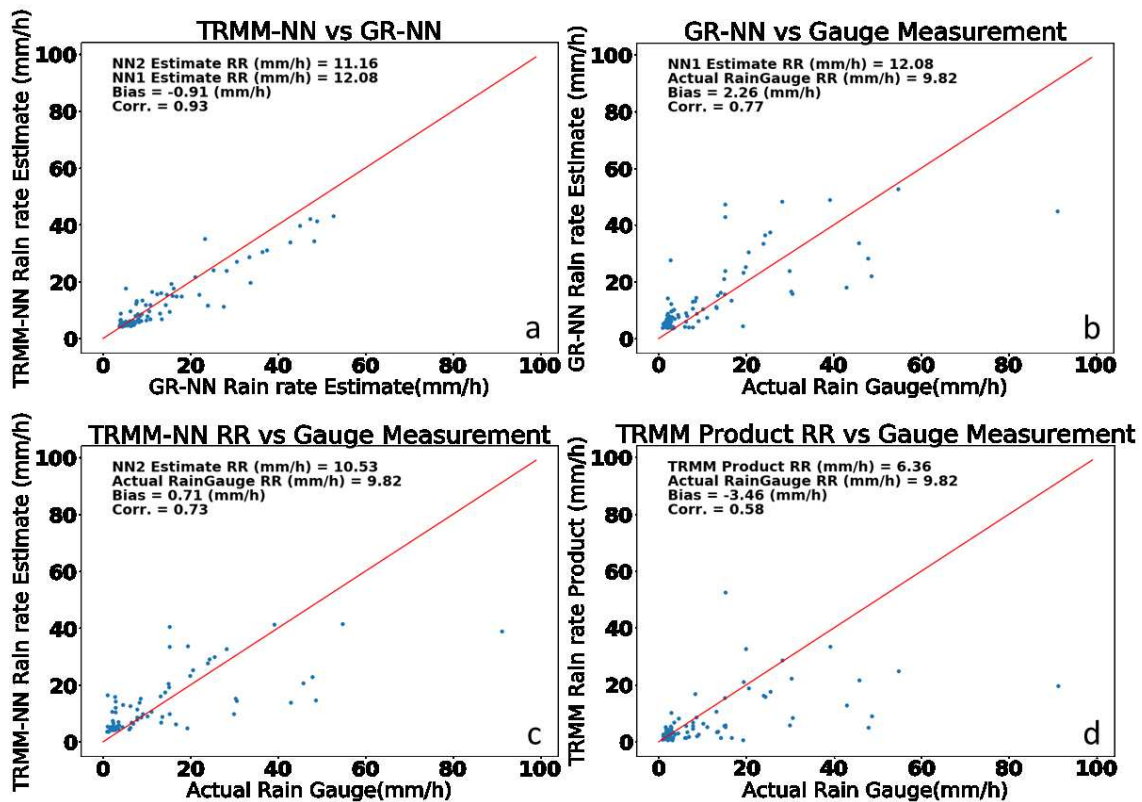


Fig. 5.9. System validation results for 2007.

Table 5.5. Performance comparison between the TRMM NN and TRMM PR products (2007).

	Estimate RR (mm/h)	RainGauge RR (mm/h)	Bias (mm/h)	Corr.
TRMM NN	10.53	9.82	0.71	0.73
TRMM Product	6.36		-3.46	0.58

Table 5.6 System validation dataset information for 2009.

Site	KMLB
Year	2009
Data coverage duration	1~12 months
GR-RG dataset size	111995
TRMM-GR dataset size	6466
System validation dataset size	169

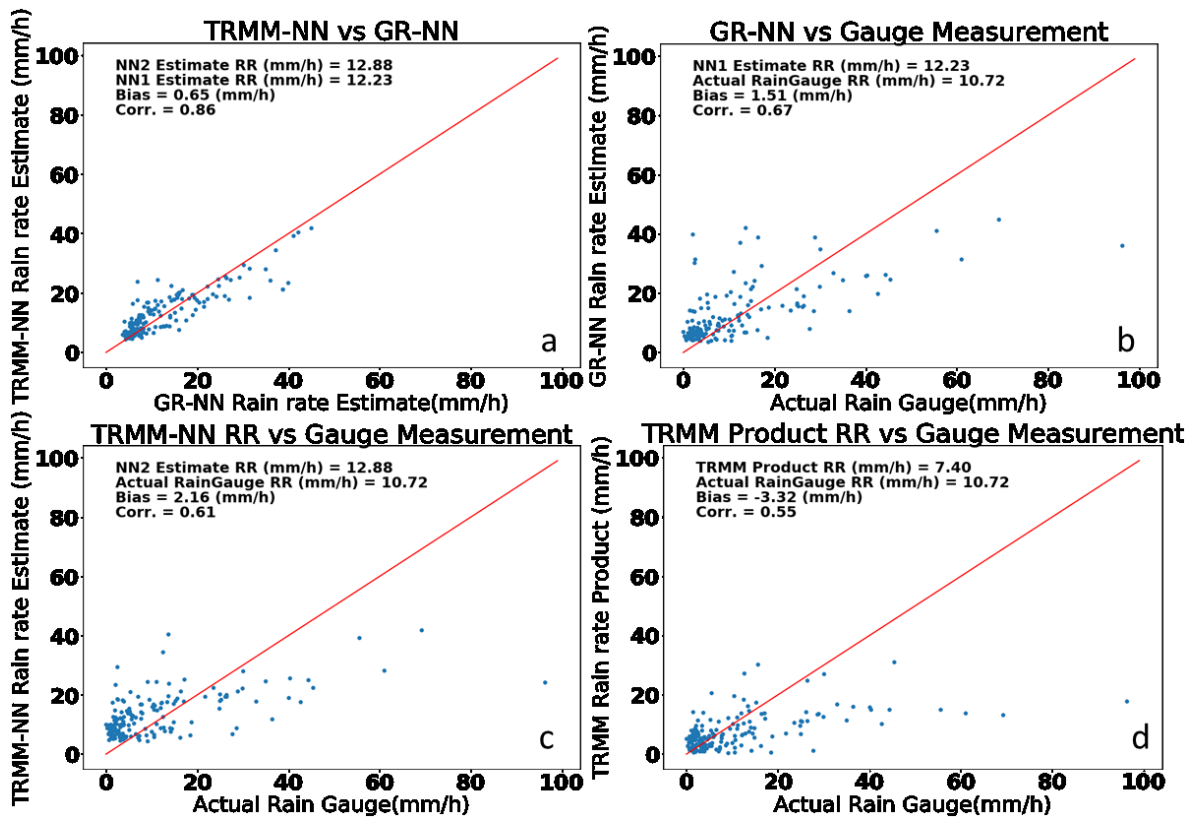


Fig. 5.10. System validation results for 2009.

Table 5.7. Performance comparison between the TRMM NN and TRMM PR products (2009).

	Estimate RR (mm/h)	RainGauge RR (mm/h)	Bias (mm/h)	Corr.
TRMM NN	12.88	10.72	2.16	0.61
TRMM Product	7.40		-3.32	0.55

Table 5.8. System validation dataset information for 2014.

Site	KMLB
Year	2014
Data coverage duration	1~12 months
GR-RG dataset size	88648
TRMM-GR dataset size	6994
System validation dataset size	130

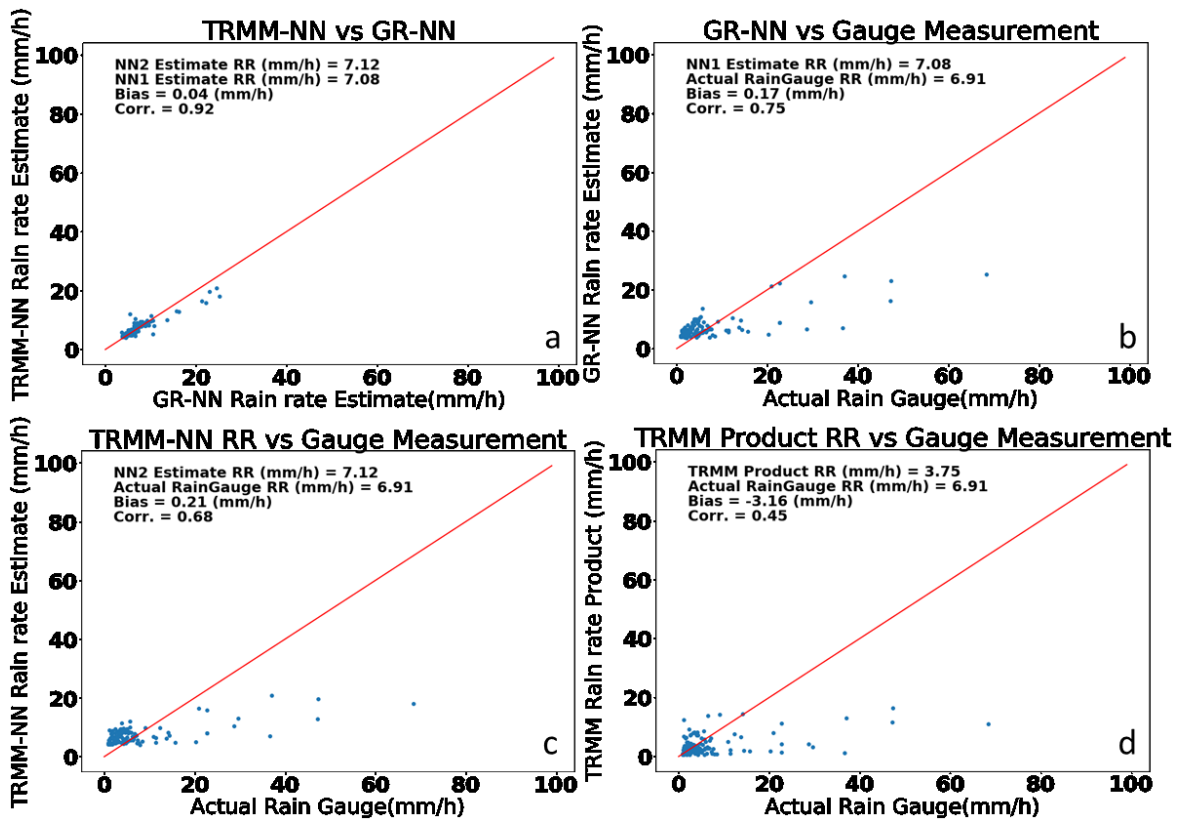


Fig. 5.11. System validation results for 2014.

