Thesis for the Degree of Doctor of Philosophy

DNN-based accuracy improvement for the rain rate from microwave satellites and the blending of the multi-satellite rain rate data using a modified Bayesian ensemble

by
Kwangjin Kim

Department of Spatial Information Engineering
The Graduate School
Pukyong National University

August, 2020

DNN-based accuracy improvement for the rain rate from microwave satellites and the blending of the multi-satellite rain rate data using a modified Bayesian ensemble

(심층신경망 기반의 다중위성 강우강도 개선 및 베이지안 앙상블을 이용한 합성장 산출)

Advisor: Prof. Yang Won Lee

by

Kwangjin Kim

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in Department of Spatial Information Engineering, The Graduate School, Pukyong National University

August, 2020

DNN-based accuracy improvement for the rain rate from microwave satellites and the blending of the multi-satellite rain rate data using a modified Bayesian ensemble

A dissertation
by
Kwangjin Kim
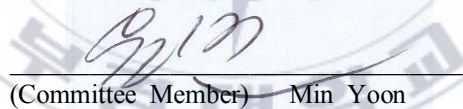
Approved by

_____
(Committee Chair)   Chul-Uong Choi

_____
(Committee Member)   Kyung-Soo Han

_____
(Committee Member)   Min Yoon

_____
(Committee Member)   Jeail Cho

_____
(Committee Member)   Yang-Won Lee

August, 2020

Contents

List of Figures

List of Tables

List of Acronyms

ACC	Accuracy
AI	Artificial Intelligence
AMSU-B	Advanced microwave sounding unit B
ANN	Artificial Neural Network
BMA	Bayesian model averaging
bRR	Blended rain rate
BT	Brightness temperature
CC	Correlation coefficient
CDF	Cumulative density function
CMORPH	Climate prediction center morphing technique
CMORPH-KF	CMORPH Kalman filter
CRM	Cloud-resolving model
CV	Cross validation
DMSP	Defense meteorological satellite program
DNN	Deep Neural Network
DPR	Dual-frequenccy precipitation radar
EBMA	Ensemble Bayesian model averaging
ECDF	Empirical cumulative distribution function
ECMWF	European centre for medium-range weather forecasts

FAR	False alarm rate
GEO	Geosynchronous Earth orbiting
GMI	Global precipitation measurement microwave imager
GPM	Global precipitation measurement
GPU	Graphic processing unit
GPROF	Goddard profiling algorithm
HSS	Heidke skill score
IMERG	Integrated multi-satellite retrievals for GPM
IR	Infrared
IWC	Ice water content
IWP	Ice water path
JAXA	Japan aerospace exploration agency
KMA	Korea meteorological administration
LEO	Low Earth-orbiting
MAE	Mean absolute error
MBE	Mean bias error
Metop	Meteorological operational satellite
MHS	Microwave humidity sounder
MW	Microwave
NASA	National Aeronautics and space administration
NESDIS	National environmental satellite, data and

	information
NOAA	National oceanic and atmospheric administration
NRMSE	Normalized root mean squares error
PCT	Polarization corrected temperature
PDF	Probability density function
PERSIANN	Precipitation estimation from remotely sensed information using artificial neural network
PERSIANN-CCS	PERSIANN-Cloud classification system
PMW	Passive Microwave
POD	Probability of detection
PPS	Precipitation processing system
RMSE	Root mean square error
RTM	Radiative transfer model
SC ratio	Stratiform-convective rain ratio
SI	Scattering index
SSMIS	Special sensor microwave imager/sounder
ReLU	Rectified linear unit
TMI	TRMM microwave imager
TRMM	Tropical rainfall measuring mission

심층신경망 기반의 다중위성 강우강도 개선 및 베이지안 앙상블을 이용한 합성장 산출

김 광 진

부 경 대 학 교　　대 학 원　　공 간 정 보 시 스 템 공 학 전 공

요　　약

강수는 수문 순환에 중요한 요소이며, 인간과 생태계에 밀접한 영향을 준다. 강수의 관측은 소규모 변동성이 매우 크고 비정상정인 통계적 변동성을 보인다. 따라서 정확하고 시공간적으로 조밀한 모니터링이 요구된다. 위성을 이용한 강우강도는 이러한 모니터링에 효율적인다. 지상에서 관측이 어려운 곳과 넓은 영역의 모니터링을 가능하게 한다. 위성의 강우강도의 산출은 수동 마이크로파(passive microwave, PMW) 센서를 널리 사용하고 있다. 이는 적외채널이나 가시광선채널에 비해 구름에 대한 투과율이 높기때문에 더 직접적인 관측이 가능하기 때문이다.

하지만 기존의 PMW를 이용하여 강우강도를 산출하는 알고리듬은 밝기온도 및 밝기온도로 구해지는 몇몇 지수에 대한 회귀분석을 통해 산출되거나 복사전달모델과 구름용해모델을 통해 대기상태를 모의한 후 베이지언 이론으로 산출되었다. 회귀분석을 이용한 통계적 기법은 비구름과 지표면의 강우에 대한 복잡하고 비선형적인 관계를 충분히 설명하기엔 제한적이었으며, 모델 자료를 이용한 물리적 기반의 강우강도 산출은 대기상태의 모의가 어렵고, 해양에 국한되어 왔다.

또한 PMW 센서는 대부분 저궤도 위성에 탑재되어, 단일 위성 혹은 센서를 이용하기에는 전지구적인 관측을 위해서 시간적, 공간적인 탐지영역에 대해 제한적이다. 따라서 다중위성 혹은 센서를 이용한 다중 위성 강우강도 합성장들이 개발 및 배포되고 있지만, 이는 중첩지역에 대해서 개별 자료에 대한 편차 보정 후 자료를 평균, 중위수 또는 최근린 값으로 사용하고 있다. 이는 개별 위성 및 센서의 불확실성의 차이를 고려하지 않는 방법으로 오히려 그 정확도를 감소시킬 수 있다.

따라서 본 연구의 목표로는 비구름과 지표면 사이의 복잡하고 비선형적인 관계를 설명하기 위해 심층신경망(Deep neural network)을 이용하여 위성강우강도 산출물을 개선하고자 한다. 또한 다중위성의 합성시 베이지언 모형 평균화(Ensemble Bayesian model

averageing, EBMA)를 이용하여 중첩지역에 대한 불확실성의 차이를 고려한 가중치 설정과 가중평균을 통해 합리적인 합성을 통해 관측영역의 확대와 중첩지역의 이중의 정확도 개선을 목표로 하였다.

DNN을 적용한 PMW센서 기반의 강우강도 산출물의 개선을 위해서 본 연구에서는 강우여부 판단, 강우강도 산출에 대한 2가지의 DNN 모델을 생성하고 이를 결합하여 강우강도를 개선하였다. 개선 대상 센서로는 Global precipitation measurement (GPM) microwave imager (GMI), Special sensor microwave imager/sounder (SSMIS), Microwave humidity sounder (MHS) 센서를 선정하였다. 3개 위성을 탑재하고 있는 8개 위성에서 자료를 취득하여 실험을 진행하였다. 반복 수행을 통해 2가지 DNN 모델의 최적화를 수행하였으며, 이를 통해 산출된 자료와 Goddard profiling algorithm (GPROF)를 통해 산출되는 강우강도 산출물과의 정확도를 비교 분석하였다. 강우여부 판단에서는 4.3%의 개선을 보였으며 강우강도 산출에서는 17.4%의 정확도 개선을 확인하였다.

EBMA를 이용한 합성장 산출을 위해 선행연구에서와 같이 로지스틱회귀와 감마분포의 결합분포로 확률밀도함수를 구성하였다. 하지만 선행연구들과 달리, 위성 강우강도에 적절히 적용하기 위해 EBMA를 변형 및 개선하여 사용하였다. 경험 누적확률분포를 이용한 합성멤버들의 일차적인 편의보정을 추가하였으며, 로지스틱회귀와 감마분포에서의 추정값의 보정계수로 사용된 지수를 불필요를 확인하고 이를 제거하였다. 그리고 가중 평균된 EBMA 확률밀도함수로부터 강우강도를 추정하기 위한 적절한 백분위수를 반복 수행 통해 최적의 백분위수를 선정하였다. 그 결과 합성장 생성시 합성멤버의 탐지영역을 모두 포함하지만, 중첩지역에 대해서 불연속적인 부분이 보이지 않으며 중첩지역에서는 입력자료인 결합DNN의 강우강도 자료보다 개선된 오차를 보였다. 또한 널리 사용되는 Intergrated multi-datellite retrievals for GPM (IMERG) 자료에 비해 44.2%의 뚜렷한 개선을 확인하였다. 또한 최종적인 합성장의 중첩지역에서는 DNN 강우강도 자료에서 다시 한번 정확도개선의 효과를 보임을 확인할 수 있었다.

CHAPTER 1 INTRODUCTION

1.1 Backgrounds

Precipitation is an important factor in the hydrological cycle, and global distribution and intensity are essential for understanding and feedback of the Earth system(Sanò et al., 2015). Therefore, precipitation and the intensity has been measured and estimated in a variety of ways. In addition, the quality of precipitation data is very important for hydrological studies because the error in observation and estimation of precipitation is the decisive factor that cause uncertainties in hydrologic models (Gottschalck et al., 2005; Jiang et al., 2012; Sanò et al., 2015; Tao et al., 2018).

The existence of precipitation shows a discontinuous spatial distribution and the variability of existence and intensity is large(Huffman et al., 2007). Although gauges are used for observing the intensity of precipitation on the surface of the earth, lack of the number, spatial distribution, and representativeness of observation space is insufficient to understand global precipitation characteristics and patterns. The estimate by remote sensing is used for measurement of the global precipitation. The estimation of precipitation using remote sensing is mainly based on ground radar and satellite.

The satellite-based estimates are generally obtained using channels of infrared (IR) or microwave (MW). The estimates of precipitation using the microwave are more direct than using IR channels. This is because the character of the IR channel estimates of the precipitation using only the information in the upper layer of the cloud. The MW observations have the great advantage of providing a more direct measurement of the precipitation due to the ability of MW radiation to penetrate precipitating clouds and interact with its liquid and ice hydrometeors (Mugnai et al., 1990; wilheit et al., 1994; Weng and Grody, 2000; Bennartz and Petty, 2001; Bauer et al., 2005, Sanò et al., 2015).

Active microwave sensors, called radar, allow more direct measure of precipitation from space than passive microwave (PMW) sensors. However, the spatial coverage of active microwave sensors is narrower than the PMW sensor. Thus, PMW are most commonly used to estimate global precipitation. The PMW techniques for the estimation of precipitation have seen advances, due largely to the increased number of radiometers available, with improved sensing capabilities and due to several theoretical studies on microwave radiative transfer modeling through precipitating clouds (Mugnai et al., 1993; wilheit et al., 1994; Smith et al., 1998, 2002; Stephens and Kummerow. 2007; Skofronick-Jackson and Johnson, 2011)

The estimation of precipitation using PMW has largely been used for statical and physical methods. The statical methodology is to calculate the rain rate by using the indexes such as polarization corrected temperature (PCT), scattering index(SI) and ice water path (IWP) obtained by using the brightness temperatures (BTs) of the sensor(Spencer et al., 1989; Ferraro and Marks, 1995, Di Tomaso et al., 2009). However, it is difficult to say that the accuracy of satisfactory rainfall intensity is expressed.

Recently, the National aeronautics and space administration (NASA) Precipitation processing system (PPS) uses the Goddard profiling (GPROF) retrieval scheme to calculate the rainfall intensity from satellites equipped with PMW sensors as a part of the global precipitation measurement (GPM) mission(Hou et al., 2014; Skofronick-Jackson et al., 2017). The GPROF scheme is basically based on Bayesian approach, and generates the conditional probability density function (PDF) between precipitation and BTs of each channel from the previous data to estimate the rainfall intensity (Kummerow et al., 2001). In version 5, it is fully parameterized (Kummerow et al., 2015). However, this algorithm basically needs ancillary data such as the temperature and surface type, and has difficulty in calculating the PDF. The PDF is derived from a reflectance transfer model using a database of previous auxiliary data and satellite BT data,

which are then parameterized and used. This may have the disadvantage that it is difficult to immediately update the parameter.

Another approach to satellite-based rainfall intensity calculation is to use artificial neural network (ANN) (Hsu et al., 1997, 2018; Yao et al., 2001; Sarma et al., 2008; Sanò et al., 2015). ANN is an algorithm that mimics the structure of a human neural network. The neural network of supervised learning builds a neural network through learning rules. Through iteration, the weights and errors of nodes and links are adjusted. It has the advantage of being able to express or simulate complex and nonlinear relationships by adjusting the hidden layer. It has the advantage of being able to express or simulate complex and nonlinear relationships by adjusting the hidden layer. However, the deeper the hidden layer in order to express this relationship, the longer the time for computing model training because of local minima problem by gradient vanishing. In addition, ANN is a kind of machine learning, and it also has the problem of overfitting machine learning. Overfitting means that errors in training data are reduced due to repetitive learning of the same training data, but errors in new data or verification data due to excessive training increase with training (Goodfellow et al., 2016).

Recently, Deep neural network (DNN) has emerged in the ANN field to express complex nonlinear relationships. DNN

complements this by using various active functions such as Rectified linear unit (ReLU) instead of sigmoid to solve the problem of gradient vanishing. In addition, in order to solve the overfitting problem, a model with the smallest error in verification data is selected in various normalization techniques or training courses. In statical and physical methodologies, rainfall intensity is calculated based on scattering and emissivity. However, the microscopic characteristics between ice and water droplets in the upper layer of clouds and rainfall intensity on the surface are difficult to express in a linear relationship. The ANN has the advantage of being able to express these nonlinear relationships (Sanò et al., 2015). Hsu et al.(1997, 1999) introduced Precipitation estimation from remotely sensed information using artificial neural network (PERSIANN), which uses ANN to calculate rainfall intensity from IR bands of geostationary satellites (Hsu et al., 1999; Scorooshian et al., 2000). Based on this, Hong et al.(2004) further developed the PERSIAN-Cloud Classification System (PERSIANN-CCS), a high-resolution cloud detection algorithm. In addition to geostationary orbit, the polar orbit satellites have been used to retrieve rainfall using ANN technique. The SI and PCT indices, which are used for the empirical method using the linear regression, are used as input data to suggest the possibility of successful application of the ANN approach(Mahesh et al., 2011; Panegrossi et al., 2013; Sanò

et al., 2015, Tan et al., 2018).

The most PMW sensors are mounted on low Earth-orbiting (LEO) satellites. Observation of rain rate using a single satellite has limitations in temporal and spatial monitoring. National oceanic and atmospheric administration (NOAA) produces and distributes synthetic fields using multiple satellites. The national environmental satellite, data and information service (NESDIS) of NOAA is providing the rain rate composite map observed from multi PMW sensor, called blended rain rate (bRR). This composite map is averaged by rainfall intensity data modified by each PMW sensor using cumulative density function (CDF) (Blended Rain Rate-Algorithm Description 2020, Mar 30). In addition, another global composite map is provided through a climate prediction center morphing technique (CMORPH) using both IR sensor and PMW sensor rain rate products. The CMORPH is a data synthesized using a time interpolation technique using motion vectors for continuous data of geostationary orbit satellites(Joyce et al., 2004). NASA also generates and distributes integrated multi-satellite retrievals for GPM (IMERG) data by synthesizing rain rate data obtained through satellites from Japan aerospace exploration agency (JAXA) and various other countries.

The reliability of this rain rate synthesis field is improved through verification evaluation with ground observation data. However, most of these satellite precipitation synthesis methods

use average, median, or recent values after timely interpolation and bias correction of data. This is unusual for the assumption that the uncertainty of satellite data is different, and it is difficult to say that rational weight distribution between members was made.

The ensemble Bayesian model averaging (EBMA) is a blending technique that considers the uncertainty differences between members, and has scalability that can be applied not only to normal distribution data such as temperature, but also to gamma distribution data such as precipitation (Raftery et al., 2005; Sloughter et. al., 2007). In the past, it was mainly used for climate model ensembles and recently began to be incorporated into blending satellite data.

1.2 Research Objectives

The purpose of this study is to improve rain rate products obtained from PMW sensor of satellite and produce the blending rain rate map of multi-sensor. More specifically, the detailed objects include addressing the following issues:

① DNN is used to establish and calculate a model for improving the rain rate.

② Calculate the blending data of multiple sensor data using EBMA.

Target sensors for constructing DNN model and retrieving rain rate were selected as GMI, SSMIS, and MHS sensors. In the case of GMI, it is a sensor on the extended line of Tropical rainfall measuring mission (TRMM) Microwave imager (TMI), and has various frequency channels for precipitation and cloud detection, thus showing high accuracy in retrieving rain rate using PMW sensors. In the case of SSMIS and MHS, they have been installed in various satellite series and holds data for a long period of time, and they are still in operation. In addition, a number of prior studies for retrieving rain rate using two sensors have been conducted.

As input data for constructing DNN model and retrieving the rain rate, brightness temperature and indexes such as SI, IWP, and SC ratio were acquired. These indexes were selected by considering them as the indices used directly for retrieving the rain rate. In addition, land-sea fraction data of the European centre for medium-range weather forecasts (ECMWF) were used as input data.