Table 5.9. Performance comparison between the TRMM NN and TRMM PR products (2014).

	Estimate RR (mm/h)	RainGauge RR (mm/h)	Bias (mm/h)	Corr.
TRMM NN	7.12	6.91	0.21	0.68
TRMM Product	3.75		-3.16	0.45

## 5.6 Local Rainfall Map Generation

One of the advantages of the two-stage satellite radar rainfall estimation is that it can be used for generating local rainfall maps from both ground-based and space-based radars. For rain map generation, each network in the hybrid network is tested for some instances of the TRMM-PRs as well as their corresponding ground radar measurements.

Figs. 5.13 to 5.18 show instances seen by the KMLB radar and overpassed by the TRMM radar. Each instance was tested by the KMLB neural network that was designed in the year the instances were measured. The top row of the figures shows two reflectivity factor maps from both TRMM PR and ground radar. The second row shows the rainfall estimation maps generated by the corresponding neural networks. The third row shows the rainfall map from the TRMM standard product for comparison purposes. From these figures, the rainfall maps generated by the hybrid neural networks show better representation of storms compared to the TRMM standard rainfall product; the TRMM product tends to underestimate the rainfall while the neural network technique captures the storm with small bias. The figures also demonstrate that outputs from both networks are very close to each other, and that the TRMM-PR standard product tends to underestimate rain rate compared to the hybrid system.



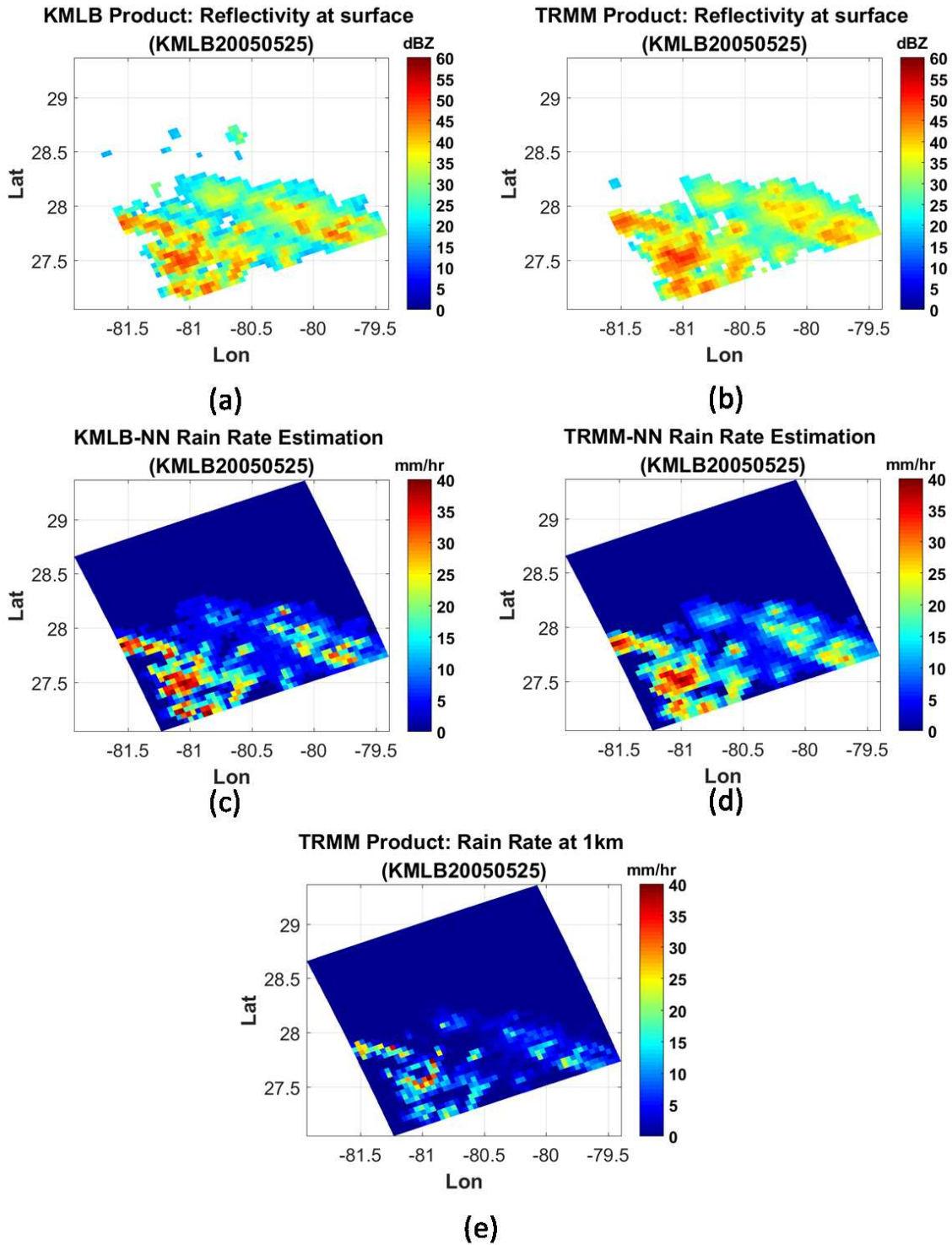


Fig. 5.12. (a) GR reflectivity at 1km height, (b) TRMM PR reflectivity at 1km height, (c) GR NN rain rate estimation, (d) TRMM NN rain rate estimation, (e) TRMM PR product rain rate estimation. (Case: 05/05/2005)