For the generation of a synthetic field using EBMA, experiments were performed to select parameters such as square number and appropriate percentile for EBMA. As the input data of EBMA, the rainfall intensity data of each sensor improved using DNN was used.

Finally, the blending map was compared with the mean and median ensemble, and IMERG data currently in operation, using the rain rate using multi-sensors, and the commonly used synthesis technique.

CHAPTER 2 LITERATURE REVIEW

2.1 Rain rate obtained from the satellite

Rain rate is observed with the various satellites. Infrared(IR) or visible(VIS) radiometers, of onboard the geosynchronous Earth orbiting (GEO) satellites, indirectly estimate rainfall based on the relationship between the cloud top temperature and surface rainfall, their observations have no direct physical connection with surface precipitation (Petty and Krajewski, 1996, Behrangi et al., 2009; Kummerow and Gilio, 1995). The observations derived from microwave (MW) provide more direct information. Because transmittance of MW about liquid and ice hydrometeors in precipitating clouds (Wilheit et al., 1994; Weng and Grody, 2000; Bennartz and Petty, 2001). The active microwave sensor has the advantage of being able to observe the size and shape of water droplets most directly. Passive microwave (PMW) has a great advantage in global rain rate due to the improvement of sensing capabilities such as increased payload and increased resolution. Various microwave radiative transfer models have been developed to convert improved PMW data to rain rate (Sephens and Kummerow, 2007; Skofronick-Jackson and Johnson, 2011). Most PMW sensors are built on low Earth-orbiting (LEO) satellites. Although various frequency bands are used, the high

transmittance 20, 36GHz bands, and 89, 150, 183GHz are used as window channels (Fig 2.1). Since this study aims at improving and synthesizing the satellite rainfall, the PMW sensor used in most satellites for the observation of rainfall data was targeted. Therefore, the study of retrieving the rain rate using the PMW sensor was intensively reviewed.

The PMW radiometers are generally divided by scanning mode. The first is cross-polarized conical scanning such as SSMIS and GMI. The second is cross-track scanning such as MHS and ATMS. Various PMW precipitation retrieval algorithms have been developed by scanning and by different conditions depending on the frequencies held by the sensors. retrieval of rain rate using PMW is classified into three types. These are (1) the "emission type" algorithms (Wilhet et al., 1991, 1994; Berg and Chase, 1992; Chang et al., 1999) use low-frequency channels to detect increases in radiance. (2) The "scattering" algorithms (Spencer et al., 1983; Grody, 1991; Ferraro and Marks, 1995) are techniques that use the correlation between reduction of radiance and precipitation caused by precipitation clouds. (3) the "multi-channel inversion" type algorithm (Oslon, 1989; Mugnai et al., 1993; Smith et al., 1994, Bauer et al., 2001; Kummerow et al., 2001). The atmospheric profile and brightness temperatures are estimated using radiative transfer model (RTM). These various algorithms are applied according to the frequency and

characteristics of the target sensor, and each algorithm has advantages and disadvantages, so it is difficult to determine which one is the "best". Each of these types of algorithms can be divided into statistical or physically retrieval algorithm based on rain rate depending on the process.
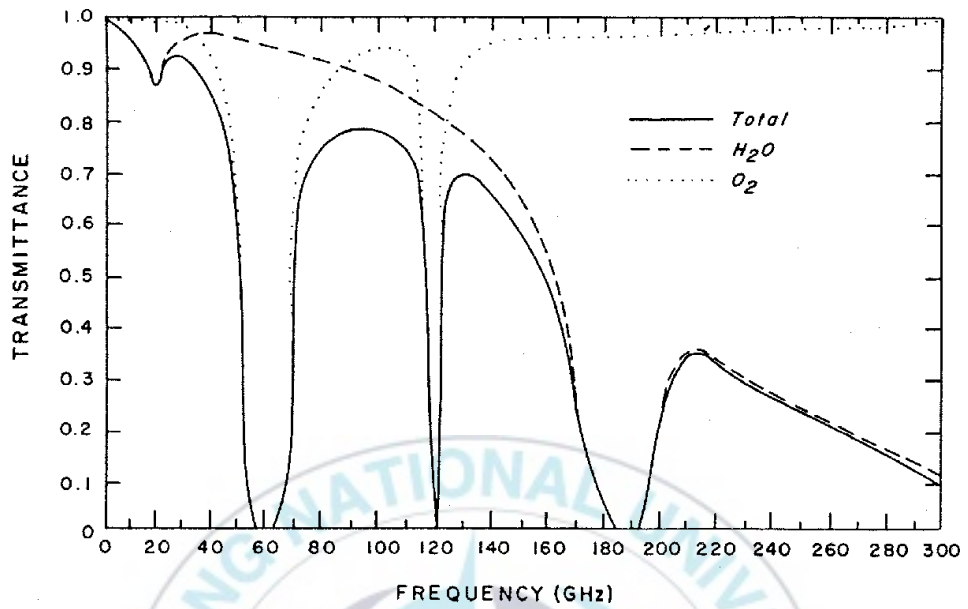
Figure 2.1. Atmospheric transmittance in the microwave region of the spectrum as a function of frequency.(Menzel, 2005)

2.1.1 Statical retrieval algorithm

The advantage of conical scanning is that it has various frequency channels and also provides polarization (V/H) information. It has been used for a long time to retrieve various rainfall from SSMI and TMI sensors, which are the previous versions of SSMIS and GMI. One of the most widely used empirical methods is to use polarization corrected temperature (PCT) defined by Spencer et al. (1989). This is an index that uses the difference between vertical polarization and horizontal polarization at specific frequency (around 89 or 36GHz). PCT is finally used to retrieve the rain rate through an empirical linear regression.

Scattering index (SI) means the difference between the estimated value using the low-frequency BT and the actual observation value with respect to the BT near 89GHz. 19 and 22GHz with high transmittance are used as the low frequency range. This is a kind of statistical technique that uses scattering around the 89GHz frequency, which is a frequency that greatly affects liquid hydrometeors due to the presence of precipitating clouds.

The above two statistical-based algorithms are limited in application to cross-track scanning sensors such as MHS. First, MHS sensors do not have low frequencies and also have only one polarization. retrieval of rain rate derived from MHS uses ice

water path (IWP)(Weng et al., 2003; Zhao and Weng., 2002; Qiu et al., 2005; Ferror et al., 2005, Kongoli et al., 2007). It is a statistical-based algorithm that finally retrieve the rain rate through a regression expressing the relationship with surface rainfall, such as PCT and SI.

However, this statistical-based algorithm consists of a linear or quadratic regression for the relationship between the exponents representing the state of each rain cloud and the surface. This cannot be said to properly represent the relationship between a rain cloud and a surface with complex and nonlinear relationships.

2.1.2 Physical retrieval algorithm

The most widely used algorithm for physically based retrieval algorithm is Goddard Profiling Algorithm (GPROF). The GPROF algorithm was introduced in Kummerow et al (1996) as a method for estimating precipitation profiles based on Bayesian formulations. Later, researchers further developed this by applying the cloud resolving model (Olson et al., 1996; Zarzano et al., 1999; Bauer et al., 2001, Viltard et al., 2006). This is described in detail in Kummerow et al (1996, 2001, 2011, 2015) and the development and description of the algorithm. In this paper, I will only briefly explain. Based on Bayesian theory, this method weights the precipitation profile using the posterior

probability. In order to obtain the posterior probability, the prior probability of the precipitation profile, the conditional probability between BT and precipitation obtained using microwaves, and the prior probability density function(PDF) are required. These are obtained by Cloud-resolving model (CRM) and RTM.

GPROF is the official algorithm for rain rate product of GPM constellation provided by NASA's PPS. This algorithm requires accurate estimation of BT for each frequency by using CRM and RTM. However, this is limited to consider in each model the impacts of the various land surfaces. Therefore, the regression of the scattering type is still applied to the rainfall on land.

2.1.3 The retrieval of rainfall intensity using neural networks

Recently, neural networks have been used as another statistically based approach for retrieving the rain rate from IR and PMW senors(Hsu et al., 1997, 2018; Yao et al., 2001; Sarma et al., 2008; Mahesh et al., 2011; Panegrossi et al., 2013; Sanò et al., 2015, Tan et al., 2018;). Neural networks have the advantage of expressing nonlinear relationships. Hus et al. (1997) developed the precipitation estimation from PERSIANN. It uses intrared satellite measurement and ground-surface information for the estimation of rain rate. Yao et al. (2001) successfully demonstrated the usefulness of SI and PCT for rainfall retrieval on the Tibetan

Plateau. Tapiador et al. (2004) also developed and evaluated a new operational procedure to produce half-hourly rainfall estimates at 0.1˚ resolution using ANN based approach utilizing PMW and IR satellite measurement. Smara et al. (2008) used an ANN based rainfall retrieval algorithm for estimating rain over Indian land and oceanic regions. However, most of the previous studies on the retrieval of rain rate using a neural network targeted IR and VIR of GEO satellites, and in the case of PMW, most of the experimental studies on a specific sensors were conducted. Also, a simple ANN has limitations in implementing complex nonlinearity.

2.2 Rain product blending

Consistent monitoring with single satellite is difficult due to spatial and temporal resolution. Blending data that synthesize multi-satellite data are being developed and researched for consistent observation of various meteorological and climatic factors. The measurement of rainfall using satellites also are developed and provided these blending data. However, while most meteorological and climate variables follow a normal distribution, it is difficult to apply general interpolation or synthetic techniques due to the nature of rainfall or distribution of intensity.

Recently, rain rate synthetic fields have been provided by various institutions. NOAA NESDIS provides a blended-rain rate (bRR), a composite field for rainfall intensity of nine satellites equipped with PMW sensor. Briefly, The bRR performs bias correction, based on the histogram of the reference rain rate over the past 5 days, on rain rate of each satellite and maps it to a 0.25° grid. When mapping, if there are more than one data in the grid, they are averaged.

The PERSSIANN-Climate Data Record (PERSIANN-CDR) from NOAA also provides quasi-global daily rainfall data. It provides a composite field through error correction by applying ANN to IR estimates of GEO satellite.

GPM provides a resource called integrated multi-satellite retrievals for GPM (IMERG). This data include PMW and IR synthesis data respectively. The synthesis field of PMW, the target of this study, was subjected to temporal interpolation using the CPC Morphing-Kalman filter (CMORPH-KF) (Joyce et al., 2011) for the rain rate product of GPM constellation, and then bias correction using CDF is performed based on data of GMI and DPR. After that, the final data is produced by distributing the nearest data to the grid.

Most of the techniques of the existing blending methods for rainfall derived from satellite are produced with mean, median and nearest interpolation, after time interpolation and bias

correction of data. This is due to the assumption that the uncertainty of each satellite or sensor is the same even through the bias-corrected data. So it is difficult to say that rational distribution of weight among the blending members was achieved.

On the other hand, EBMA is a synthetic technique that considers differences in uncertainty between members, and has the scalability that can be applied not only to normal distribution of data such as temperature, but also to distribution such as precipitation (Raftery et al., 2005; Sloughter et al., 2007). Previous studies have been mainly used for climate model ensembles, and recently applied to satellite data synthesis (Kim et al., 2016).
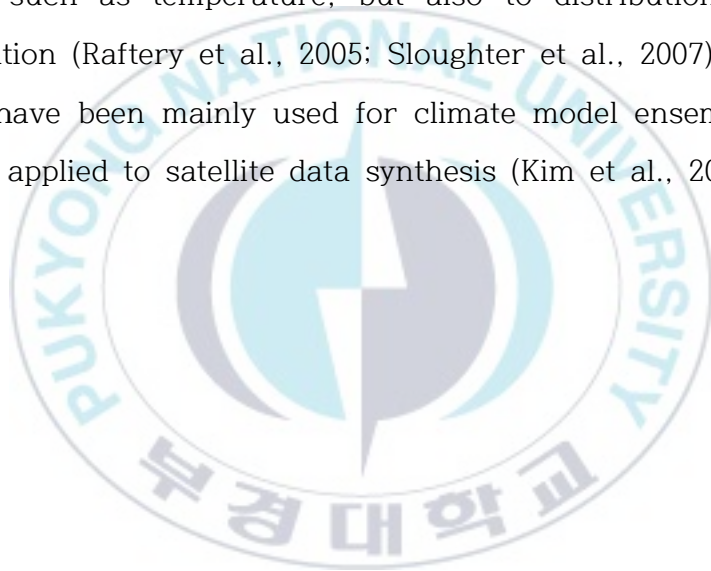
CHAPTER 3 DATA AND STUDY AREA

3.1 Data

In this study, the data obtained from Global precipitation measurement (GPM) microwave imager (GMI), Special Sensor Microwave Imager/Sounder (SSMIS), and the microwave humidity sounder (MHS) sensors were used. SSMIS and MHS are mounted on multiple satellites. This can make it possible to increase the time resolution. data obtained from multiple satellites were synthesized according to each sensor with half-hourly unit of time and used as a single sensor data.

3.1.1 Sensors

Global precipitation measurement (GPM) microwave imager (GMI) instrument which has multi-channel, is conical-scanning. This sensor provides quasi-global data. GMI has 13 frequency bands from 10GHz to 183GHz. GMI provides various information about atmospheric clouds and water droplets by possessing various frequencies.

Special Sensor Microwave Imager/Sounder (SSMIS) is mounted on the satellites of Defense meteorological satellite program (DMSP)'s F16-19 series. These satellites are LEO satellite that observes the globe. In order to match up with reference data

and to secure the number, data obtained through multi-satellite observing the research area at timely were used. SSMIS has 24 channels from 19GHz to 183GHz. It is a sensor that has been used in many previous studies due to the advantage of being able to use information on various atmospheres and water droplets through various channels.

The last sensor used is the microwave humidity sounder (MHS). MHS has channels in the frequency domain similar to advanced microwave sounding unit B (AMSU-B) instruments. MHS is also a PMW sensor and is mounted on various NOAA satellite series and meteorological operational satellite (Metop) series. In this study, data obtained from MHS of NOAA-18, 19, Metop-A, and B satellites were used. These are satellites that are still in operation.

The GMI and SSMIS are imager and conical scanning sensors. The imager aims for a two-dimension image. The observation of rainfall using an imager uses scattering in the atmosphere and emission from the surface. To estimate the rain rate, SSMIS had been calculated through regression analysis using scattering. And it can be retrieved using the GPROF algorithm from data obtained from GMI. The method of estimating the rain rate using scattering and emission is easy to estimate the amount of rain. The MHS is a sounder with 5 high-frequency channels and a cross-scanning sensor. The sounder aims to observe the

atmospheric vertical structure. The MHS retrieve the rainfall intensity based on estimating the amount of moisture in the atmosphere. This has the advantage of detecting the existence of rainfall.

In this study, BTs, SI, IWP, and SC ratio for each sensor were acquired and generated as input data used to improve rainfall intensity using DNN. ECMWF Land-sea fraction data was used as ancillary data.

3.1.2 Brightness temperature

Brightness temperature is a key factor used in both physical and statistical estimation. Through previous studies, BT data of frequencies known to be closely related to rainfall intensity calculation was acquired. The data was obtained by using level 1C data provided by NASA's PPS.

Level 1C is the BT data converted from the radiance obtained from each sensor after the common calibration process. The Level 1C algorithms contain the following major components:

● Orbitzation.

● Satellite intercalibration.

● Quality control.

● Ancillary data calculations.

The detail of Level 1C algorithms is described in the official document (ATBD NASA GPM Level 1C Algorithms version 1.8,

2017). Since the frequencies held for each sensor are different, the frequency of BT used for each sensor was selected differently.

3.1.3 Scattering index(SI)

The SI obtained by the Tb of the 19V, 22V, and 85V channels was introduced by Grody (1991). Research that are estimating rain rate through empirical relationship between this index and surface rainfall have been ongoing. Recently, many sensors use 89V or 91V channel instead of 85V channel. And accordingly, SI are also calculated by 89V or 91V channels instead of 85V. The formula for calculating SI is as follows.

$$F = a \times BT_{19v} + b \times BT_{22v} + c \times (BT_{22v})^2 + d \qquad (4.1)$$

$$SI = F - BT_{89v} \qquad (4.2)$$

Where F is the BT of the 89V channel estimated using the 19V and 22V channels. BT represent the Tb of each channel. a, b, c, and d represent empirically obtained coefficients and constants. SI means the difference from estimated BT of 85V channel using BT of the 19V and 22V channel and the BT obtained by 85V. GPM does not provide a SI index. The SI of the GMI was calculated using the BT data acquired for the input of this study. GMI uses 89V channels instead of 85V channels. The coefficients

in Equation 4.1 are needed to calculate SI. In this study, we used the coefficients used to calculate the SI using the SSMIS sensor of the DMSP satellite (Grumman, 2002).

3.1.4 Ice water path (IWP)

The ice water path (IWP) defined as the integral of the ice water content (IWC) through the depth of an ice cloud layer. It is that uses BT frequency around 150GHz that is the more sensitive to weak rain than BT of 89GHz range. Like SI, it has been widely used for statistical rain rate estimation, and it is the main factor of the algorithm that produces the NOAA's rainfall product

3.1.5 Stratiform-convective rain ratio (SC ratio)

Stratiform-convective rain ratio is the ratio of convective rainfall to total rainfall. We calculated the SC ratio using convective rain rate which is the output of the GPROF algorithm. Depending on whether it is convective rainfall or stratiform rainfall, the intensity of rainfall varies greatly even in areas of similar clouds. While the convective rainfall generally shows the strong rain rate, the stratiform rainfall shows weaker rain rate than convective rainfall.