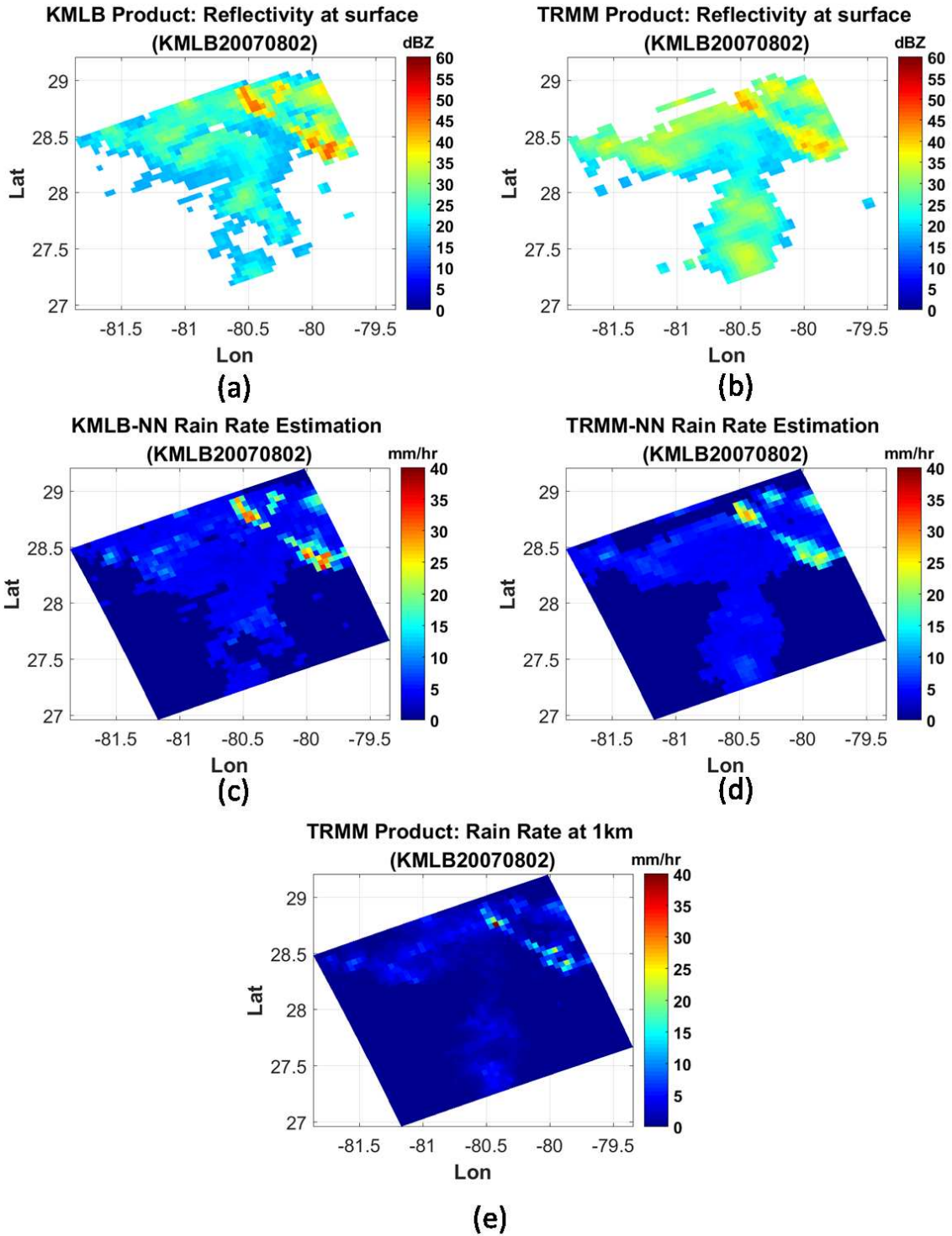


Fig. 5.13. (a) GR reflectivity at 1km height, (b) TRMM PR reflectivity at 1km height, (c) GR NN rain rate estimation, (d) TRMM NN rain rate estimation, (e) TRMM PR product rain rate estimation. (Case: 08/02/2007)

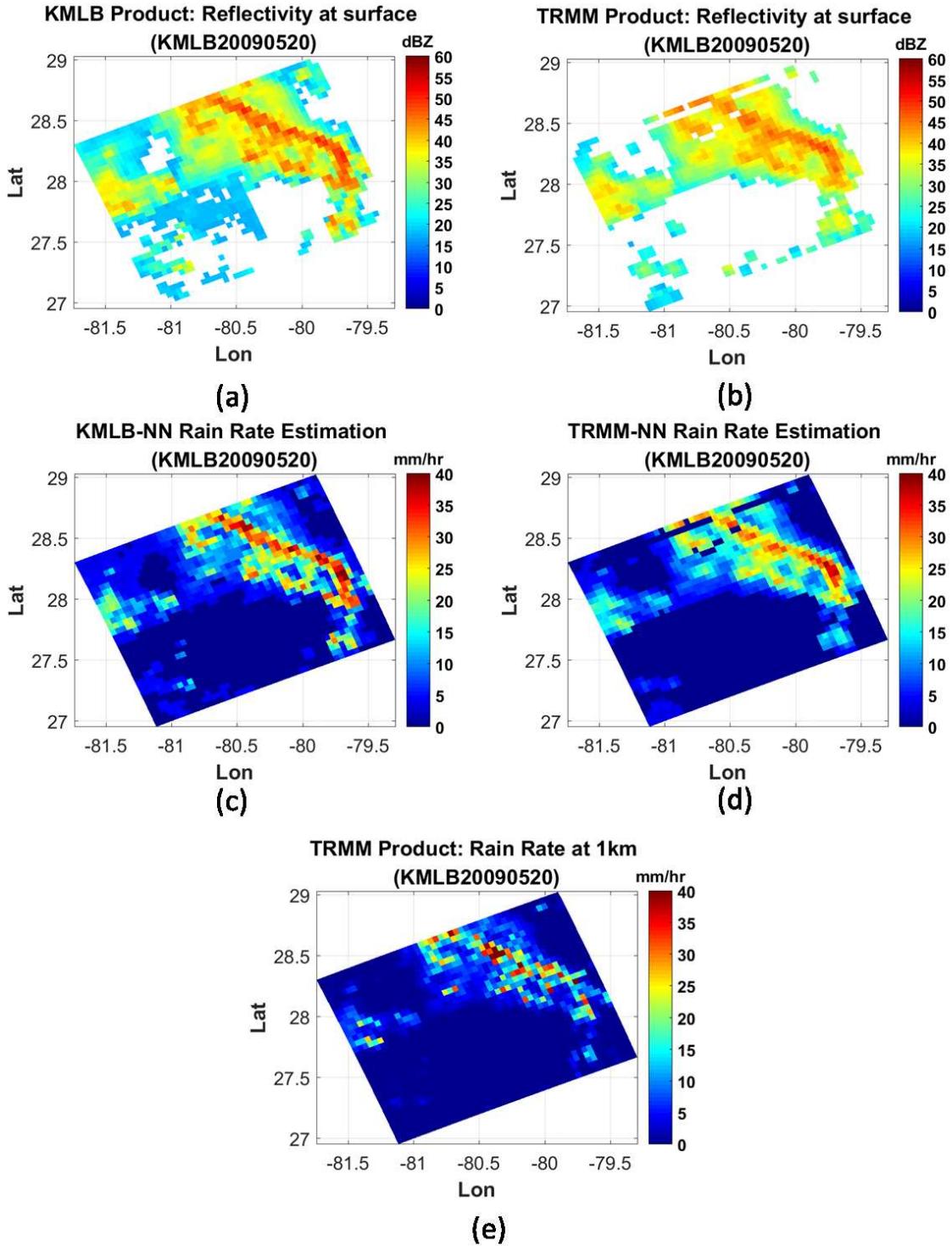


Fig. 5.14. (a) GR reflectivity at 1km height, (b) TRMM PR reflectivity at 1km height, (c) GR NN rain rate estimation, (d) TRMM NN rain rate estimation, (e) TRMM PR product rain rate estimation. (Case: 05/20/2009)

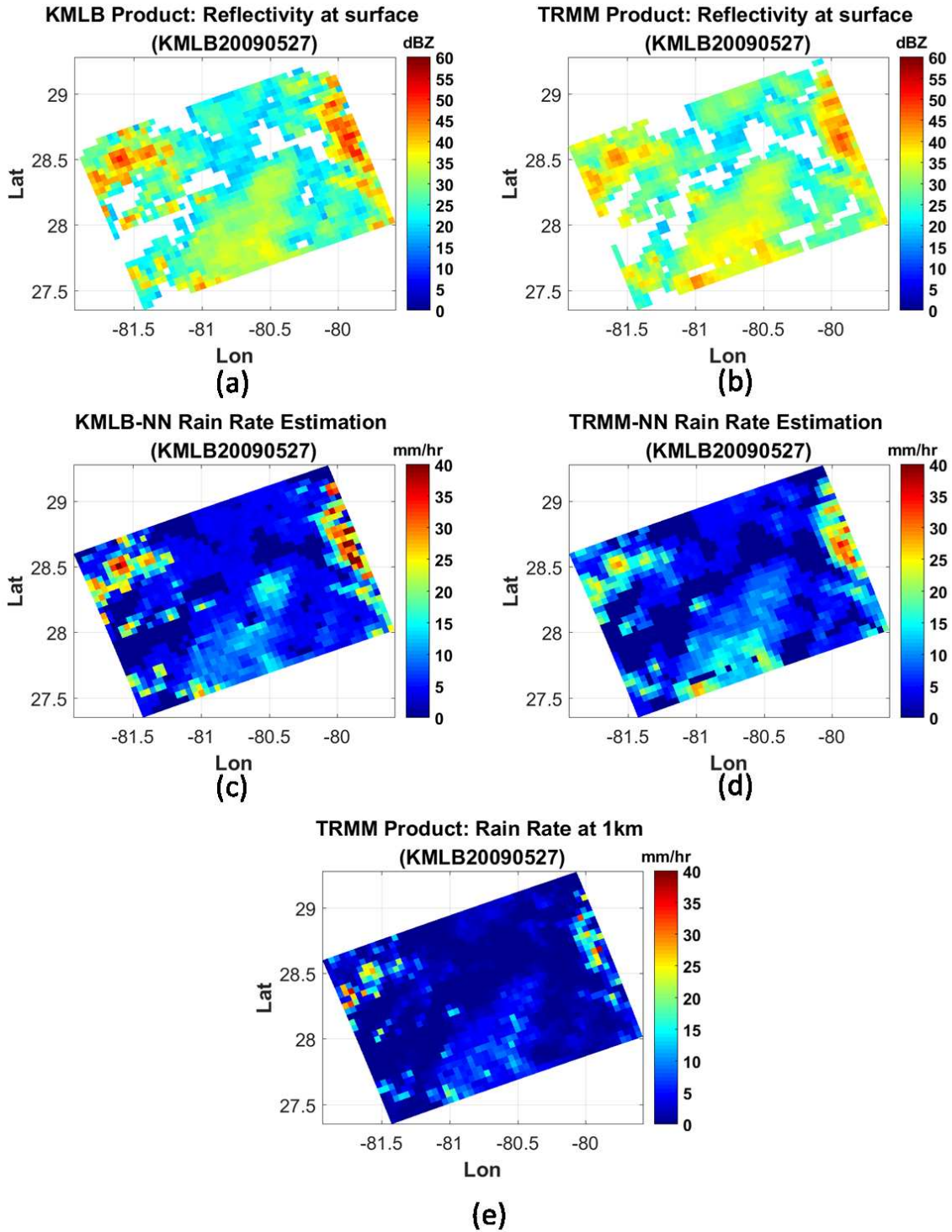


Fig. 5.15. (a) GR reflectivity at 1km height, (b) TRMM PR reflectivity at 1km height, (c) GR NN rain rate estimation, (d) TRMM NN rain rate estimation, (e) TRMM PR product rain rate estimation. (Case: 05/27/2009)

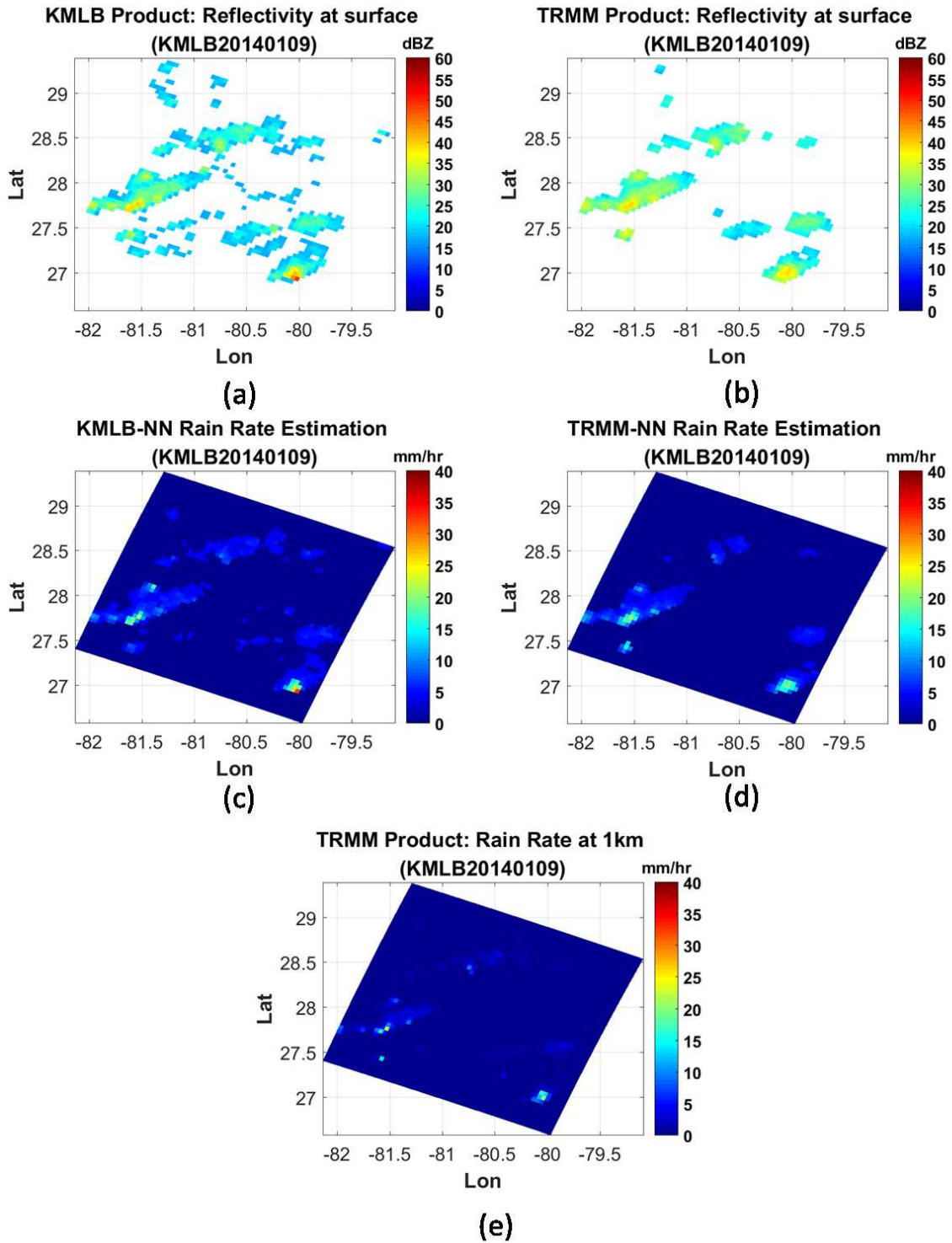
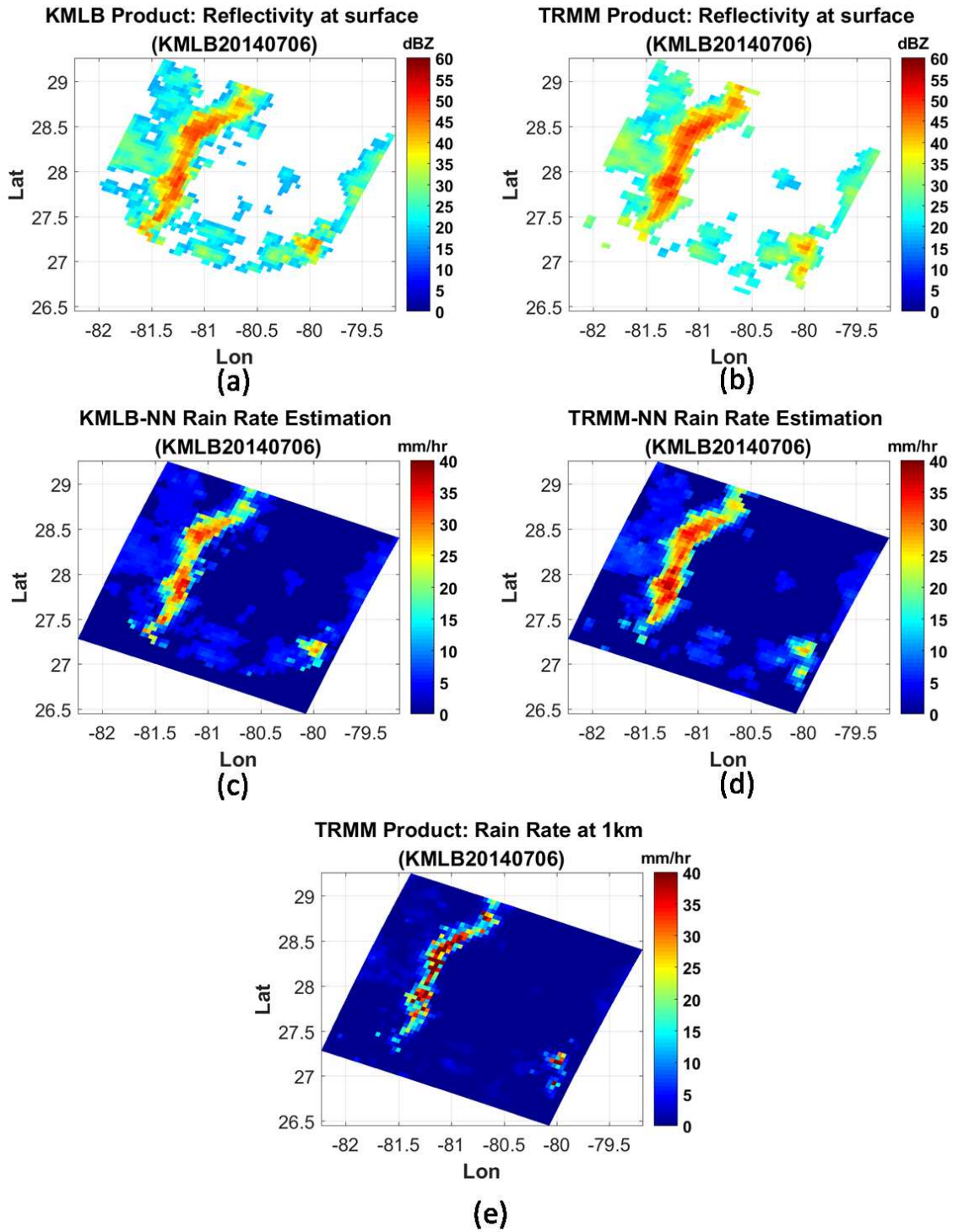


Fig. 5.16. (a) GR reflectivity at 1km height, (b) TRMM PR reflectivity at 1km height, (c) GR NN rain rate estimation, (d) TRMM NN rain rate estimation, (e) TRMM PR product rain rate estimation. (Case: 01/09/2014)



0  
 Fig. 5.17. (a) GR reflectivity at 1km height, (b) TRMM PR reflectivity at 1km height, (c) GR NN rain rate estimation, (d) TRMM NN rain rate estimation, (e) TRMM PR product rain rate estimation. (Case: 07/06/2014)

## 6 DEVELOPMENT OF THE DEEP NEURAL NETWORK MODEL FOR GPM RAINFALL ESTIMATION

### 6.1 The Two-Stage Hybrid System on GPM

As discussed in chapter 3, the TRMM satellite finished its duty in 2015 and stopped providing PR data in October 2014. The GPM core observatory satellite, launched in February 2014, took over the TRMM duty of offering global rainfall observations. In order to further investigate the potential of our new approach, the entire hybrid system has been transferred from TRMM to GPM. We continue to focus on Melbourne, Florida, as the area of study. Thus ground radar, rain gauge data, and the first stage neural network (NN1) from chapter 5 will be used in this chapter. All the components from the first stage of the system remain the same. At the second stage of the system the input dataset, which is from satellite radar, is changed from TRMM to GPM. The GPM DPR overpasses are used for rainfall estimation instead of using the TRMM PR overpasses at the region of interest. The training target is still the rainfall estimation result generated by the ground radar system. The neural network in stage two is modified to the other models because of the difference in the input dataset between GPM and TRMM. Fig. 6.1 shows the structure change in the system migration from TRMM PR to GPM DPR.