3.1.6 Land-sea fraction

In addition, ECMWF's Land-sea fraction data, which are ratio

of land in the grid, were used as ancillary data. It was used to take into account the difference between oceans with homogeneous surface characteristics and lands with complex surface characteristics.

3.1.7 Reference data

The ground radar or rainfall gauge is mainly used as a reference data for validating the satellite-based rainfall. It is considered that the observation of the point using the rainfall gauge is the most accurate(Qin et al., 2014). However, this observations is only possible on land and have limitations due to the density of data and spatial and temporal gaps(Huffman et al., 1995; Huffman and Klepp, 2011). Satellite radar is somewhat less accurate for rainfall than point observation of the rain gauge, but it can observe both land and ocean. It is showing similar accuracy to ground radar, and is steadily improving by many researchers. Since this study focuses on the methodological on improving and blending of rain rate, satellite radar data is used as reference data to conduct research on wide spaces.

In this study, the rain rate retrieved by the Dual-frequency precipitation radar (DPR) sensor was used as the reference data for the improvement and blending of rainfall. The DPR operates the Ku-band (13.6GHz) and Ka-band (35.5GHz) to monitor rainfall. DPR has a resolution of 5 km (Figure 3.1). The swath of the

Ku-band is 245 km, and Ka-band is 120 km (Figure 3.2). The scan area that observes the rainfall using the Ku band is called normal scan (NS), and it shows twice the width of the area observed with the Ka-band. In this study, the rainfall data retrieved from Ku-band was used to improve the rainfall and retain a sufficient number of match-ups of reference data for training the composite model.

Kidd et al. (2018) investigated the level 2 precipitation data produced by GPM using ground radar and rainfall gauge in Europe. When comparing the occurrence frequency according to rainfall intensity in PDF, it was confirmed that the distribution of instantaneous rainfall and accumulative rainfall in Europe (Figure 3.3) and the United States (Figure 3.5) was most similar to DPR. And The product of DPR rain rate was confirmed to be more dense on the 1:1 line compared to the products based from microwave radiometers (PMW sensors) in Europe(Figure 3.4) and the United States(Figure 3.5).

Baiswas and Chandraekar (2018) compared the reflectance of DPR and ground radar using volume matching at five ground radar points. The reflectance of the DPR showed high correlation at all five points(Table 3.1) compared to the ground radar(Table 3.2). However, when comparing the instantaneous rainfall, it was confirmed that the correlation decreased. This is thought to be due to the difference in the rain rate retrieval algorithm between

the ground radar and the satellite radar.

Tan et al (2018) evaluated the accuracy of the DPR sensor. Compared to the gauge data on the surface, it varies depending on the region, but has a correlation coefficient of 0.57 to 0.80, and NRMSE is 0.71 to 1.68, which is more accurate than the PMW sensors obtained through the GPROF sensor.
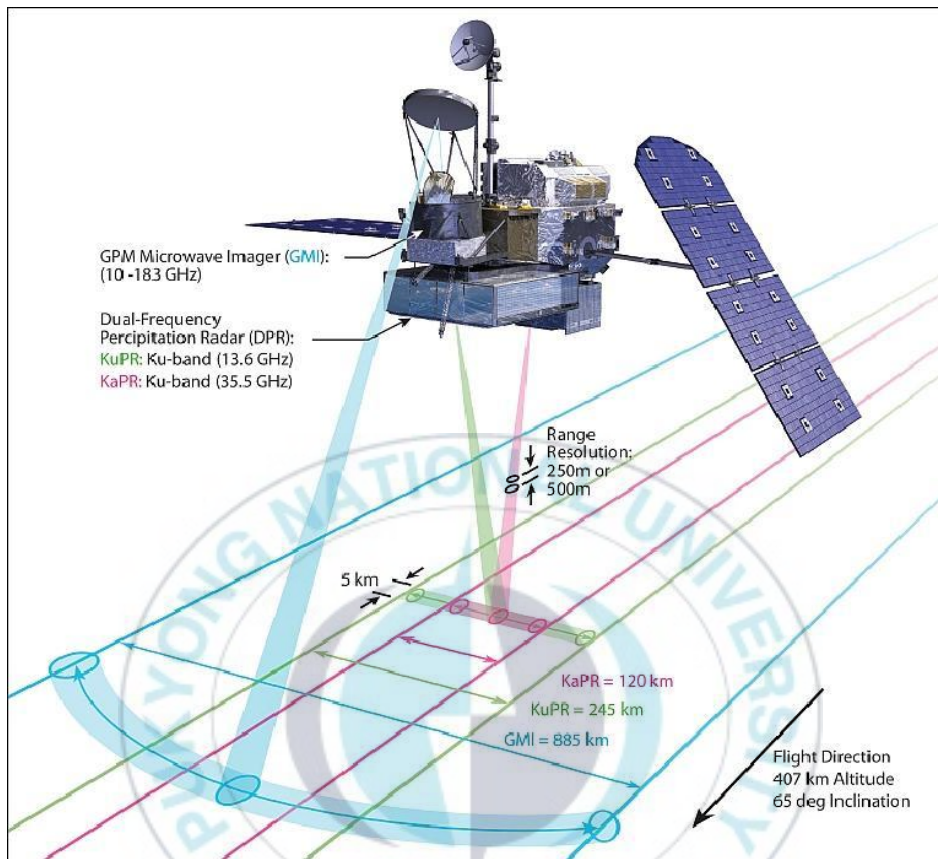
Figure 3.1. The GMI and DPR instruments aboard the GPM Core Observatory and their respective swaths – the area of Earth's surface observed by the instrument. (https://earth.esa.int/web/eoportal/satellite-missions/g/gpm)
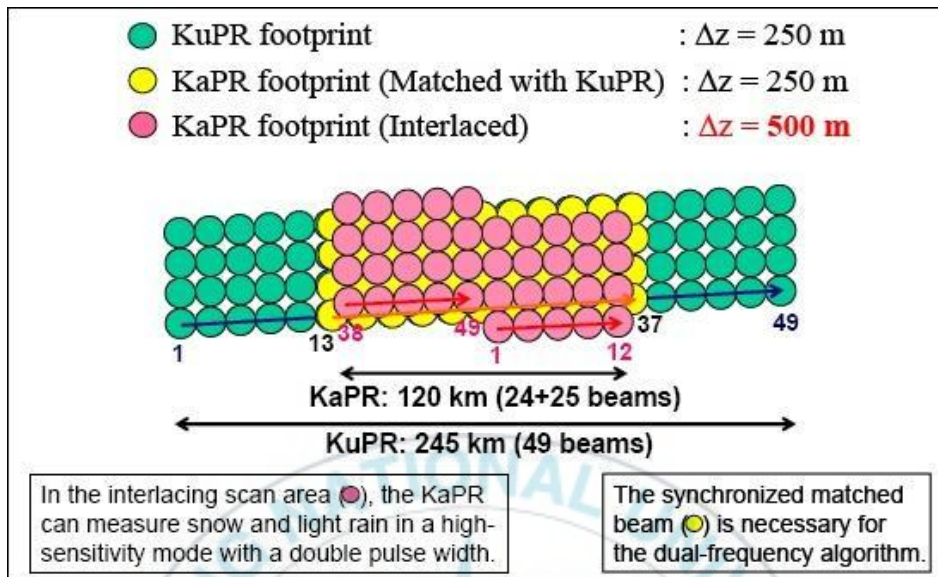
Figure 3.2. The DPR antenna scanning concept.
(https://earth.esa.int/web/eoportal/satellite-missions/g/gpm)
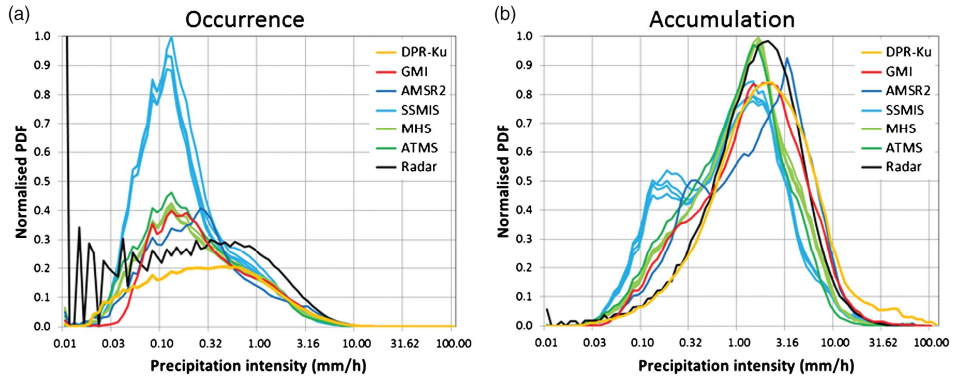
Figure 3.3. Normalized distribution fo the (a) occurrence and (b) accumulation of precipitation by intensity for GPROF and DPR Ku products over western Europe (Kidd et al.,2018)
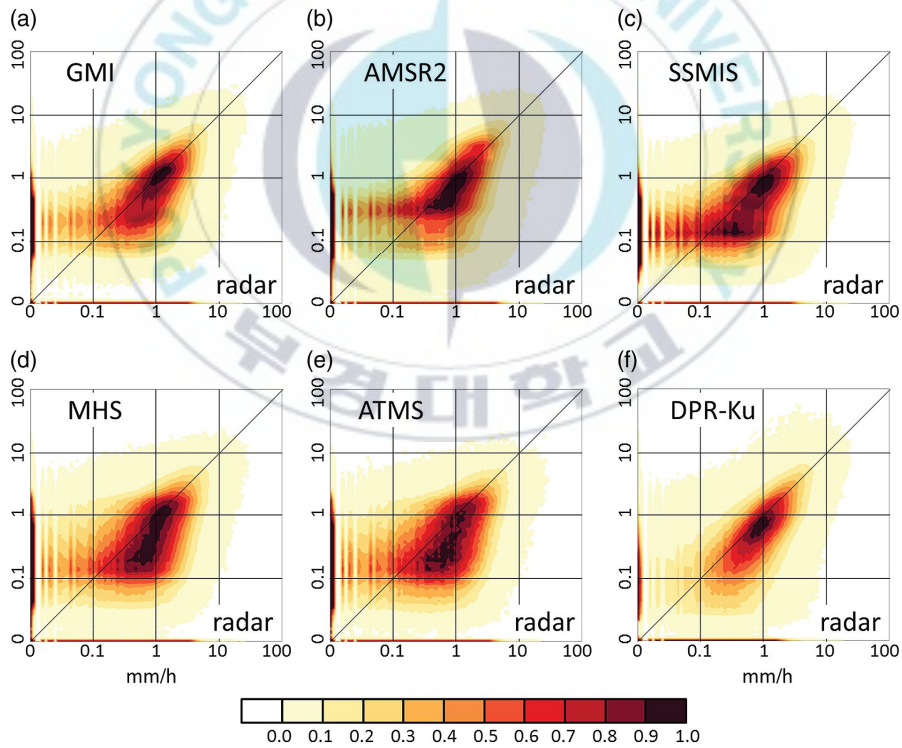
Figure 3.4. Normalized density scatterplot of GPROF(a-e) and DPR-Ku(f) precipitation products versus ground radar(western European) (Kidd et al.,2018)
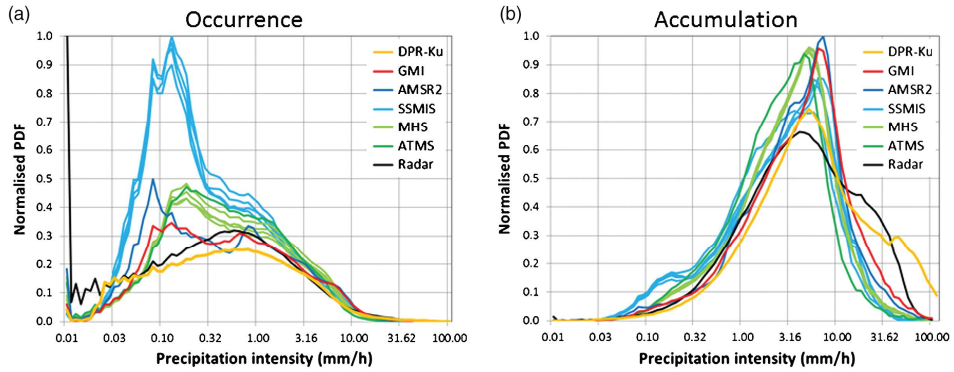
Figure 3.5. Normalized distribution fo the (a) occurrence and (b) accumulation of precipitation by intensity for GPROF and DPR Ku products over eastern United State(Kidd et al.,2018)
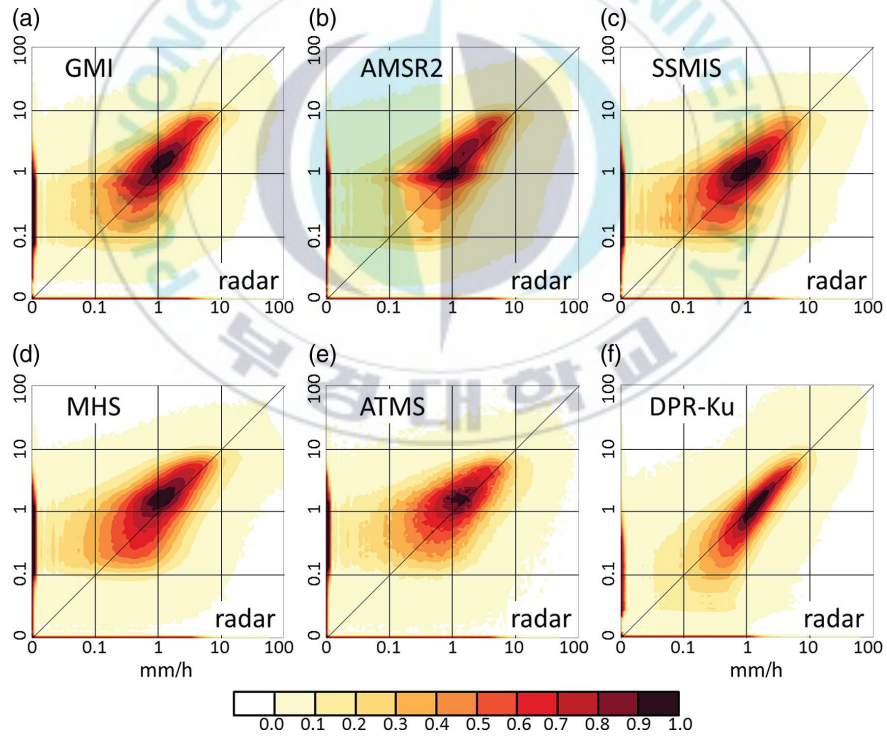
Figure 3.6. Normalized density scatterplot of GPROF(a-e) and DPR-Ku(f) precipitation products versus ground radar(eastern United State)(Kidd et al.,2018)

Table 3.1. Next-generation radar(NEXRAD) stations' name, location and latitude.(Biswas and Chandrasekar, 2018)

Name	Location	Lat(deg.)	Lon(deg.)	Altitude(m)
KFWS	Dallas/Ft. Worth, TX	32.5731	-97.3031	208.17
KHGX	Houston/Galveston, TX	29.4719	-95.0792	5.48
KSHV	Shreveport, LA	32.4508	-93.8414	83.21
KLIX	New Orleans, LA	30.3367	-89.8256	7.31
KMLB	Melbourne, FL	28.1133	-28.6542	10.66

Table 3.2. Summarizing different error statistics for comparison of DPR-Ku and NEXRAD S-band reflectance and rain rate product

Radar	GPM Ku-band reflectance			GPM Ku-band Rain rate product		
	MB (dBZ)	MAE (dBZ)	CC	MB (mm/hr)	MAE (mm/hr)	CC
KFWS	0.49	2.38	0.90	-0.50	1.66	0.68
KHGX	0.18	2.44	0.91	-1.07	2.44	0.66
KSHV	0.64	2.40	0.90	0.30	1.75	0.66
KLIX	0.58	2.67	0.88	-0.61	2.56	0.60
KMLB	0.31	2.58	0.87	-0.29	1.90	0.54

3.1.8 Input data for DNN

The PMW frequencies possessed by each sensor are different (Table 3.3). Through previous studies, frequencies that were thought to be closely related to rainfall were selected. For GMI, BT of 18.7 (V/H), 23.8 (V), 36.6 (V/H), 89.0 (V/H), 166.0 (V/H), 183 ± 3 (V/H), 183 ± 7 (V/H) GHz, SI, IWP, SC ratio and land-sea fraction were used as input data to the DNN models for rain discrimination and rain rate. For SSMIS, BTs of 19.35 (V / H), 22.235 (V), 37.0 (V / H), 91.66 (V / H), 183.31±1 (H), 183.31±3 (H), 183.31±6 (H) GHz, SI, IWP, and SC ratio were used as input data. MHS does not have the low-frequency range. SI cannot be calculated because it does not have the low frequency range of 19 and 22GHz. Therefore, unlike GMI and SSMIS, BT of IWP, SC ratio and 5 frequencies was used except SI. For the five frequencies, BTs obtained at 89 (V), 157 (V), 183±1 (H), 183±3 (H), and 190 (H) GHz were used.

Table 3.3. The frequency and resolution of GMI, SSMIS, MHS channels

Frequency(GHz)								
GMI	10.65 V/H	18.7 V/H	23.80 V	36.50 V/H		89.00 V/H	165.6V/H	183.31 V
SSMIS		19.35 V/H	22.235 V	37.0 V/H	50.3-63.28 V/H	91.65 V/H	150 H	183.31 H
MHS						89 V	157 V	183.311 H 190311 V
Resolution(km)								
GMI	26	15	12	11		6	6	6
SSMIS		59	59	36	22	14	14	14
MHS						17	17	17

3.2 Study area

Latitude from 20 to 50 degrees and a longitude from 100 to 150 degrees was selected as the study area. Data for each satellite passing through this area was acquired. This study area is considered to be a lack of research on improvement and verification of rainfall in previous studies compared to Europe and America. The region is also part of the monsoon domain and has a significant impact on East Asia's summer climate (Nitta 1987; Wu and Wang, 2000). The precipitation characters in Japan and Korea were influenced by transport of water vapor in this area(Tomita et al., 2004; Lee et al., 2005). In early summer East Asia, rain continues to exist for a period of time. The rainy season is called changma in South Korea, baiu in Japan, and mei-yu in China. The changma front forms around the boundary between mid-latitude and subtropical air masses and is characterized by a strong gradient of equivalent potential temperature (Ninomiya 1984; Ninomiya and Akiyama 1992). The rainy season shows strong rainfall as well as a wide variety of precipitation characteristics(Takayabu and Hikosaka 2009; Xu et al. 2009; Xu and Zipser 2011; Yokoyama et al. 2014; Park et al. 2016).

Depending on its characteristic and amount, precipitation associated with the rainy season has a large impact on the East Asian region. Therefore, the accurate rainfall observations in the

East Asian region of this period are required (Yokoyama et al., 2017). Also, Prakash et al. (2016) suggested the improvement of IMERG precipitation for intensive rainfall in India during the southwest monsoon period. Lee and Lee (2018) carried out verification of GPM IMERG data and ground observation data for daily precipitation, and showed regret for the overall area of East Asia compared to accuracy for South Korea.

For the study period, data from January 2016 to December 2018, which are common periods of eight satellites, were acquired. During the study period, Korea recorded less precipitation than the normal year in 2017, and 2018 was similar to the normal year, but in October there was the most rainfall event for 30 years (Korea meteorological administration (KMA), 2017, 2018).
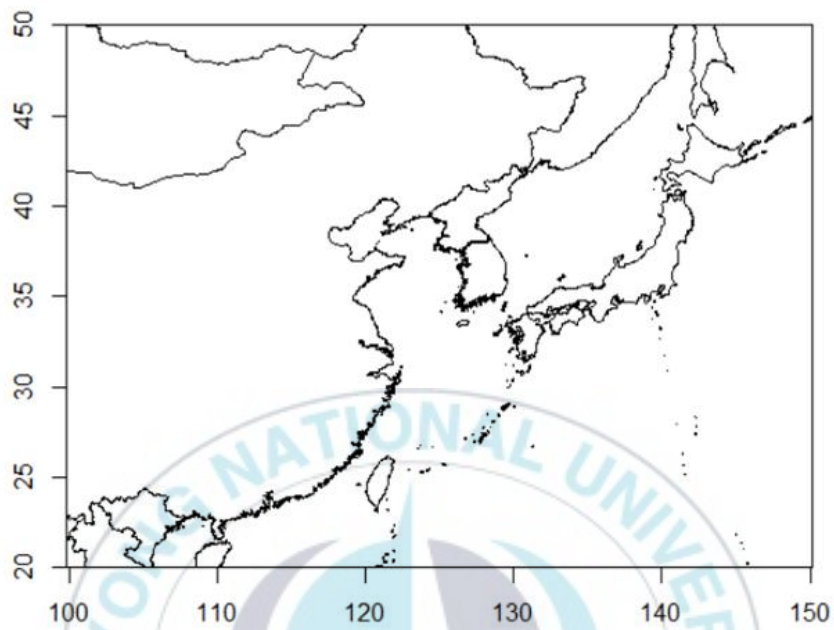
Figure 3.7. Study area

CHAPTER 4 METHODS

4.1 Overview of research procedure

Figure 4.1 shows the overall flow of this study. The GMI data of GPM CO, the SSMIS of DMSP F16, 17 and 18, and the MHS data of NOAA-18, 19 and Metop-A, B were acquired. Each satellite data was synthesized as half-timely data respectively. The same coordinate system was set up to, and match-up data between each sensor data and reference data was constructed. The constructed match-up data was divided into training data and verification data. These data were used for determining the rain discrimination DNN classification model and rain rate retrieval DNN regression model. The improved rain rate of each sensor obtained through the combination of the two DNN models was extracted again with overlapping match-up data from all three sensors with reference data DPR for generating EBMA model. Empirical cumulative distribution function (ECDF) was constructed using the DPR rain rate. Based on the constructed ECDF, bias-correction was performed once again for the rain rates of GMI, SSMIS, and MHS. EBMA was repeatedly performed to select the appropriate percentile and constructed the final EBMA model. Finally, the weights were redistributed according to the number of members in each grid in the study area to create

an EBMA blending map. In this chapter, we will explain the improvement of rain rate and the blending process in detail.
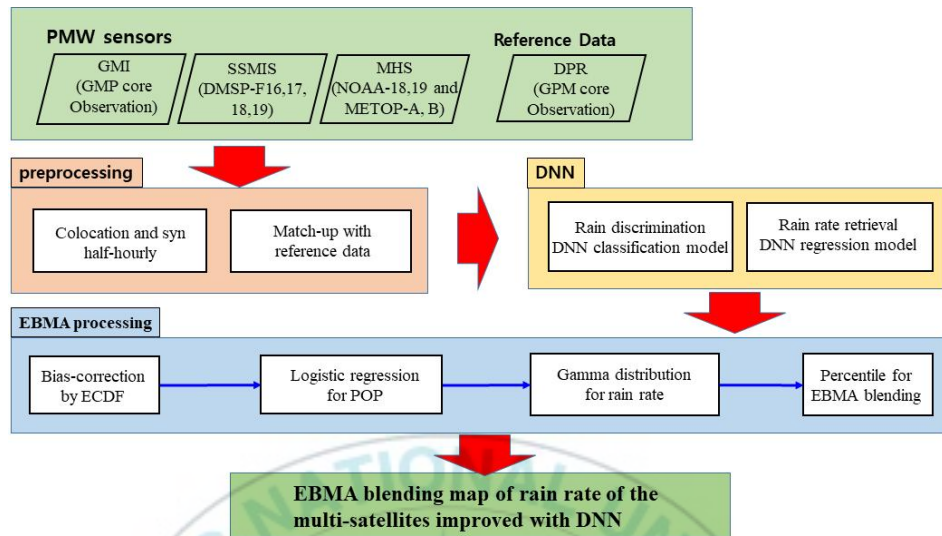
Figure 4.1. Research flow

4.2 Data preprocessing

4.2.1 Preprocessing

The focus of this study is on general rainfall issues. January, February and December are winter period in the study area. In winter, precipitation data includes snowfall data along with rainfall data. To exclude this, only data from March to November were used.

Also, during the period from March to November, data less than 20mm/hr was extracted to target general rainfall. In matchup between input sensors and reference sensor respectively, rain rate over 20mm/hr in the study area from 2016 to 2018 are considered to be outliers at about 0.01%. The criterion for determining rainfall event was 0.1 mm/h, and numerically, data below 0.1 mm/h were judged as non-rain. It referred to the standards of the Korea meteorological administration (KMA).

As a preprocessing task, sensors including multi-satellites were first generated as half-hourly data. The first half represents data observed from 00 to 29 minutes, and the second half represents data from 30 to 59 minutes. Since each sensor has different resolutions for each frequency, gridding and gap filling were performed for matching. Bilinear interpolation was used for the gridding. However, since most frequencies have a resolution

lower than 0.1°, an empty space is generated. Gap filling was performed using the surrounding pixel values. The value of the surrounding 24 pixels (5×5) was averaged and filled with the pixel having no value. The coordinate system was set to the World Geodetic System 1984(WGS84) coordinate system. We obtained ECMWF's land-sea fraction data of 0.5° and downscaled it to 0.1° with bilinear interpolation.

4.2.2 Matchup and sampling

After performing the preprocessing of each data, a match-up was performed between each sensor's hourly data and DPR data. The GMI, which shares the same instrument with DPR, had the largest number(approximately 31 millions) of match-ups for the study period. MHS, which has the largest observation width, was the next (approximately 4 millions). The SSMIS had the smallest number (approximately 2 millions) of match-ups. Among match-ups of SSMIS, about 180,000 event were judged to be DPR-based rainfall. To generate the DNN model for each sensor, the same number of data was used as input data through random sampling in the match-up data of GMI and MHS based on the number of data of SSMIS that showed the smallest number of rainfall events.

In the case of the DNN classification model for rain discrimination, the ratio of the number of rainfall and non-rain

was the same. Therefore, a model was generated using a total of 200,000 data by adding 100,000 non-rain and 100,000 rainfall events. The test data for model verification were also composed of rain and non-rain in the same number. The number of rainfall events in the test data was 20,000, and the number of non-rains was 20,000. For the training data of the DNN regression model for rainfall intensity calculation, the rainfall data of the rain discrimination DNN classification model was used. That is, 100,000 training data and 20,000 test data were used. In the study area, relatively strong precipitation and many rainfall events occur during the summer months. To reflect this, random sampling was performed considering the ratio of the number of rainfall events in the entire matchup to the number of rainfall events in the month.

4.3 Deep neural network

4.3.1 Improvement of DNN

The ANN structure is based on the human brain's biological neural system, and can be used to develop empirically based rainfall models. The ANN architecture has been used in the study of various remote sensing data and has achieved better results than statistical techniques (Foody, 1996; Jensen et al., 1999; Ji, 2000; Aitkenhead and Aalders, 2011; Santi et al., 2014).

ANN consists of three layers: input layer, hidden layer and output layer. The input layer means the explanatory variables in this study, and one of the main advantages of the ANN is that it can use various kinds of data. By using neurons in hidden layers, the nonlinear patterns of input data can be simulated. Setting the number of neurons in the hidden layer is one of the most fundamental issues in recent research (Brown et al., 2008; Hu and Weng, 2009). The neurons in the output layer can represent each of classes classified by ANN, and ANN can predict continuous variables using the single neuron (Jiang et al., 2004; Gonzalez Vilas et al., 2011). The factors affecting performance of ANN include the hidden layer, learning rate, training tolerance, and others. The number of hidden nodes determines the number of connections between inputs and outputs and may vary depending on the specific problem under

study. The learning rate determines the amount the weights change during a series of iterations so as to bring the predicted value within an acceptable range of the observed value; furthermore, the training tolerance refers to the maximum error rate at which the network must converge during training (Ji et al., 2007).

DNN, also known as deep learning, is a machine learning method similar to ANN, but capable of processing the complicated, huge input data by learning tasks by using the feed-forward multi-layer network (Ali et al., 2015; You et al., 2017). In the case of ANN, if the layers of the neural networks are deepened, then its performance is reduced. It can be divided into three problems. First, the gradient vanishing occurs that the output error of neural network is not reflected as further away from the output layer. The back-propagation algorithm trains the neural network by reversing the error of the output layer to the hidden layer, but the error is hardly transmitted to the hidden layer in the front layer and then learning do not properly. This problem can be solved by using the Rectified Linear Unit (ReLU) function as an activation function of nodes. When using the Sigmoid function, the output of neural network nodes does not exceed 1 even if the input value increases. However, ReLU improves the learning performance by removing these restriction (Agostinelli et al., 2015; LeCun et al., 2015). Secondly, the

overfitting problem of a neural network is encountered. In order to improve the performance of neural network, it is necessary to deepen the structure by increasing the number of hidden layers. In this case, the overfitting problems arises, and which can be solved by applying the "dropout" (Pham et al., 2014; Srivastava et al., 2014). "Dropout" is an efficient and uncomplicated method that randomly learns for some nodes without learning the entire neural network. Applying dropout effectively prevents the neural network from becoming overfitting because the nodes and weights are changed each learning (Pham et al., 2014). However, the dropout ratio is empirically set because there is no definite criterion for setting dropout ratio. The last problem is that ANN requires extensive computation. As the number of hidden layers increases, the connection weights increase exponentially and, given that much learning data is needed, the computational complexity increases rapidly. Due to these problems, it was difficult to learn the deep neural network in the past. In recent years, however, due to the development of high-performance hardware such as Graphic processing unit (GPU) and improvement for several algorithms, the application of DNN has become available.

Training process of DNN usually consists of pre-training and fine-tuning. Pre-training is the phase of data processing that includes unsupervised learning so as to improve the

generalization error of the trained deep architectures. Fine-tuning by supervised learning is performed to improve the classification error (Erhan et al., 2010).
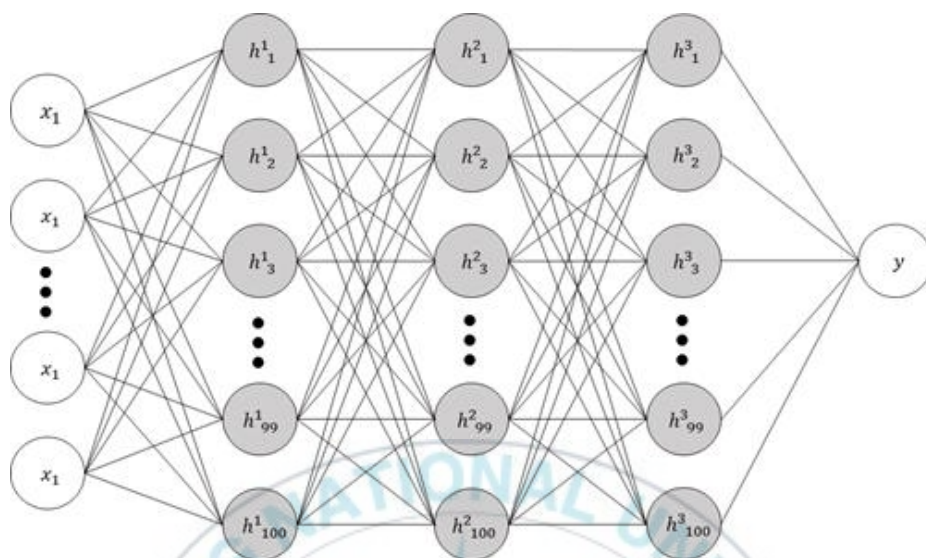
Figure4.2. The diagram of DNN

4.3.2 Integration of DNN for rain rate

Two models were generated to apply DNN to retrieval of satellite rainfall. The first is the DNN model to determine whether there is rainfall event. This model determines whether it is rain or non-rain by using the data obtained through the satellite sensor as input data. Next, The DNN model was constructed to retrieve the rain rate. The rain rate retrieval DNN model generated by data obtained through satellite sensors as input data which were judged to be rain.