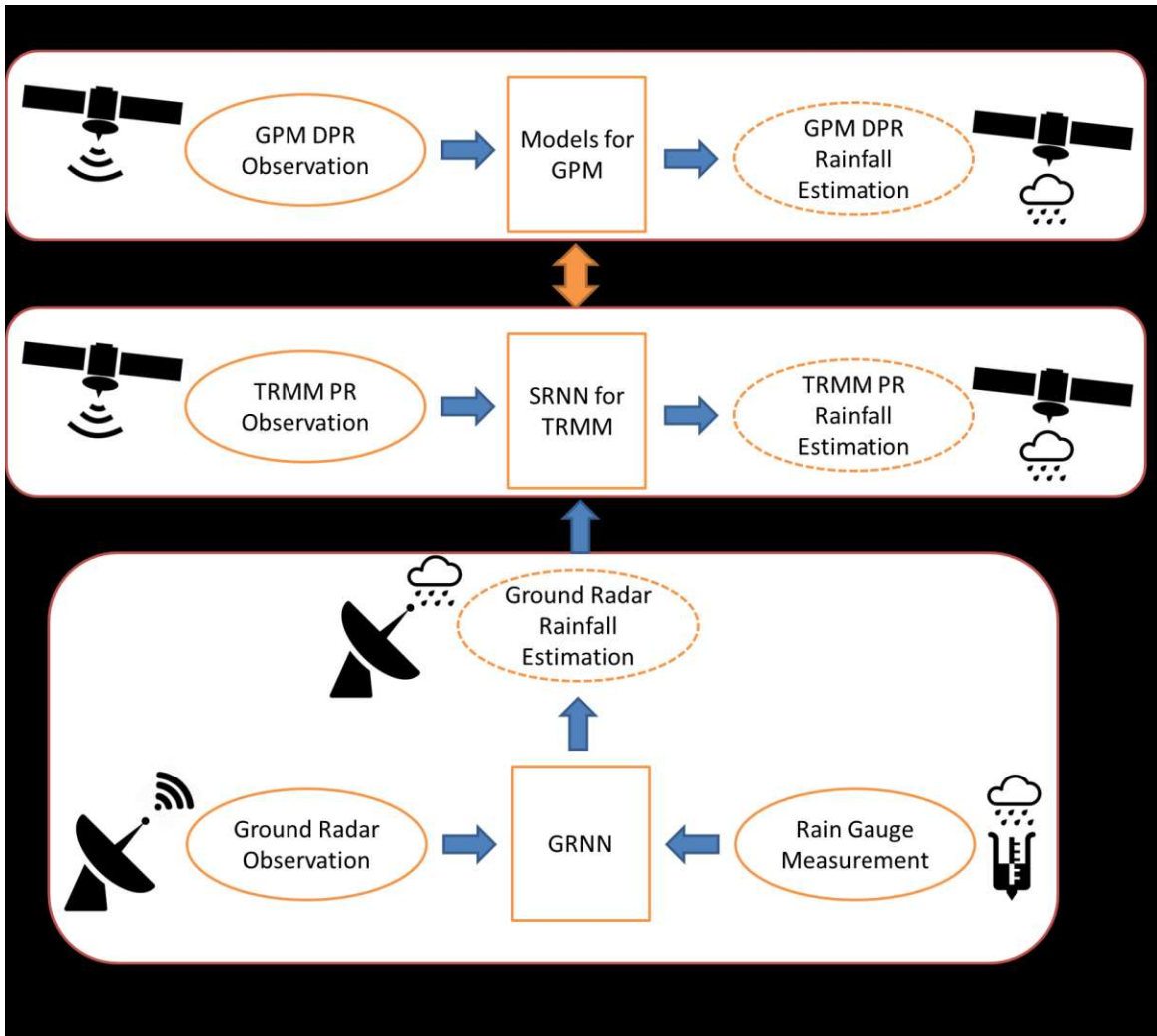


Fig. 6.1. To apply the hybrid system to GPM, the first stage remains the same. The second stage is modified to adapt the new input from GPM DPR instead of TRMM PR.

## 6.2 The Difference between GPM Observations and TRMM Observations

The most significant difference between the GPM DPR and TRMM PR observations is that GPM DPR has dual bands (Ku and Ka) and TRMM PR only has a single band. The Ku band in GPM is very similar to the TRMM PR. They have a similar frequency (13.6 GHz). Besides the Ku band, the GPM has another band, Ka band, at 35.5 GHz, which can detect light rain and snow from orbit. Because the GPM DPR have two bands with different frequencies (Ku and Ka band), two neural networks with a similar architecture but different parameters replaced the single neural network at the second stage of the hybrid system. See the system diagram in Fig. 6.2.



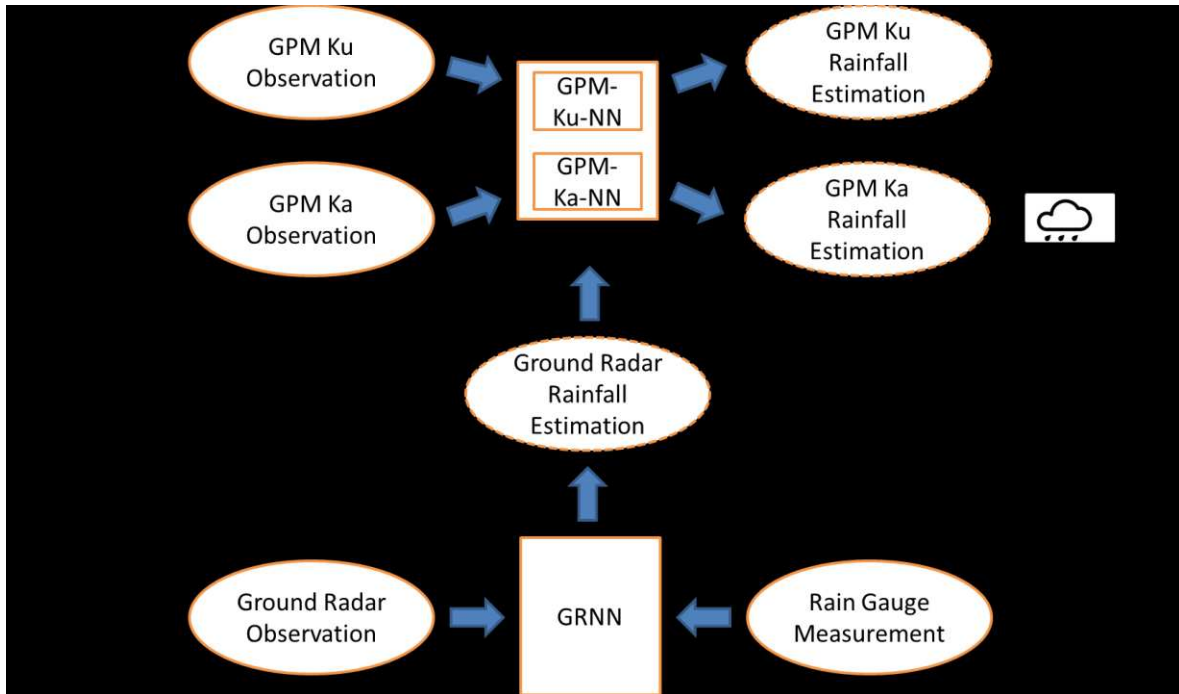


Fig. 6.2. The diagram of two-stage hybrid system for GPM rainfall estimation in Ku and Ka band. Another difference between the GPM and TRMM systems is in spatial coverage. While the

TRMM satellite only covered the tropical region from -38.0 deg to 38.0 deg latitude, the GPM can cover a much wider area, from -70.0 deg to 70.0 deg latitude. This large area decreases the number of overpasses from the GPM at a single location as compared to TRMM. As we mentioned in chapter 5, TRMM provided 44 overpasses for a precipitation case while GPM only covered 7 cases for our research in 2014. The limited number of cases causes fewer data samples for our training and evaluation process. Below is table 6.1, which describes the differences between the datasets used for training and evaluation between the GPM and TRMM observations.

Table 6.1. The sample size information between TRMM PR and GPR DPR.

	TRMM PR	GPM Ku	GPM Ka
Site	KMLB		
Year	2014		
Data coverage duration	1~10	3~12	3~12
SR-GR dataset size	5416	712	723
System validation dataset size	106	27	32

### 6.3 GPM Neural Network Implementation

The entire implementation process is similar to that described in chapter 5, section 4. The only difference is that the neural networks for GPM Ku and GPM Ka bands will be trained and validated separately. Fig. 6.3 shows the scatter plot density between the SR and GR neural networks.