To optimize these models, iterative experiments were performed on various hidden layers and nodes, and the appropriate structures were selected. There were 9 hidden units, from 2 to 4 hidden layers, and the number of nodes in each hidden layer increased from 100 to 300 by 100. In this study, the k-fold cross validation(CV) technique was used as a means for exploring the optimal hyperparameters and generalizing the model. The k-fold CV technique divides some of the training data into k without overlapping to generate k CV models, and each CV model constructs a model using it as validation data(Figure 4.4). Finally, the predicted values of k models performed individually are equal to the total number of training data. Eventually, a final model is generated through the verification matrix using the predict. And, to avoid overfitting and find the optimal epochs of the model, we use the early stopping. Five

cross-validation models, that is 5-fold, were constructed to create the model.
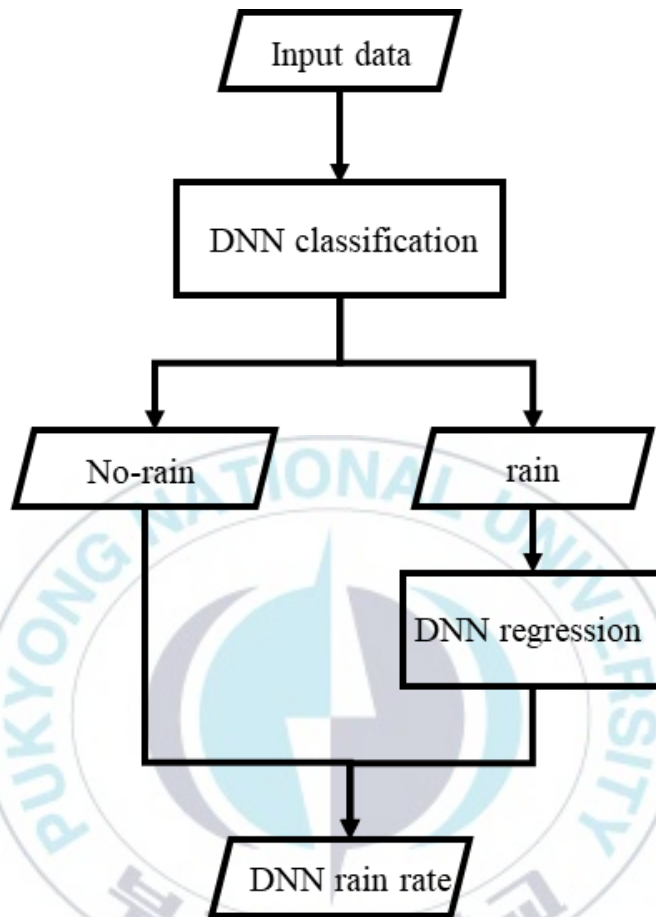
Figure4.3. Schematic diagram of configuration for DNN model for improvement of rainfall retrieval.
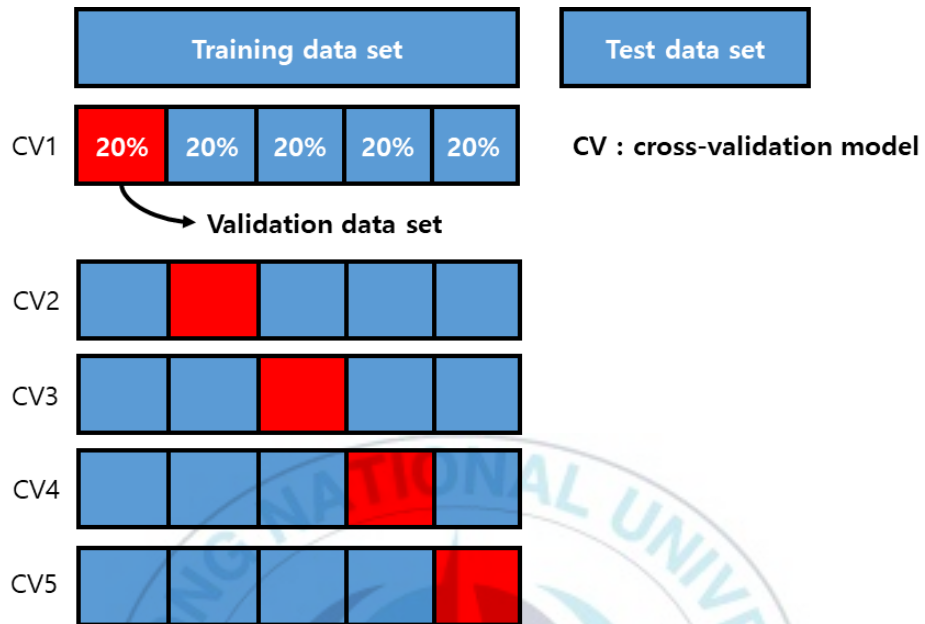
Figure 4.4. Concept of K-fold cross validation technique

4.3 Ensemble Bayesian model averaging

In this study, EBMA technique was used to blend the rain rate. In order to apply the theoretical BMA technique, EBMA with mixed model technique was used. The EM algorithm was used as the mixed model technique. In the sections below, explanations of each theory and procedures for application to blending rainfall are given.

4.3.1 Theoretical BMA

In BMA for ensemble forecasting, each ensemble member predict f_k is associated with a conditional PDF $h_k(y|f_k)$, which can be thought of as the PDF of the rain rate quantity y given f_k, conditional on f_k being the best predict in the ensemble. The BMA predictive PDF is then expressed as Equation 4.1.

$$p(y|f_1, ..., f_K) = \sum_{k=1}^{K} w_k h_k(y|f_k) \tag{4.1}$$

Where w_k is the posterior probability of predict k being the best one, and is based on predict k's relative performance in the training period. The w_k's are probabilities and so they are non-negative and add up to 1. Here K is the number of ensemble members.

However, it is difficult to calculate the posterior probability and generate PDF of each member using only rainfall data. Therefore,

theoretically, BMA has limitations in its application to synthesis. To overcome these computational difficulties, the EM algorithm was applied to estimate the PDF and posterior probability used in BMA.

4.3.2 EM algorithm

EM algorithm is a kind of mixture model and is an algorithm that estimates PDF through maximum likelihood estimation(MLE). It repeats e step and m step and updates the mean and variance of probability density function.

Expectation step :

$$\widehat{Z}_{kst}^{(j)} = \frac{w_k^{(j-1)} g(y_{st}|\widetilde{f}_{kst}, \sigma_k^{2(j-1)})}{\sum\limits_{i=1}^{K} w_i^{(j-1)} g(y_{st}|\widetilde{f}_{ist}, \sigma_k^{2(j-1)})} \tag{4.2}$$

Maximization step:

$$w_k^{(j)} = \frac{1}{n} \sum_{s,t} \widehat{Z}_{kst}^{(j)} \tag{4.3}$$

$$\sigma_k^{2(j)} = \frac{\sum\limits_{st} \widehat{Z}_{kst}^{(j)} (y_{st} - \widetilde{f}_{kst})^2}{\sum\limits_{s,t} \widehat{Z}_{kst}^{(j)}} \tag{4.4}$$

4.3.3 Improvement of EBMA for blending the rainfall

In order to apply EBMA to blending rainfall derived from satellite, based on previous studies, appropriate improvement of

satellite rainfall data is required. First, in the previous study, bias-correction through the linear regression equation required in the EBMA process was replaced with a method using ECDF. After that, the PDFs of the bias-corrected members need to be adjusted to the rainfall. The PDF used in this study was divided into probability for rain and probability for non-rain, and was composed of the sum.

$$h_k(y|f_k) = P(y=0|f_k)I[y=0] + P(y>0|f_k)g_k(y|f_k)I[y>0] \qquad (4.5)$$

Where, as I is rain classification factor, ($I[y>0]$) means the rain state and ($I[y=0]$) denote the non-rain state. The first component as the non-rain probability is consisted by logistic regression. The probability for non-rain is calculated. Logistic regression is expressed as follows(3.6).

$$P(y=0|f_k) \equiv \log\frac{P(y=0|f_k)}{P(y>0|f_k)} \qquad (4.6)$$
$$= a_{0k} + a_{1k}f_k + a_{2k}\delta_k$$

In previous studies, in Equation 3.6, there was exponent index for correcting the value of the ensemble member. This exponent was selected based on the correlation coefficient of ensemble members and reference materials. This was a factitious means to increase the accuracy of the ensemble between model data and actual data. In this study, it was removed and used. The second factor is the probability of precipitation, which is described by the gamma distribution. The total probability of this gamma distribution is the probability minus the probability of the specific precipitation

calculated from the first factor. In general, the gamma distribution is expressed as follows using shape parameters (α) and scale parameters (β).

$$g_k(y|f_k) = \frac{1}{\beta_k^{\alpha_k}\Gamma(\alpha_k)}y^{\alpha_k-1}\exp(-y/\beta_k) \qquad (4.7)$$

The average and variance of the gamma distribution is determined using shape parameters and scale parameters.

$$\mu_k = b_0 + b_1 f_k = \alpha_k\beta_k \qquad (4.8)$$
$$\sigma_k^2 = c_0 + c_1 f_k = \alpha_k\beta_k^2$$

Finally, PDF of EBMA is expressed as Equation 4.9 as a combination of logistic regression and gamma distribution as follows.

$$p(y|f_1,\ldots,f_k) = \sum_{k=1}^{K} w_k[P(y=0|f_k)I[y=0] + P(y>0|f_k)g_k(y|f_k)I[y>0]] \quad (4.9)$$

Here, the weight, w_k of each member and the parameters of the logistic regression coefficients and gamma distribution, which are components of the PDF of each member, are selected from the EM algorithm using the training data.

Finally, since the result of EBMA is PDF, an appropriate percentile is required to calculate the rain rate. Therefore, the EBMA result was repeatedly performed at a constant percentile gap to calculate the rain rate, and the optimal percentile was selected by comparing error statistics with reference data.

CHAPTER 5 IMPROVEMENT AND
BLENDING OF RAIN RATE

5.1 Improvement rain rate using DNN

5.1.1 The rain discrimination DNN classification model

The DNN classification model for rain discrimination classifies whether it is rain or non-rain using input data. ReLU was used for the active function of each hidden layer, and sigmoid was used for the active function of the output layer. Eventually, the value between 0 and 1, which is the result of the output layer, can be represented by probability. That is, it can be considered as a probability of the rainfall event. Therefore, in the end, it is judged that rainfall is less than threshold and non-rain is less. This threshold was also selected by selecting the most suitable threshold for each sensor through repeated results.

To generate the optimal model, combination of various hidden layers and the number of nodes was constructed. DNN Models were generated according to the combination of hidden units, and the optimal hidden units were selected by comparing their verification statistics. In addition, to avoid overfitting and find the optimal epochs of the model, early stopping was used. If the accuracy of the model's prediction was maintained or decreased

during the 5 updates, the model was stopped. Logloss was used as the prediction accuracy for early stopping, and probability of detection (POD), false alarm rate (FAR), Heidke skill score (HSS), and accuracy(ACC) were used to compare the dichotomous accuracy of the model.

$$POD = H/(H+M) \qquad (5.1)$$

$$FAR = F/(H+F) \qquad (5.2)$$

$$HSS = 2(HC-FM)/[(H+M)(M+C)+(H+F)(F+C)] \qquad (5.3)$$

$$Accuracy = (H+C)/N \qquad (5.4)$$

Here, H means the number of cases judged to be rainy in both the calculated data and the reference data. M is data that is judged to be non-rain in the model, but is judged to be rainfall in the reference. C means the number of data judged to be non-rain in both model and reference. N is the total number equal to H + F + M + C.

First, the accuracy of prediction, according to the combination of the number of nodes and the hidden layer (hyperparameter) for GMI, were compared (Table 5.1). As the 5-fold CV technique, the overall performance showed similar values. In the case of HSS and ACC, almost the same result was obtained, and POD and FAR are slightly different for each hidden unit case, but do not appear to be a big difference. Based on the accuracy, the

hidden unit composed of four hidden layers with 200 nodes that showed the highest ACC was selected as the optimal model structure. When the constructed CV model was confirmed, it was confirmed that most of the models had finished updating the model in about 20 epochs.

Table 5.2 shows the accuracy of prediction according to the hidden unit of SSMIS. Overall accuracy showed Slightly lower performance than GMI. In SSMIS, the optimal model structure was selected based on the accuracy of 9 cases in the same way as the optimal model structure selection of GMI. Four hidden layers with 300 nodes that showed the highest accuracy was selected as the optimal model structure.

Table 5.3 compares the accuracy of prediction according to the hidden unit case of MHS. MHS showed the highest accuracy in the structure of a model composed of four hidden layers in 200 nodes, similar to GMI.

Table 5.4, Table 5.5, and Table 5.6 show the difference in accuracy according to the classification threshold of each optimal model of GMI, SSMIS, and MHS. Commonly, the higher the threshold, the lower the values of POD and FAR. This is because the number of H and F decreases in equations 5.1 and 5.2 as the threshold increases. In this study, appropriate thresholds(GMI: 0.54, SSMIS: 0.46 and MHS: 0.51) of the models for individual sensors were selected based on HSS and ACC, which can be

considered comprehensively.

Table 5.1. POD, FAR, HSS and ACC of estimates derived from rain discrimination DNN of GMI against reference data(DPR)

Hidden units	POD	FAR	HSS	ACC
100x2	0.900	0.090	0.811	0.906
200x2	0.899	0.090	0.810	0.905
300x2	0.896	0.086	0.812	0.906
100x3	0.905	0.094	0.811	0.905
200x3	0.898	0.088	0.811	0.905
300x3	0.879	0.074	0.809	0.904
100x4	0.893	0.083	0.813	0.906
200x4	0.895	0.082	0.815	0.907
300x4	0.891	0.079	0.815	0.907

Table 5.2. POD, FAR, HSS and ACC of estimates derived from rain discrimination DNN for SSMIS against reference data(DPR).

Hidden units	POD	FAR	HSS	ACC
100x2	0.866	0.109	0.760	0.880
200x2	0.891	0.133	0.754	0.877
300x2	0.870	0.114	0.758	0.879
100x3	0.850	0.104	0.751	0.876
200x3	0.871	0.113	0.760	0.880
300x3	0.848	0.094	0.759	0.880
100x4	0.855	0.103	0.756	0.878
200x4	0.861	0.106	0.758	0.879
300x4	0.849	0.094	0.761	0.881

Table 5.3. POD, FAR, HSS and ACC of estimates derived from rain discrimination DNN for MHS against reference data(DPR).

Hidden units	POD	FAR	HSS	ACC
100x2	0.830	0.110	0.727	0.863
200x2	0.847	0.122	0.729	0.865
300x2	0.857	0.129	0.730	0.865
100x3	0.837	0.119	0.724	0.862
200x3	0.850	0.128	0.725	0.862
300x3	0.837	0.115	0.728	0.864
100x4	0.859	0.130	0.731	0.865
200x4	0.855	0.126	0.733	0.866
300x4	0.839	0.116	0.728	0.864
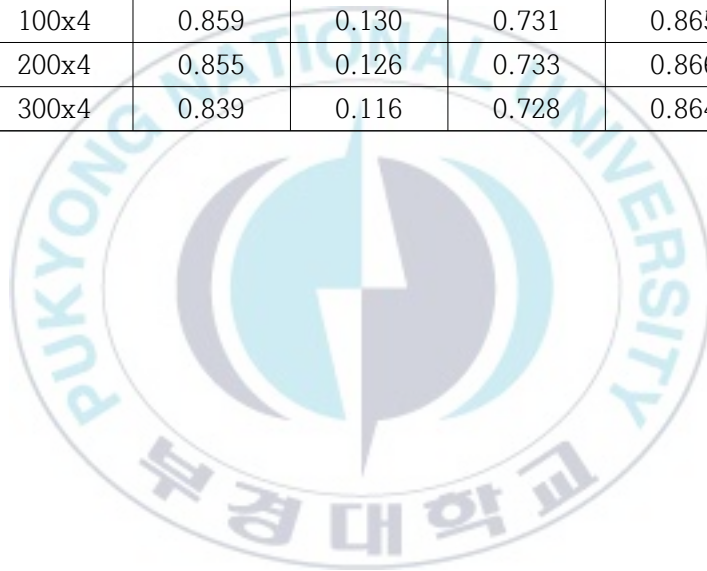
Table 5.4. POD, FAR, HSS and ACC according to threshold for GMI.

Threshold	POD	FAR	HSS	ACC
0.45	0.906	0.093	0.814	0.907
0.46	0.905	0.091	0.814	0.907
0.47	0.902	0.089	0.814	0.907
0.48	0.900	0.087	0.815	0.907
0.49	0.898	0.084	0.816	0.907
0.5	0.895	0.082	0.815	0.907
0.51	0.893	0.081	0.814	0.907
0.52	0.891	0.080	0.813	0.907
0.53	0.888	0.077	0.814	0.907
0.54	0.886	0.075	0.815	0.907
0.55	0.884	0.073	0.814	0.907
0.56	0.881	0.071	0.813	0.906
0.57	0.878	0.070	0.813	0.906
0.58	0.876	0.068	0.813	0.906
0.59	0.874	0.066	0.812	0.906
0.6	0.872	0.064	0.812	0.906

Table 5.5. POD, FAR, HSS and ACC according to threshold for SSMIS.

Threshold	POD	FAR	HSS	ACC
0.45	0.859	0.102	0.762	0.881
0.46	0.858	0.100	0.762	0.881
0.47	0.855	0.098	0.762	0.881
0.48	0.854	0.097	0.762	0.881
0.49	0.852	0.096	0.761	0.881
0.5	0.849	0.094	0.761	0.881
0.51	0.847	0.093	0.760	0.880
0.52	0.844	0.091	0.759	0.880
0.53	0.841	0.089	0.759	0.879
0.54	0.838	0.087	0.758	0.879
0.55	0.835	0.086	0.756	0.878
0.56	0.832	0.084	0.755	0.878
0.57	0.829	0.081	0.756	0.878
0.58	0.825	0.079	0.754	0.877
0.59	0.820	0.077	0.752	0.876
0.6	0.817	0.075	0.750	0.875

Table 5.6. POD, FAR, HSS and ACC according to threshold for MHS.

Threshold	POD	FAR	HSS	ACC
0.45	0.871	0.139	0.730	0.865
0.46	0.868	0.136	0.732	0.866
0.47	0.865	0.134	0.731	0.866
0.48	0.862	0.131	0.732	0.866
0.49	0.858	0.128	0.732	0.866
0.5	0.855	0.126	0.733	0.866
0.51	0.853	0.123	0.733	0.867
0.52	0.849	0.121	0.732	0.866
0.53	0.847	0.119	0.732	0.866
0.54	0.843	0.116	0.732	0.866
0.55	0.839	0.114	0.731	0.865
0.56	0.836	0.112	0.730	0.865
0.57	0.833	0.111	0.729	0.865
0.58	0.828	0.109	0.727	0.864
0.59	0.826	0.106	0.728	0.864
0.6	0.822	0.103	0.727	0.864

5.1.2 DNN regression model to retrieve rain rate

Rain rate retrieval DNN regression model uses the input data to retrieve the rain rate. ReLU was used for the activation function of the hidden layer and the output layer. The rain rate retrieval DNN regression model training and verification data are targeted for rainfall. That is, the number of non-rain events was excluded. It is performed sequentially with the rain discrimination DNN classification model, and the rain rate is calculated based only on the data determined to be rain in the rain discrimination DNN classification model.

The process for selecting the optimal model was the same as that of the rain discrimination DNN model. Mean bias error (MBE), mean absolute error (MAE), root mean square error (RMSE), and correlation coefficient (CC) were used as statistical for comparing accuracy.

Table 5.7 shows the error statistics for each hidden unit to select the optimal rain rate retrieval DNN regression model structure for GMI. After comparing the results of 9 cases and 5-fold CV technique, such as the rain discrimination DNN classification model, the optimal model structure was selected. The MBE, MAE, RMSE, and CC according to the hidden units appeared to be similar, although there were some differences. In this study, MAE was used as the first criterion for model selection. Therefore, the structure of the model consisting of two

hidden layers at 200 nodes was selected as the optimal structure of the rain rate retrieval DNN regression model. GMI used more BTs and is known to have better sensor accuracy. Overall, it showed a high correlation coefficient of 0.726 and the lowest MAE and RMSE among the three sensors.

The error statistics according to the hidden units for SSMIS are shown in Table 5.8. The error statistics of GMI showed better results than SSMIS. In SSMIS, the hidden unit composed of the 3 hidden layer with 300 nodes with the lowest MAE was selected as the optimal model configuration.

Table 5.9 shows the error statistics according to hidden unit for MHS. MHS showed lower correlation than GMI and SSMIS. This is due to the fact that MHS only has 5 high frequencies, and the basic ability of sensor is inferior to the previous two sensors. Accordingly, MAE and RMSE were also slightly higher than the previous two sensors.

In the case of MHS, when three hidden layers were composed of 100 nodes, MAE was the lowest, and this was selected as the optimal model structure.

Table 5.10 and 5.11 compare the results of the final selected DNN models with the accuracy of the rain rate of each sensor produced by the GPROF algorithm. It can be seen that the POD, FAR, HSS, and accuracy have improved for most of the three sensors. In particular, it can be seen that the POD showed a

marked improvement. As a result, HSS and ACC also improved.

On the basis of Accuracy, the improvement was 5.2% for GMI ((0.907-0.862) /0.862) and 4.9% for SSMIS ((0.881-0.847) /0.847), and 6.0% for MHS ((0.867-0.818) /0.818). Even in the retrieval of the rain rate, the error statistics (MB, MAE, RMSE, and CC) were clearly improved in all sensors when compared to the rain rate product of the GPROF algorithm. In terms of performance improvement, It showed improvement of 0.2mm/h overall based on MAE.

Table 5.12 shows the top two variables and the importance among the variable importance of the rain discrimination DNN classification model. The variable importance is obtained using the Gedeon method (Gedeon, 1997). In DNN, perceptrons are connected by a weight called a link. The links (weights) connected to each input data are synthesized to calculate importance. The BT of 183.31±7 GHz and BT of 36.64 GHz(V) for GMI, IWP and BT of 183.31±6 GHz for SSMIS, and IWP and BT of 157 GHz were ranked first and second. The region around 183 GHz is a frequency with low transmittance to water vapor. It is used as a frequency for observing water vapor. 36GHz is a region where the water vapor transmission rate increases, and is used to calculate SI along with 19 and 22GHz frequencies. IWP is the index that represents the total amount of water in the vertical space. 150GHz is the index that calculates IWP and SI,

and is a frequency related to rainfall. Table 5.13 shows the top two variables and the importance among the variable importance of the rain rate retrieval DNN regression model. The BTs of frequencies closely related to SI was selected for all sensors. As with previous studies, it was confirmed that SI has strength in estimating the rain rate.

Table 5.7. MBE, MAE, RMSE and CC of estimates derived from rain rate retrieval DNN of GMI against reference data(DPR)

Hidden units	MBE	MAE	RMSE	CC
100x2	−0.122	0.816	1.538	0.717
200x2	0.146	0.926	1.547	0.711
300x2	−0.151	0.807	1.518	0.725
100x3	−0.128	0.815	1.529	0.722
200x3	−0.053	0.813	1.506	0.726
300x3	0.025	0.875	1.518	0.724
100x4	−0.089	0.824	1.524	0.723
200x4	0.015	0.848	1.502	0.734
300x4	−0.017	0.853	1.536	0.720

Table 5.8. MBE, MAE, RMSE and CC of estimates derived from rain rate retrieval DNN of SSMIS against reference data(DPR)

Hidden units	MBE	MAE	RMSE	CC
100x2	0.285	1.082	1.758	0.582
200x2	0.293	1.092	1.758	0.582
300x2	0.041	0.976	1.722	0.587
100x3	0.078	0.996	1.729	0.583
200x3	0.067	0.999	1.736	0.579
300x3	−0.366	0.917	1.770	0.580
100x4	0.064	0.996	1.733	0.581
200x4	0.067	0.997	1.740	0.576
300x4	0.010	0.968	1.743	0.577

Table 5.9. MBE, MAE, RMSE and CC of estimates derived from rain rate retrieval DNN of MHS against reference data(DPR)

Hidden units	MBE	MAE	RMSE	CC
100x2	−0.064	1.085	1.941	0.505
200x2	−0.028	1.098	1.951	0.499
300x2	−0.033	1.109	1.946	0.507
100x3	−0.209	1.028	1.948	0.506
200x3	−0.006	1.125	1.953	0.503
300x3	−0.073	1.097	1.950	0.498
100x4	−0.294	1.036	1.961	0.506
200x4	−0.170	1.044	1.937	0.512
300x4	−0.076	1.077	1.935	0.510
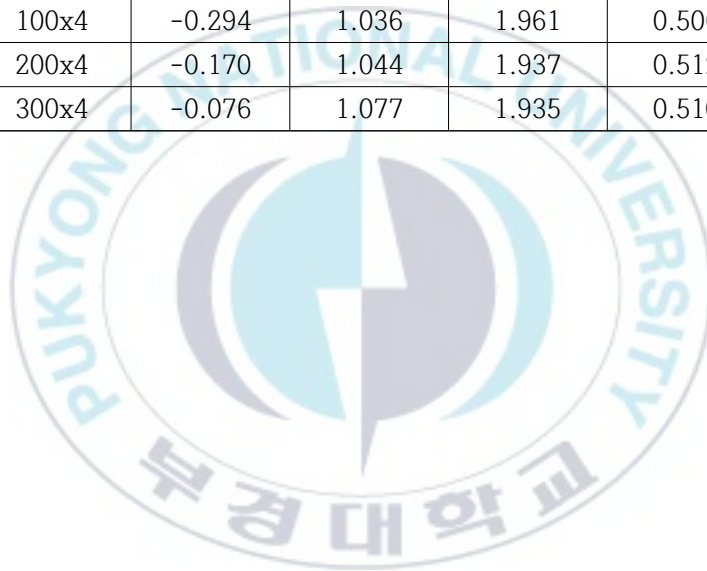
Table 5.10. POD, FAR, HSS and ACC of rain discrimination DNN classification model according to sensor

Rain discrimination DNN classification model					
		POD	FAR	HSS	ACC
GPROF	GMI	0.788	0.075	0.725	0.862
	SSMIS	0.815	0.129	0.694	0.847
	MHS	0.726	0.110	0.636	0.818
DNN	GMI	0.886	0.075	0.815	0.907
	SSMIS	0.858	0.100	0.762	0.881
	MHS	0.853	0.123	0.733	0.867

Table 5.11. MBE, MAE, RMSE and CC of rain discrimination DNN regression model according to sensor

Rain rate retrieval DNN regression model					
		MBE (mm/h)	MAE (mm/h)	RMSE (mm/h)	COR
GPROF	GMI	−0.182	1.040	1.901	0.637
	SSMIS	−0.330	1.103	2.074	0.460
	MHS	−0.358	1.260	2.254	0.444
DNN	GMI	−0.053	0.813	1.506	0.726
	SSMIS	0.041	0.976	1.722	0.587
	MHS	−0.209	1.038	1.948	0.506

Table 5.12. The top two ranked variable importance of rain discrimination DNN classification model according to sensor

ranking	GMI		SSMIS		MHS	
1	BT(183.31±7GHz)	10.7%	IWP	13.2%	IWP	18.3%
2	BT(36.64 GHz V)	9.7%	BT(183.31 ±6GHz)	9.8%	BT(157GHz)	17.7%

Table 5.13. The top two ranked variable importance of rain rate retrieval DNN regression model according to sensor

| ranking	GMI		SSMIS		MHS	
1	BT(23.8 GHz V)	13.2%	BT(37.0 GHz V)	13.2%	BT(89 GHz V)	18.9%
2	BT(36.64 GHz V)	12.7%	BT(22.235 GHz)	10.8%	BT(157 GHz)	18.6%

Rainfall has limitations in analysis by simple error analysis, with large errors occurring in strong rainfall and small errors occurring in light rainfall depending on the season. In general, the overall error increases in a season in which the range of rain rate is wide, which shows the seasonality of the error in time series. Even if the error is the same value, the influence varies depending on the season. This is also seen in Kidd et al. (2018). Figure 5.1 shows the bias and CCs of precipitation products of DPR and GPROF over time series in Europe and the United States as part of the study. To take this into account, normalized root-mean squares error (NRMSE) were obtained monthly and compared (Figure 5.2). GPROF was high in summer and relatively low NRMSE in spring and autumn. On the other hand, DNN showed a consistent degree of NRMSE according to the month. This means that a more stable rainfall intensity can be retrieved.

As an additional accuracy evaluation, differences in accuracy between land and ocean were compared (Figure 5.3). Land sea fraction data, one of the input data, was used to divide the results for land and ocean and analyze them individually. Both GPROF and DNN were more accurate than land in the ocean. This is thought to be because land is more complex than the ocean. In addition, the DNN model showed higher accuracy in both land and ocean. The accuracy improvement by optimizing

the combined DNN model was confirmed.

In Sanò et al (2018), rain rate was retrieved through GMI sensor using a neural network-based algorithm, and compared with the GPROF GMI product based on the ground radar data. Although it differed by region, it showed similar CC(0.51-0.60) and HSS(0.51-0.68) of GPROF algorithm analyzed in this study. The rain rate developed in the study showed accuracy (CC:0.56-0.61, HSS:0.51-0.7) similar to that of GPROF. In this study, direct comparison is difficult due to the difference in reference data, but it is considered that it showed better accuracy by relative comparison based on the accuracy of GPROF.
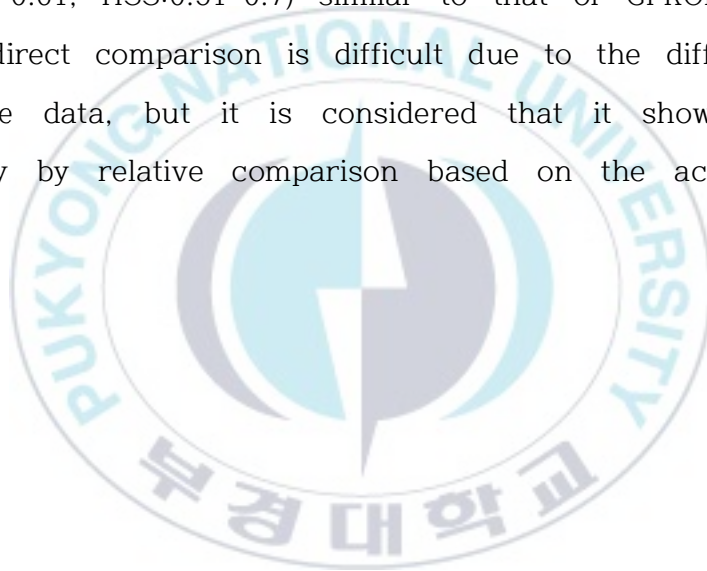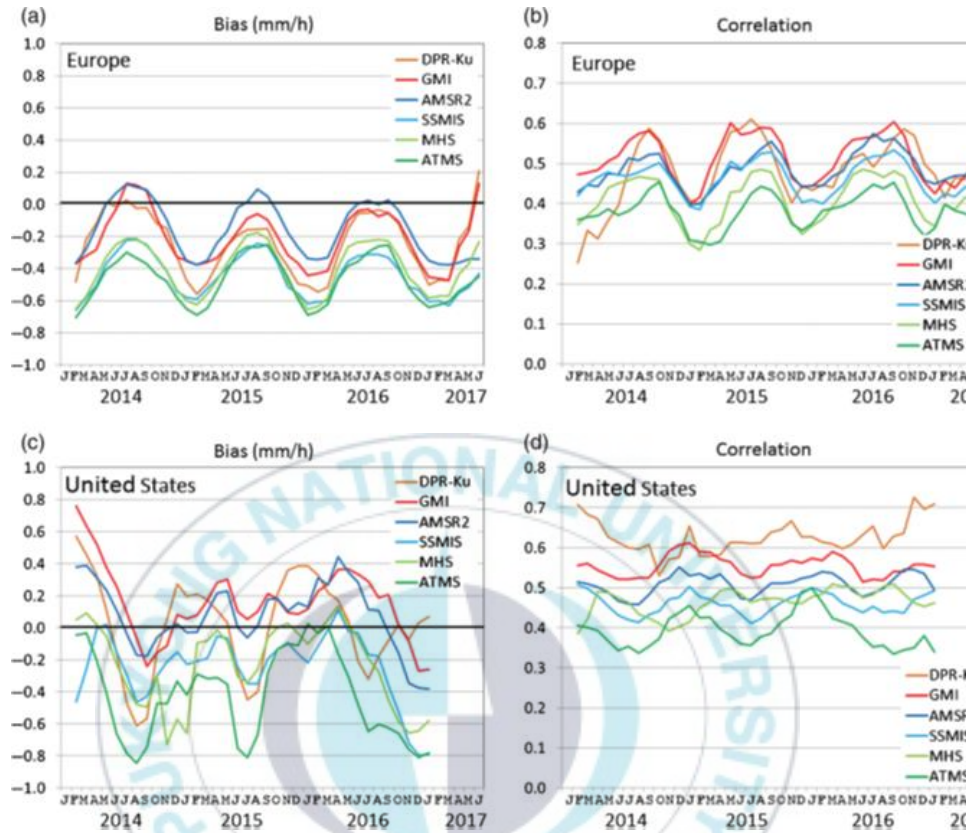
Figure 5.1 Monthly performance of instantaneous retrievals: bias in mm/hr (a, c) and (b, d) correlation for (a, b) the European region and (c, d) the United States.
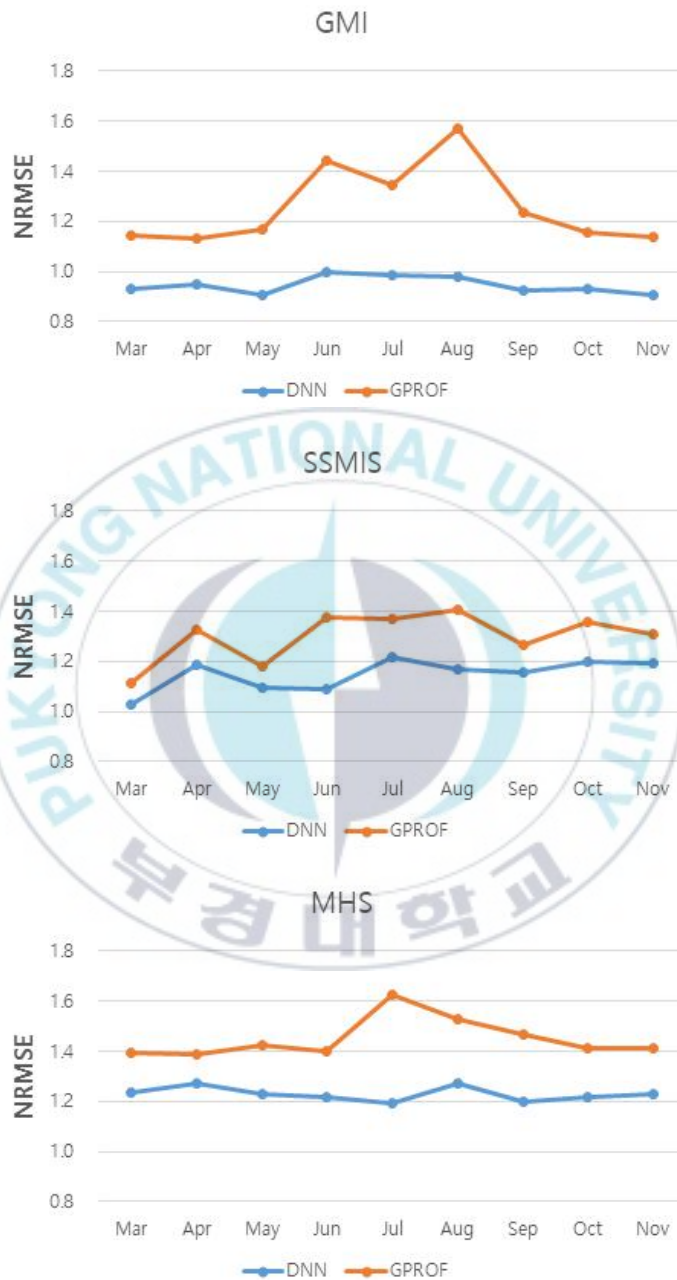
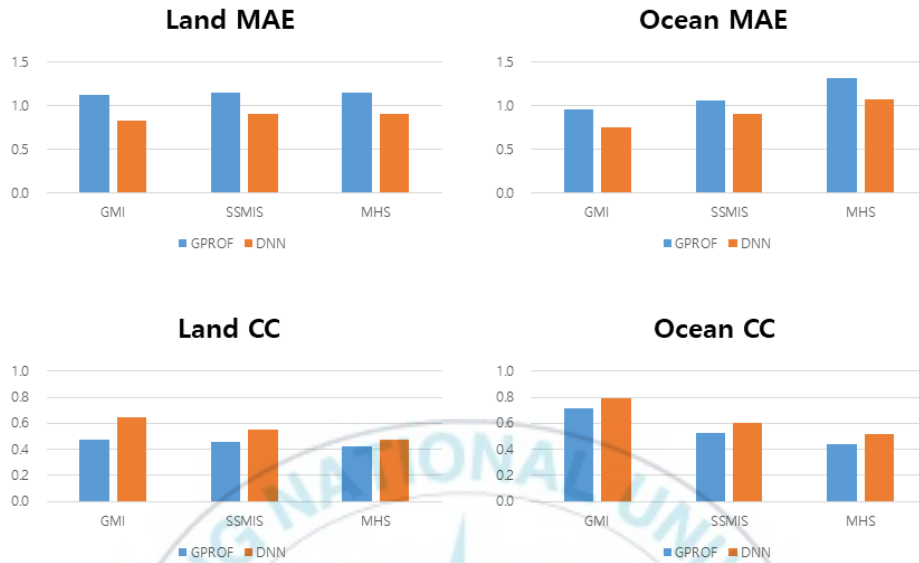**Figure 5.2** Monthly NRMSE of rain rate retrieved from DNN

Land MAE

Ocean MAE

Land CC

Ocean CC

Figure 5.3. MAE of rain rate retrieved from DNN by land and ocean

5.2 Blending rain rate of multi-sensor using the modified EBMA

In order to synthesize using EBMA, a preliminary work was first performed to generate an EBMA model.