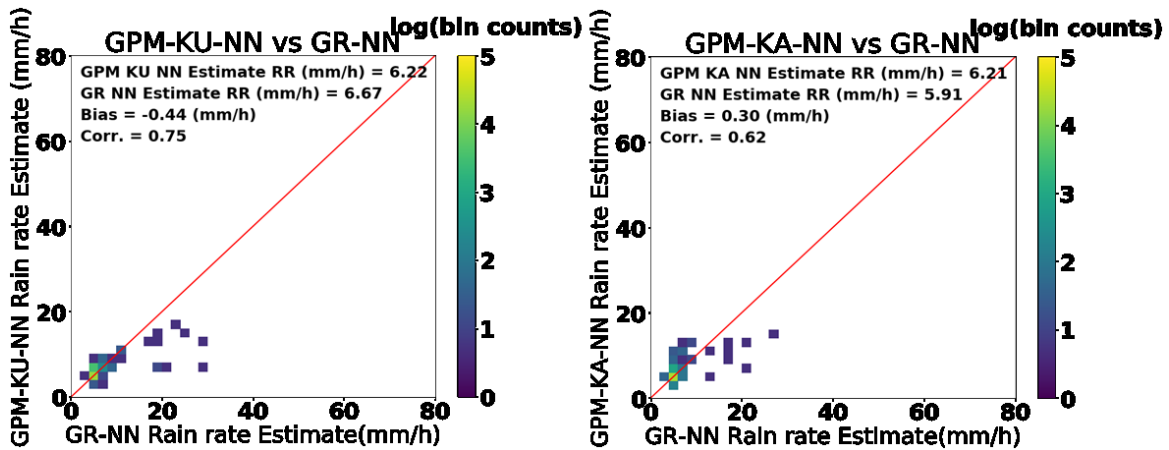


Fig. 6.3. Validation results between the GR NN and SR NN in 2014 (left: Ku, right: Ka).

From fig. 6.3, it can be seen that the bias is almost the same, and the GPM Ku neural network shows a slightly better performance in correlation than GPM ka. This slight improvement could be considered as the difference between the dataset samples in Ku and Ka bands.

### 6.4 System Validation

The system validation process was described in chapter 5, section 5. For system validation, the GPM DPR observations (including both Ku and Ka band), GPM DPR products, ground radar observations, and rain gauge measurements were all collected from the overpass cases during 2014 over Melbourne,

Florida. The data alignment process in SR and GR for two bands in GPM is similar to the process for TRMM. As mentioned above, the GPM only had seven overpass with rain events in that year. Three of the seven can be considered good cases that offered more than 80% of the data points for training the SR neural networks. The limited number of overpass cases directly restricted the size of the validation dataset. The GPM DPR rain rate product used for comparison was the surface rain rate derived by that parametrical method from the GPM DPR L2 products. Table 6.1 shows that the sample sizes of validation for the datasets were 27 for Ku band and 32 for Ka band, which is extremely small for validation purposes.

Tables 6.2 and 6.3 and figs. 6.4 and 6.5 show that the performance of the hybrid system is better than the performance of the GRM product in terms of lower bias and higher correlations at both Ku and Ka bands. Notice that the ground radar neural network has a relatively higher bias compared to the satellite radar neural network. It indicates our hybrid system has the ability to make automatic bias corrections between these two neural networks.

Table 6.2. Performance comparison between GR NN, GPM Ku NN, and GPM ku products (2014).

	GR NN	GPM Ku NN	GPM Ku Product
RainGauge RR (mm/h)	5.48		
Estimate RR (mm/h)	8.22	6.63	3.61
Bias (mm/h)	2.74	1.14	-1.87
Corr.	0.85	0.88	0.25

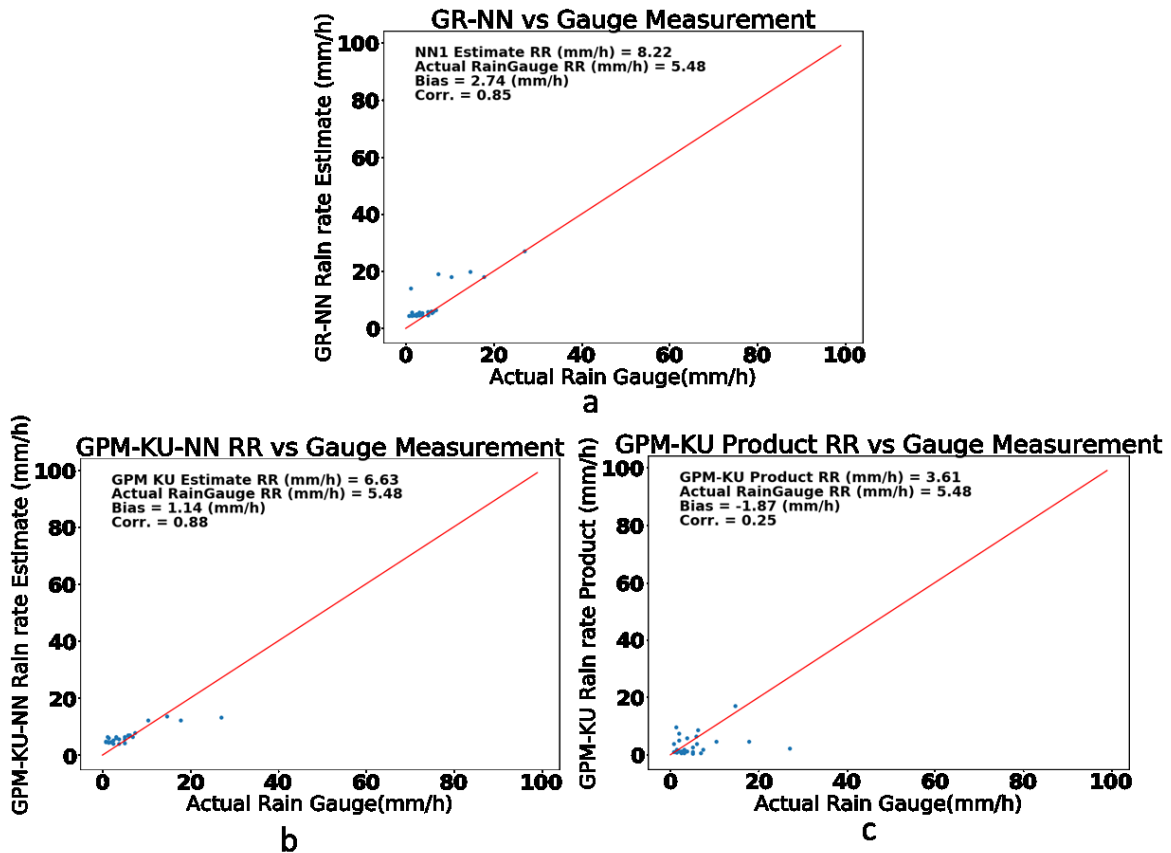


Fig 6.4. Scatter plot of system validation results for 2014: a) Ground radar NN estimation vs. gauge measurement, b) GPM ku NN estimation vs. gauge measurement, c) GPM product vs. gauge measurement.

Table 6.3. Performance comparison between GR NN, GPM Ka NN, and GPM ka products (2014).

	GR NN	GPM Ka NN	GPM Ka Product
RainGauge RR (mm/h)	6.09		
Estimate RR (mm/h)	8.98	7.40	4.32
Bias (mm/h)	2.89	1.31	-1.77
Corr.	0.82	0.89	0.62