As the first preliminary work, we redeployed the matchup between the blending members and the reference material. The matchup for DNN previously refers to a matchup between each sensor and reference. For blending using EBMA, all three members and reference data need to be commonly observed, that is, data detected by overlapping all three sensors and reference data is required. Here, the member means rain rate data of each sensor improved using DNN. The rain rate retrieved using DNN for each sensor was used to configure the number (311303) of overlapping 4 sensors as matchup data.

Next, a bias-correction process was performed. As a preliminary work, the three ECDFs by season(spring, summer, autumn) for the rain rate of the reference data were generated through the configured matchup. The ECDFs according to the season of the study area showed a similar distribution in summer and autumn, and there was a slight difference in spring due to the high rate of weak rainfall. Using the CDF, the rain rate of each member is once again bias-corrected.
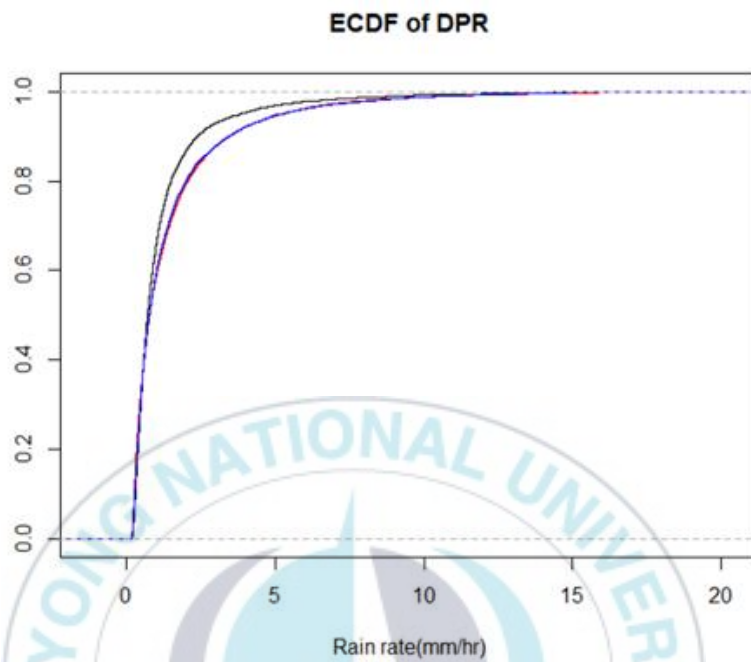
ECDF of DPR

Figure 5.4. ECDF of reference by season (black line: Spring, red line: Summer, blue line : Autumn)

The next preliminary work is to select the appropriate percentile. The result of EBMA is the weighted average PDF of each member's PDF. Therefore, in order to calculate the rainfall intensity through this, it is necessary to determine the appropriate percentile and to estimates the rain rate with the quantile value of the EBMA CDF. We calculated the rain rate from 40% to 80% in 5% increment and selected the appropriate percentile through error statistics in the reference data (Table 5.14). The CC according to percentiles were almost the same. However, the MAE, and RMSE decrease from 40% to 70% as the percentile increases, and it showed gradually increases again above 80%. From 65% to 70% as an inflection point, it was seen as an appropriate percentile.

EBMA blending was performed by applying the parameters for the EBMA obtained in the preliminary work. In order to compare the performance of EBMA blending, the average and median ensembles used for the blending of rain rate were calculated as comparison group and compared. Improved rain rate using DNN was used as input data for these two ensembles. As another comparison group, IMERG data, a multi-satellite rain rate synthesis map provided by NASA, were obtained and compared together (Table 5.15). It can be seen that EBMA blending shows better performance in MAE and RMSE than other blending techniques. EBMA blending was found to show about 44.2%

((1.181-0.819)/0.819) performance improvement compared to IMERG based on MAE. Figure 5.5, 5.6, 5.7, and Figure 5.8 show the scatter plots of the results of each blending techniques (EBMA, ensemble mean, ensemble median, and IMERG) according to the intensity of rainfall and the reference data. In the matchup for blending, the ensemble mean and ensemble median were overestimated at scatter plots below 5mm/h and 2.5mm/h where the data were dense, and the IMERG data showed low correlation. On the other hand, in the scatter plot of EBMA, high correlation was confirmed as the overestimation was relaxed compared to the ensemble mean and median using the same input data. This confirmed that there is an effect of improving accuracy through proper weight distribution and additional bias-correction. Table 5.16 and 5.17 show the errors of the composite maps in relatively strong rain rate. Although it is not as good as the error statistics of the overall match-up data, it can be seen that EBMA consistently shows the highest accuracy compared to other blending techniques. In addition, to understand the performance of EBMA blending, the PDF of the frequency of occurrence according to rainfall intensity was compared with other blending techniques(Figure 5.9). The PDF of the frequency of occurrence of EBMA was most similar to the distribution of DPR as a reference. In the ensemble mean and ensemble median, the frequency of rainfall under 1mm/hr was

small and the average was higher than DPR. This means that there is an overestimation in light rainfall, as seen in the scatter plot. IMERG had similar PDF averages, but the frequency distribution was different. Through various experiments, the conditions where the EBMA blending technique shows the highest performance were explored. Even under various conditions, EBMA showed the highest accuracy. In particular, "light" rainfall (0.1-2.5mm/hr) and "moderate" rainfall (2.5-10mm/hr) showed the largest difference compared to IMERG data. As a seasonal comparison of rainfall in this range, Figures 5.10, 5.11 and 5.12 are PDFs of the incidence of rainfall intensity over the season. Such as in the overall data, EBMA showed the most similar distribution to DPR. Tables 5.18, 5.19 and 5.20 show error statistics for rainfall below 10mm/hr depending on the season. While IMERG showed relatively large variability in CC and error according to the season, EBMA showed the best performance in all seasons, and the variability was almost consistent. Also, in the autumn, all blending techniques showed relatively high accuracy compared to other seasons, and EBMA also showed a high CC above 0.7.

Table 5.14. MBE, MAE, RMSE and CC of EBMA blending according to percentile on all matchup.

Percentile	MBE (mm/h)	MAE (mm/h)	RMSE (mm/h)	CC
0.4	−1.005	1.091	1.809	0.632
0.45	−0.927	1.047	1.750	0.638
0.5	−0.834	0.998	1.692	0.642
0.55	−0.720	0.941	1.634	0.643
0.6	−0.582	0.880	1.579	0.643
0.65	−0.426	0.832	1.536	0.642
0.7	−0.256	0.819	1.510	0.640
0.75	−0.064	0.844	1.510	0.637
0.8	0.159	0.919	1.550	0.634

Table 5.15. MBE, MAE, RMSE and CC of rain rate blended by EBMA, mean ensemble, median ensemble and IMERG

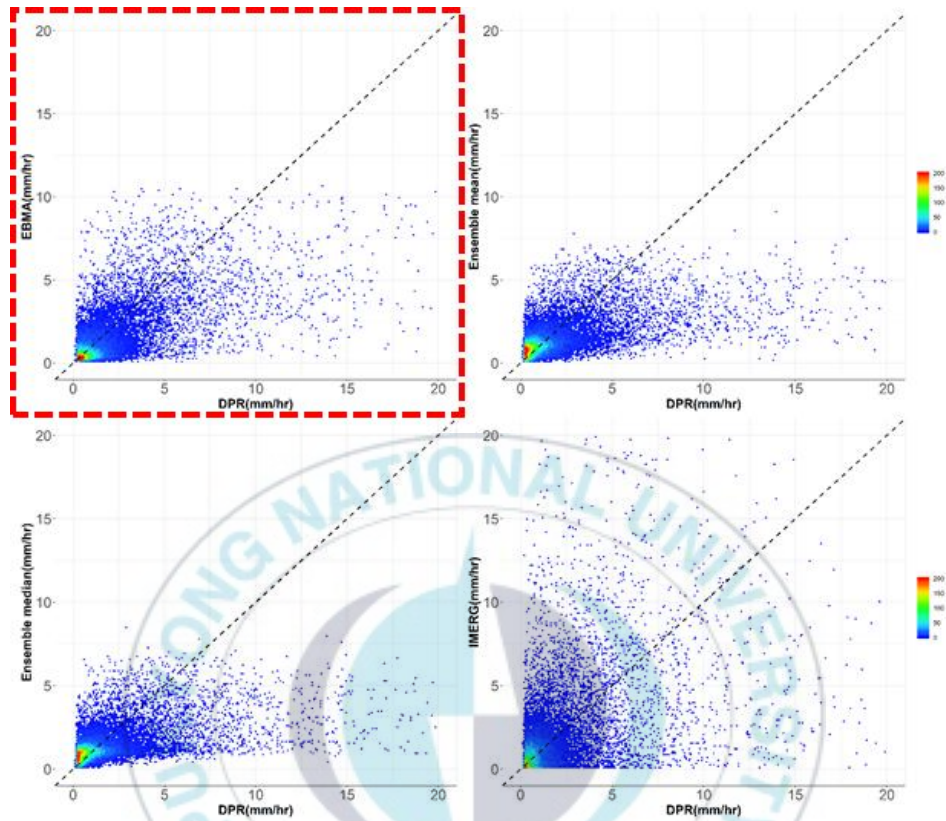| blending method	MBE (mm/h)	MAE (mm/h)	RMSE (mm/h)	CC
EBMA	−0.256	0.819	1.510	0.640
Ensemble Mean	−0.261	0.837	1.597	0.575
Ensemble Median	−0.321	0.885	1.670	0.519
IMERG	−0.197	1.181	2.427	0.432

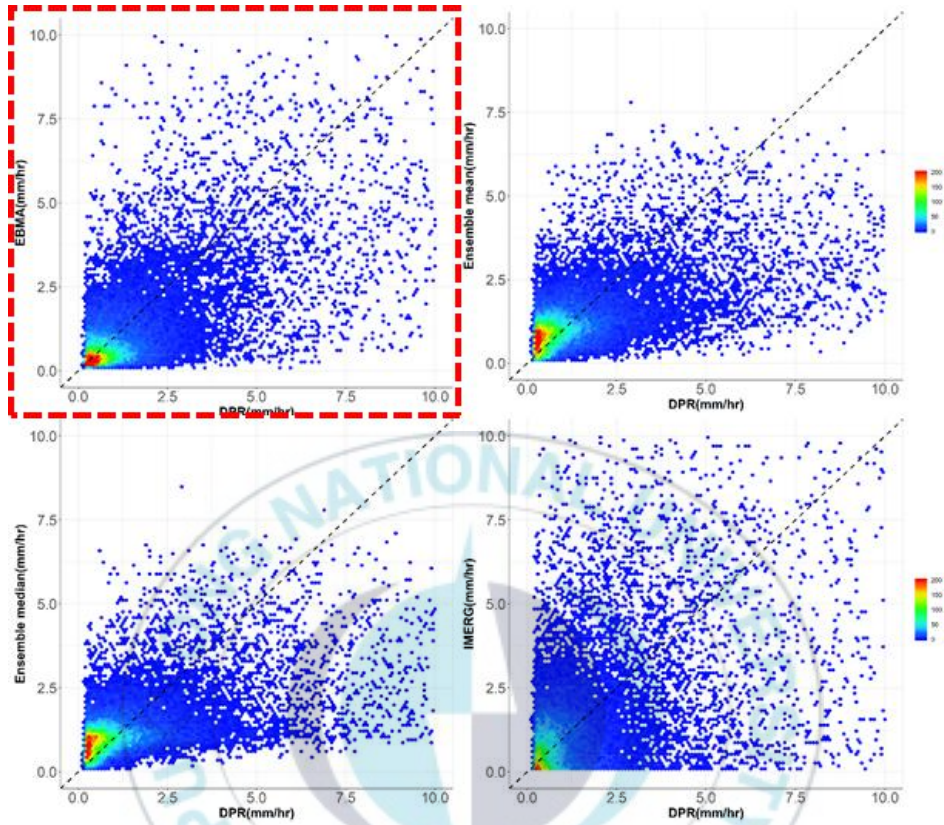**Figure 5.5.** Scatter plot of EBMA, mean ensemble, median ensemble and IMERG and reference (all Data)

Figure 5.6. Scatter plot of EBMA, mean ensemble, median ensemble and IMERG and reference (< 10mm/hr)

Table 5.16. MBE, MAE, RMSE and CC of rain rate blended by EBMA, mean ensemble, median ensemble and IMERG between 10mm/hr and 20mm/hr

10<RR<20 (mm/hr)	MBE	MAE	RMSE	CC
DNN with BMA	-8.629	8.629	9.263	0.159
Ensemble Mean	-10.013	10.013	10.397	0.125
Ensemble Median	-10.337	10.337	10.728	0.082
IMERG HQ	-7.282	9.211	10.757	0.048

Table 5.17. MBE, MAE, RMSE and CC of rain rate blended by EBMA, mean ensemble, median ensemble and IMERG between 5mm/hr and 10mm/hr

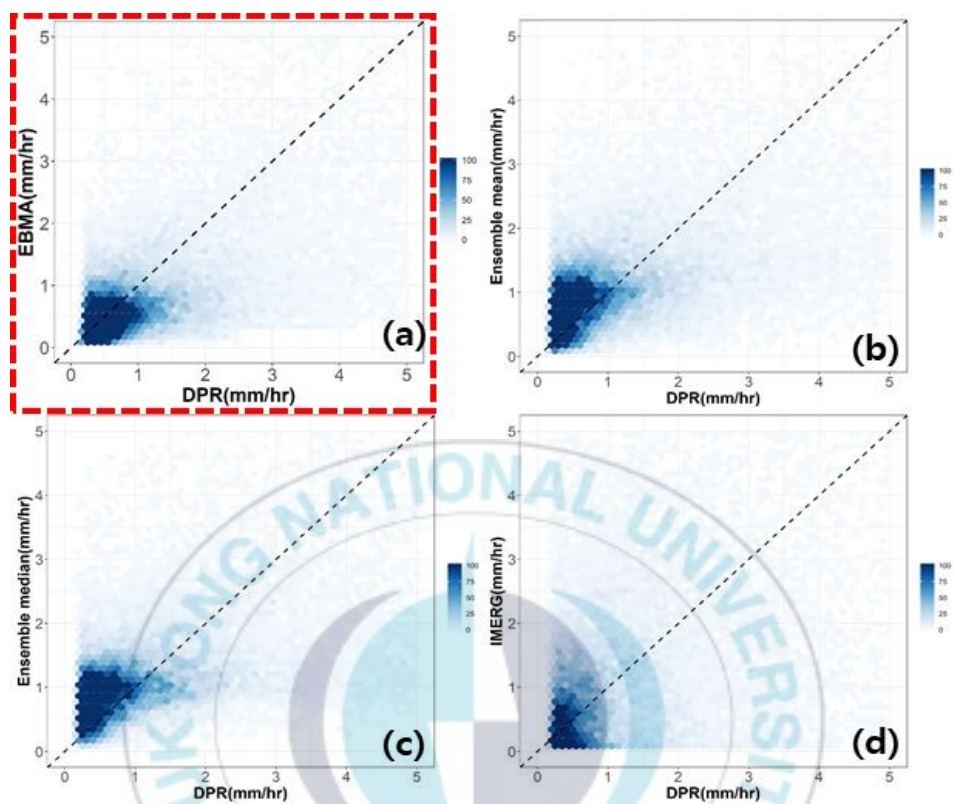| 5<RR<10 (mm/hr)	MBE	MAE	RMSE	CC
DNN with BMA	-2.922	3.268	3.813	0.185
Ensemble Mean	-3.904	3.934	4.335	0.137
Ensemble Median	-4.104	4.137	4.556	0.111
IMERG HQ	-1.697	4.789	6.375	0.045