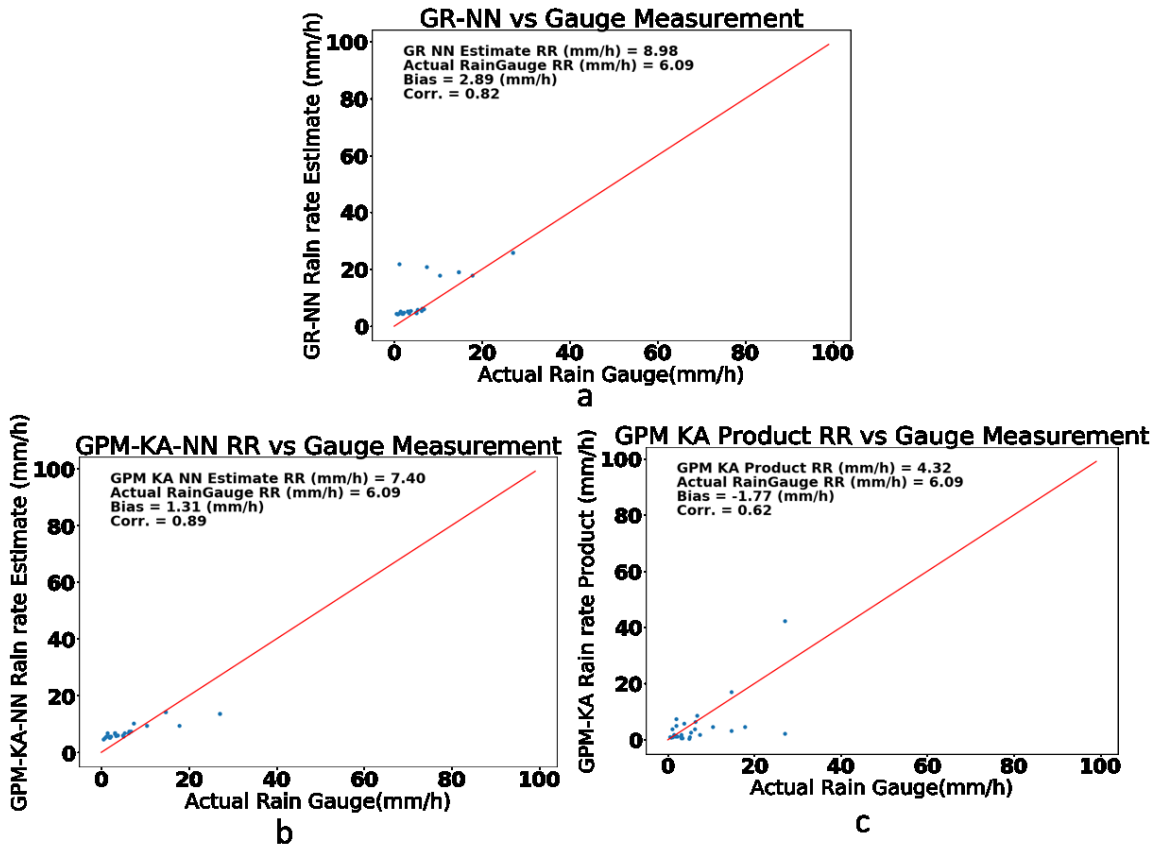


Fig. 6.5. Scatter plot of system validation results for 2014: a) Ground radar NN estimation vs. gauge measurement, b) GPM ka NN estimation vs. gauge measurement, c) GPM product vs. gauge measurement.

## 6.5 Case Study

The GPM overpass at July 22, 2014, with orbit number 002249, was selected as the case with which to study the performance of that hybrid system. In this study, an instantaneous rain rate map would be derived from Ku and Ka neural networks separately, as shown in Fig. 6.6.

The top row of the figure shows both reflectivity factor and rainfall estimation maps from ground radar. The second and third rows show the similar item from GPM Ku and Ka bands. The last row shows the rainfall map from the GPM standard product for comparison purposes. From these figures, the rainfall maps generated by the hybrid neural networks prove to be a better representation of the precipitation event as compared to the GPM DPR standard rainfall product. The GPM product tends to underestimate the rainfall while the neural network technique captures the storm with a small bias. The figures also demonstrate that outputs from both networks at Ku and Ka are very close to each other, and that the GPM-DPR standard product tends to underestimate rain rate compared to the hybrid system.

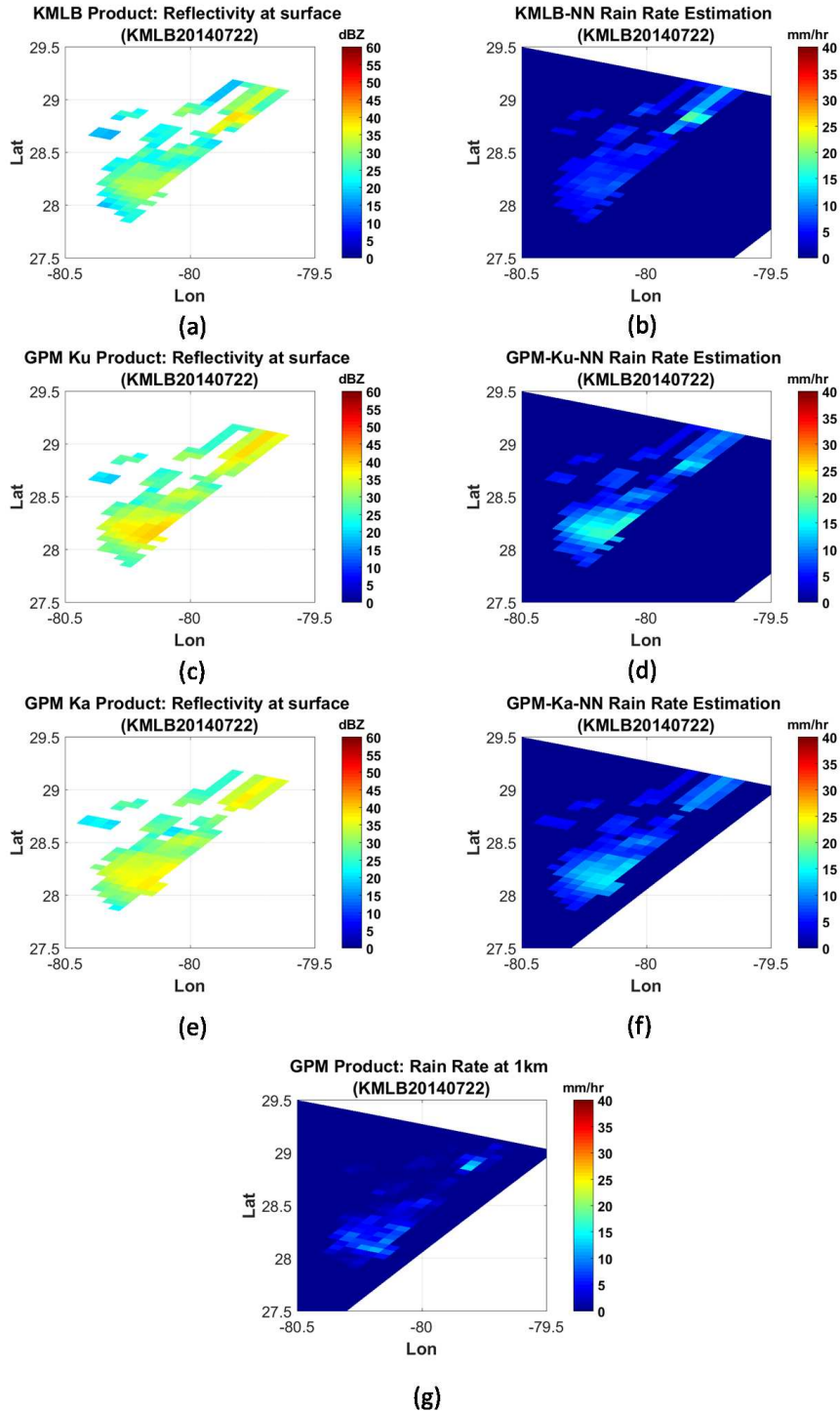


Fig. 6.13. (a) GR reflectivity at 1km height, (b) GR rainfall estimation at 1km height, (c) GPM Ku reflectivity at 1km height, (d) GPM Ku NN rain rate estimation, (e) GPM Ka reflectivity at 1km height, (f) GPM Ka NN rain rate estimation, (g) GPM DPR product rain rate estimation. (Case: 07/22/2014)

## 7 SUMMARY AND FUTURE PLAN

### 7.1 Summary

In this research, a hybrid system based on a machine learning technique was designed to estimate rainfall from ground radar and space radar observations. The key technique was the deep neural network called a multi-layer perceptron.

This hybrid system has two stages of approach to rainfall estimation. The first stage is a neural network built from ground radar observations and rain gauge measurements. This approach can provide an alternative method for carrying out a ground radar validation process. Data from observations during 2005, 2009, and 2014 over the KMLB site in Melbourne, Florida, were used to evaluate the performance of this approach in relation to direct rain gauge measurements and compare it to the traditional parametric rainfall estimation algorithm (Z-R relation). It was found that the neural network performance was better in most of cases due to less bias, lower NSE and RMSE, and higher correlation. The Z-R relation underestimated the rain rate and was unable to capture the storm variations in most cases.

The second approach was based on a hybrid system with a two-stage neural network. The first network was designed based on ground radar observations and rain gauge measurements. This network was used to map the relation between the ground radar reflectivity vertical profiles and the rain gauge measurements as the first stage of a two-stage hybrid network where TRMM observations were aligned with the ground radar observations that had been used in the first stage to train a second network with rainfall estimated based on the ground radar measurements from the first network as a target.

Subsequent to the training process, the evaluation process was performed among the ground radar estimations, the TRMM PR estimations, and the TRMM PR product and rain gauge measurements



during 2005, 2007, 2009, and 2014 at the KMLB site. It was shown that the performance of this hybrid system is better than the performance of the TRMM PR rain rate standard product when both were compared with the rain gauges. The TRMM PR product tends to underestimate the rain rate. When comparing the performance of the two stages of the neural networks to each other, it can be shown from their high correlations that both have a close performance. This indicates that the first neural network successfully passes the relation between the radar reflectivity factor and the rain gauges to the second neural network. The mapping was done by aligning the ground radar observations with the TRMM-PR observations.

In addition, the hybrid system was used to generate local rainfall maps. These instantaneous rainfall maps were generated by applying the observations from the ground radar or TRMM. The ground radar data were applied to the network that was built based on ground radar observation, while the TRMM PR data were applied to the network that was built based on the TRMM observations. The rainfall maps generated by both neural networks were very similar to each other and were shown to achieve a better representation than the TRMM product through higher correlation and less bias.

In our second approach, we transferred the process from TRMM PR to GPM DPR to investigate the possibility of using machine learning for rainfall estimation on other satellite radars. A similar training, testing, and validating process was completed on the GPM Ku and Ka datasets. Due to the limited sample size, we cannot conclude that the second approach worked perfectly in the GPM DPR. However, the results have shown the great potential of our hybrid system in space radar rainfall estimation.

## 7.2 Future Plan

Several topics are suggested for further study. They can be divided into two main points: the first one improving the model's performance and the second validating the model's performance.

### 7.2.1 Performance Improvement

First, we plan to use the model ensemble technique. This could be achieved by designing different neural networks based on different data sets for rain rate estimation and then summing the output of each network with different weights. We are going to call this the “ensemble neural network” technique.

Second, we plan to add more input features to bring additional information into the rainfall estimation. Those features include wind direction, wind speed, rainfall type, reflectivity profile at the neighborhood in space, the reflectivity profile at the last time step, and so on.

Third, we plan to use another type of neural network. So far the convolutional neuron network (CNN) and the recurrent neuron network (RNN) have been suggested. The CNN is known to have an amazing performance in computer vision problems. The RNN is known to very efficiently process time series data. By combining the two techniques, we will be able to estimate the rainfall rate from the time series 3D reflectivity observations.

Last, we plan to develop a model based on unsupervised learning such as the self-organizing map or auto-encoder in order to detect the existence of rain before going to estimation.

### 7.2.2 Performance Validation

For performance validation, we plan to the validate hybrid system based on other ground validation radars. Suggested ground radars are the KFWS WSR-88DP radar in Texas and the CSU-CHILL radar in Colorado as well as the NASA NPOL radar data collected during the Olympic Mountain Experiment (OLYMPEX) on the Olympic Peninsula of Washington State. The additional three domains can bring different climatological properties into our study, which can test model generalization in other areas.

## REFERENCES

Atlas, D., and C. Ulbrich, "Path- and area-integrated rainfall measurement by microwave attenuation in the 1–3 cm band," *Journal of Applied Meteorology*, vol. 16, no. 12, pp. 1322-1331, 1977.

Battan, L., *Radar Observation of the Atmosphere*. Chicago, The University of Chicago Press, 1973.

Bolen, S., and V. Chandrasekar, "Methodology for aligning and comparing spaceborne radar and ground-based radar observations," *Journal of Atmospheric and Oceanic Technology*, vol. 20, no. 5, pp. 647-659, 2003.

Chandrasekar, V., A. Alqudah, and Y. Wang, "Hybrid neural network technique to estimate rainfall from TRMM measurements," In *IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium*, Boston, MA, pp. IV - 291-IV - 294, 2008.

Goodfellow, I., Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, Massachusetts: The MIT Press, 2016.

GPM Data Utilization Handbook, <http://www.eorc.jaxa.jp>, 2017. [Online]. Available: [http://www.eorc.jaxa.jp/GPM/doc/data\\_utilization/GPM\\_data\\_util\\_handbook\\_E.pdf](http://www.eorc.jaxa.jp/GPM/doc/data_utilization/GPM_data_util_handbook_E.pdf). [Accessed: 09-Sep- 2017].

GPM Data Utilization Handbook, <http://www.eorc.jaxa.jp>, 2017. [Online]. Available: [http://www.eorc.jaxa.jp/GPM/doc/data\\_utilization/GPM\\_data\\_util\\_handbook\\_E.pdf](http://www.eorc.jaxa.jp/GPM/doc/data_utilization/GPM_data_util_handbook_E.pdf). [Accessed: 09-Sep- 2017].

GPM Data Utilization Handbook, 2017. [Online]. Available: [http://www.eorc.jaxa.jp/GPM/doc/data\\_utilization/GPM\\_data\\_util\\_handbook\\_E.pdf](http://www.eorc.jaxa.jp/GPM/doc/data_utilization/GPM_data_util_handbook_E.pdf). [Accessed: 09-Sep- 2017].

GPM/DPR Level-2 Algorithm Theoretical Basis Document, 2016. [Online]. Available: [http://www.eorc.jaxa.jp/GPM/doc/algorithm/ATBD\\_DPR\\_V5\\_201705final.pdf](http://www.eorc.jaxa.jp/GPM/doc/algorithm/ATBD_DPR_V5_201705final.pdf). [Accessed: 09- Sep- 2017].

Iguchi, T., T. Kozu, R. Meneghini, J. Awaka, and K. Okamoto, "Rain-profiling algorithm for the TRMM precipitation radar," *Journal of Applied Meteorology*, vol. 39, no. 12, pp. 2038-2052, 2000.

JAXA | Global Precipitation Measurement/Dual-frequency Precipitation Radar (GPM/DPR), JAXA | Japan Aerospace Exploration Agency, 2017. [Online]. Available: <http://global.jaxa.jp/projects/sat/gpm/>. [Accessed: 09- Sep- 2017].

LeCun, Y., Y. Bengio, and G. Hinton, "Deep learning." *Nature*, vol. 521, no. 7553, pp. 436-444, 2015.

Li. W., and V. Chandrasekar, "Rainfall estimation from vertical profiles of reflectivity using neural networks." *In IGARSS 2002 - 2002 IEEE International Geoscience and Remote Sensing Symposium*, IEEE International, vol. 6, pp. 3477-3479, 2002.

Li, W., and V. Chandrasekar, "Investigations in radar rainfall estimation using neural networks", *In IGARSS 2003 - 2003 IEEE International Geoscience and Remote Sensing Symposium*, IEEE International, vol. 4, pp. 2347-2349 vol.4, 2003.

Liu, H., V. Chandrasekar, and G. Xu, "An adaptive neural network scheme for radar rainfall estimation from WSR-88D observations," *Journal of Applied Meteorology*, vol. 40, no. 11, pp. 2038-2050, 2001.

Marshall, J., and W. Palmer, "The distribution of raindrops with size," *Journal of Meteorology*, vol. 5, no. 4, pp. 165-166, 1948.

Mitchell, T. *Machine Learning*. New York: MacGraw-Hill, 1997.

National Weather Service Doppler Radar Images, Radar.weather.gov, 2017. [Online]. Available: [https://radar.weather.gov/index\\_lite.htm](https://radar.weather.gov/index_lite.htm). [Accessed: 09- Sep- 2017].

NEXRAD Technical Information, Roc.noaa.gov, 2017. [Online]. Available: <https://www.roc.noaa.gov/wsr88d/Engineering/NEXRADTechInfo.aspx>. [Accessed: 09- Sep- 2017].

Orlandini, S., and I. Morlini, "Artificial neural network estimation of rainfall intensity from radar observations," *Journal of Geophysical Research: Atmospheres*, vol. 105, no. 20, pp. 24849-24861, 2000.

Rongrui, Xiao, and V. Chandrasekar, "Multiparameter Radar Snowfall Estimation Using Neural Network Techniques " In IGARSS 1996 - 1996 IEEE International Geoscience and Remote Sensing Symposium, pp. 566-568, 1996.

Rongrui, Xiao, and V. Chandrasekar, "Development of a neural network based algorithm for rainfall estimation from radar observations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 1, pp. 160-171, 1997.

Rongrui Xiao, V. Chandrasekar, Liu H. and Gorgucci E.: "Detection of Rain/No Rain Condition on ground From Radar Data Using A Kohonen Neural Network". *Geoscience and Remote Sensing Symposium*, 1998.

Skolnik, M., *Introduction to Radar Systems*. Boston: McGraw-Hill, 2007.

Teschl, R., W. Randeu, and F. Teschl, "Improving weather radar estimates of rainfall using feed-forward neural networks," *Neural Networks*, vol. 20, no. 4, pp. 519-527, 2007.

Trmm-fc.gsfc.nasa.gov, 2017. [Online]. Available: [https://trmm-fc.gsfc.nasa.gov/trmm\\_gv/data/gv\\_maps/MELB\\_map.gif](https://trmm-fc.gsfc.nasa.gov/trmm_gv/data/gv_maps/MELB_map.gif). [Accessed: 09- Sep- 2017].

TRMM Data User Handbook, <http://www.eorc.jaxa.jp>, 2006. [Online]. Available: [http://www.eorc.jaxa.jp/en/hatoyama/satellite/satdata/trmm\\_pdf/trmm\\_handbook\\_e.pdf](http://www.eorc.jaxa.jp/en/hatoyama/satellite/satdata/trmm_pdf/trmm_handbook_e.pdf). [Accessed: 09- Sep- 2017].

TRMM-PR Manual, Version 7, <http://www.eorc.jaxa.jp>, 2011. [Online]. Available:

[http://www.eorc.jaxa.jp/TRMM/documents/PR\\_algorithm\\_product\\_information/pr\\_manual/PR\\_Instruction\\_Manual\\_V7\\_L1.pdf](http://www.eorc.jaxa.jp/TRMM/documents/PR_algorithm_product_information/pr_manual/PR_Instruction_Manual_V7_L1.pdf). [Accessed: 09- Sep- 2017].