Figure 5.7. Scatter plot of EBMA, mean ensemble, median ensemble and IMERG and reference (< 5 mm/hr)

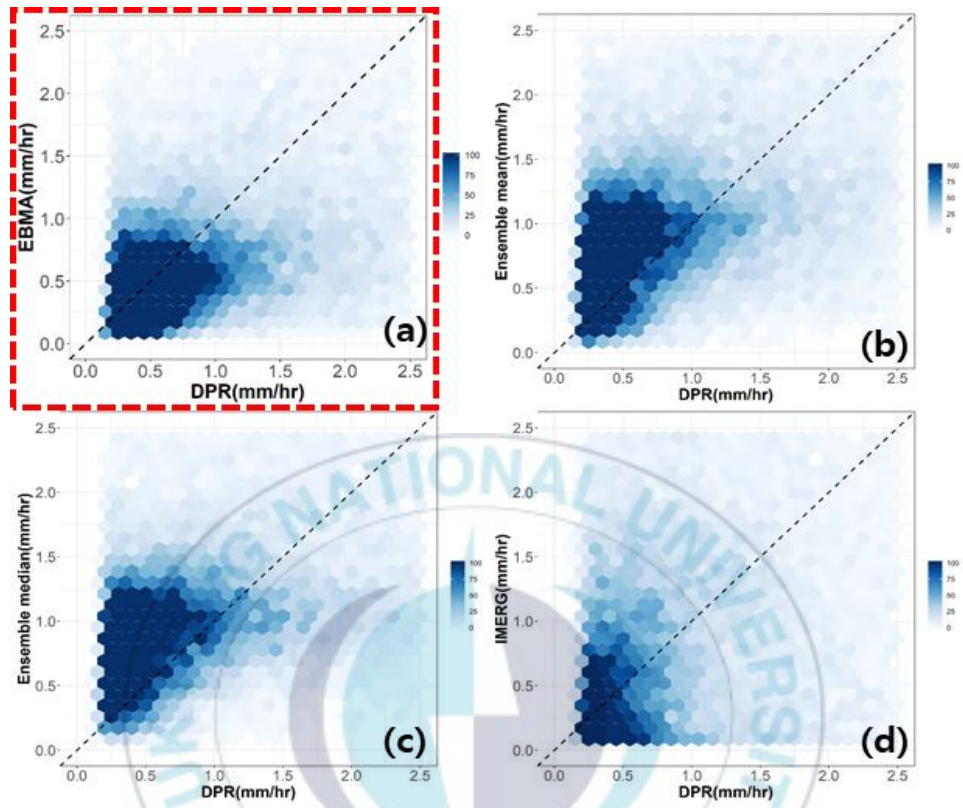**Figure 5.8.** Scatter plot of EBMA, mean ensemble, median ensemble and IMERG and reference (< 2.5 mm/hr)

All
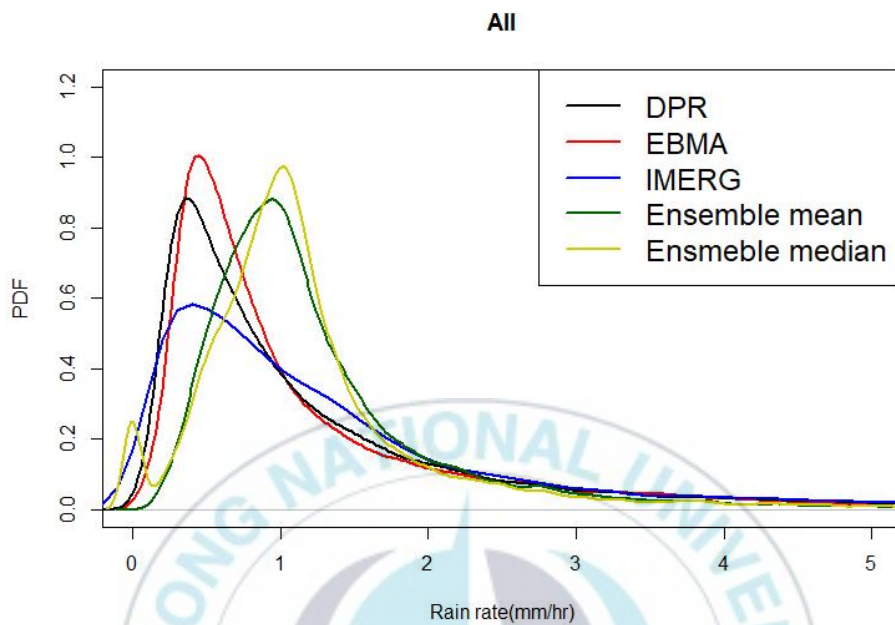
Figure 5.9. Distribution of the events of rainfall by intensity (mm/hr) for DPR, EBMA, IMERG, ensemble mean, and ensemble median (all data)
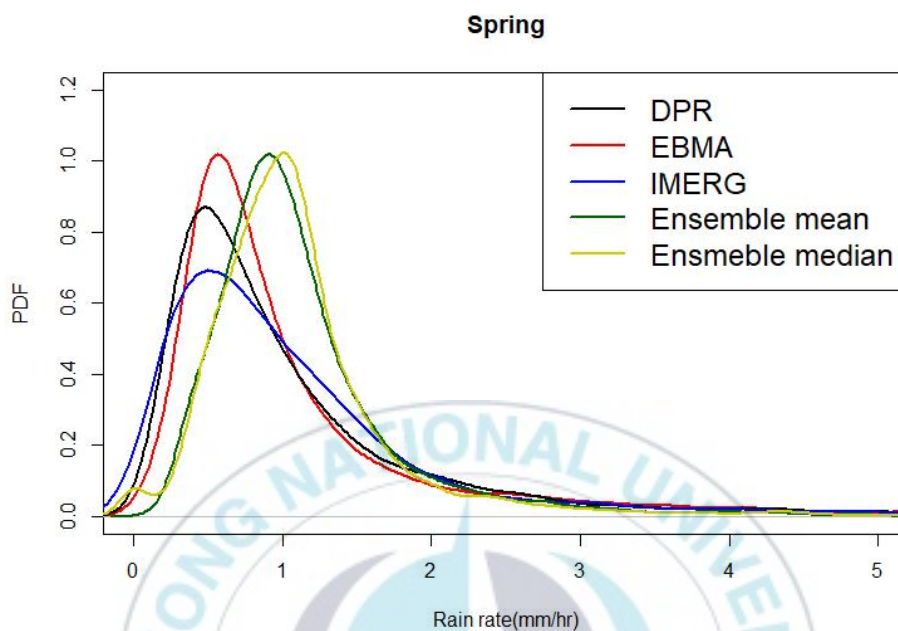
Spring

Figure 5.10. Distribution of the events of rainfall by intensity (mm/hr) for DPR, EBMA, IMERG, ensemble mean, and ensemble median (Spring)

Table 5.18. MBE, MAE, RMSE and CC of rain rate blended by EBMA, mean ensemble, median ensemble and IMERG (Spring)

|	MBE (mm/hr)	MAE (mm/hr)	RMSE (mm/hr)	CC
EBMA	−0.053	0.698	1.252	0.644
Ensemble mean	−0.111	0.709	1.321	0.576
Ensemble median	−0.139	0.742	1.388	0.499
IMERG	−0.237	0.987	2.009	0.342

Summer

Figure 5.11. Distribution of the events of rainfall by intensity (mm/hr) for DPR, EBMA, IMERG, ensemble mean, and ensemble median (Summer)

Table 5.19. MBE, MAE, RMSE and CC of rain rate blended by EBMA, mean ensemble, median ensemble and IMERG (Summer)

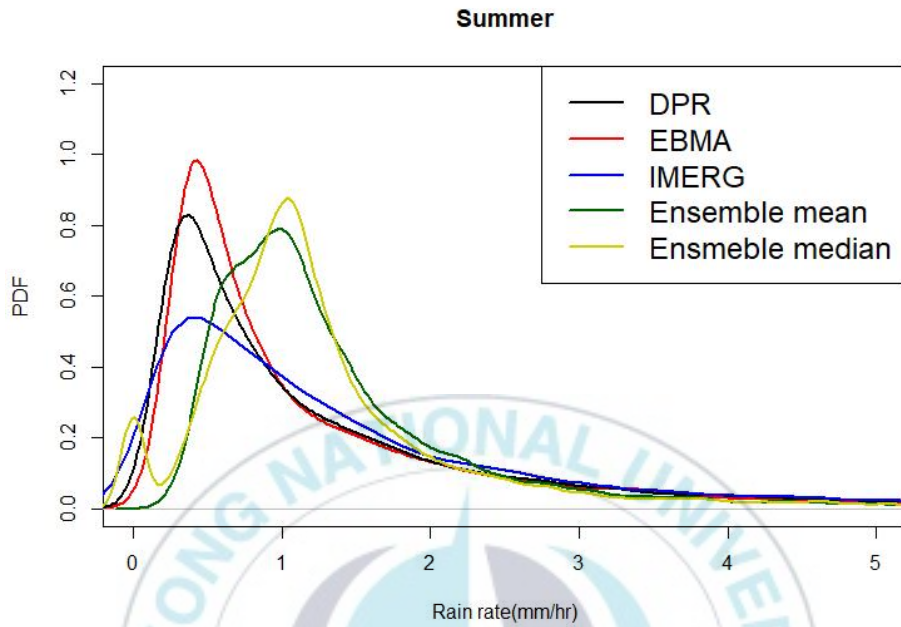|	MBE (mm/hr)	MAE (mm/hr)	RMSE (mm/hr)	CC
EBMA	−0.179	0.868	1.536	0.627
Ensemble mean	−0.134	0.930	1.582	0.574
Ensemble median	−0.221	0.982	1.664	0.514
IMERG	−0.040	1.373	2.790	0.417
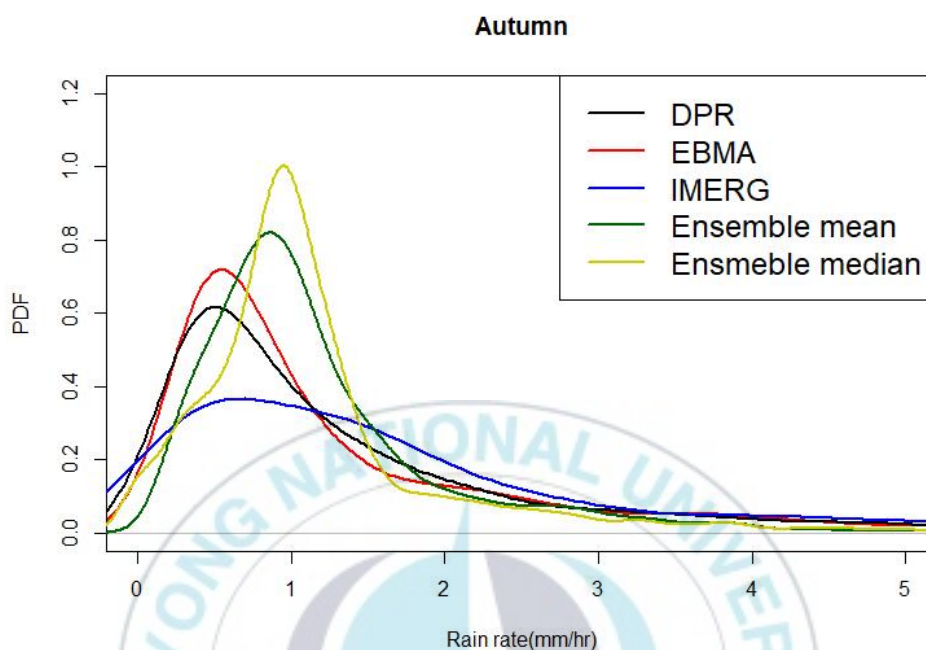
Autumn

Figure 5.12. Distribution of the events of rainfall by intensity (mm/hr) for DPR, EBMA, IMERG, ensemble mean, and ensemble median (Autumn)

Table 5.20. MBE, MAE, RMSE and CC of rain rate blended by EBMA, mean ensemble, median ensemble and IMERG (Autumn)

	MBE	MAE	RMSE	CC
EBMA	−0.167	0.817	1.405	0.716
Ensemble mean	−0.410	0.853	1.604	0.666
Ensemble median	−0.457	0.894	1.670	0.616
IMERG	0.007	1.263	2.318	0.566

5.3 EBMA blending map

In the previous section, the performance of EBMA blending was identified through error verification of regions where members are overlapped. However, the main goal of the blending is to expand the observation area based on multiple data. Therefore, a procedure for generating blending data for areas that do not overlap or areas where only some of the data overlap is required. Figure 5.13 shows an overview of the generation of blending map for this. For regions where all 3 members exist, the weights obtained through EBMA are applied to weighted-average PDFs of each member. When two members are present, the weights are re-distributed according to the weight ratio of the existing members. When only one member exists, bias-corrected was only applied.

Through the above process, an EBMA composite field was finally produced hourly. Two EBMA products are produced. The first is a composite field for rainfall probability and the other is a blending map for rain rate.

Figure 5.14 is a rainfall probability composite field for two specific dates(at 22:30 UTC on 7, August 2016 and at 07:30 UTC on 4 October 2018). It is a composite field showing the probability of rainfall more than 0.1mm/h on two dates. The probability of having a value above a certain rain rate can be expressed as a map.

Figures 5.15, 5.16, 5.17, and 5.18 are rain rate data for each DPR, ground radar, EBMA, ensemble mean, ensemble median, and IMERG at 08:30 UTC on 4, July, 2016, at 23:00 UTC on 10, July, 2016, at 19:00 UTC on 5, May, 2018, and at 22:00 UTC on 5, October, 2018. The rain rate of the reference data, DPR, is narrow, making it difficult to compare between blending data. Therefore, rain rate obtained through the ground radar provided by the Korea Weather Radar Center were additionally compared to compare the rainfall distribution. A common part was found in blending maps of the four days. The four blending maps showed almost the same distribution of rainfall. And there was some overestimation of IMERG data in areas suspected of strong rainfall based on the ground radar, while some in the ensemble mean and ensemble median underestimated in the same area. Relatively, EBMA showed similar rainfall distribution with small error compared to other blending map.

M = The number of available members of corresponding pixel
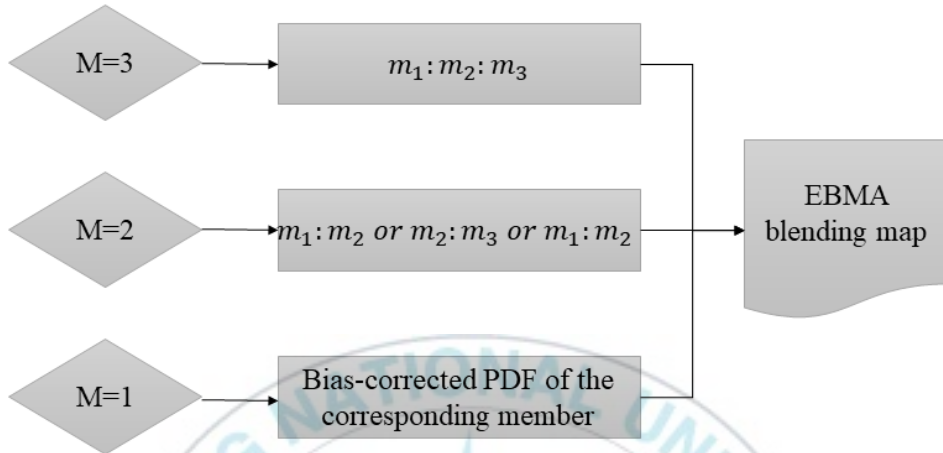
Figure 5.13. Procedure for generating EBMA blending map.
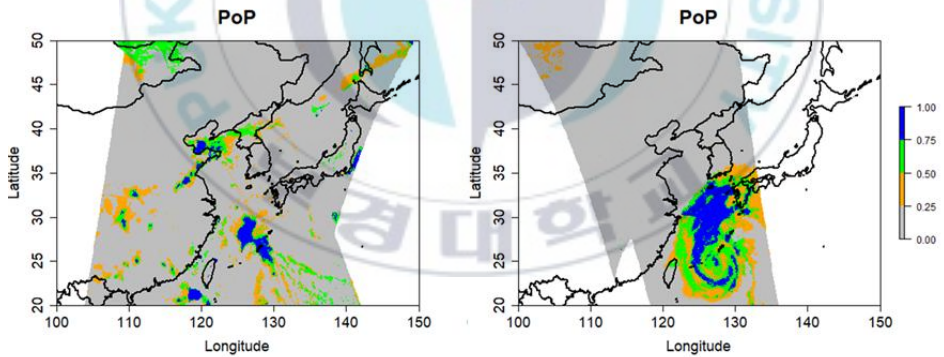
Figure 5.14. The blending probability of precipitation (PoP) using EBMA on (a) at 22:30 UTC on 7, August, 2016, and (b) at 21:00 UTC on 23, October, 2018
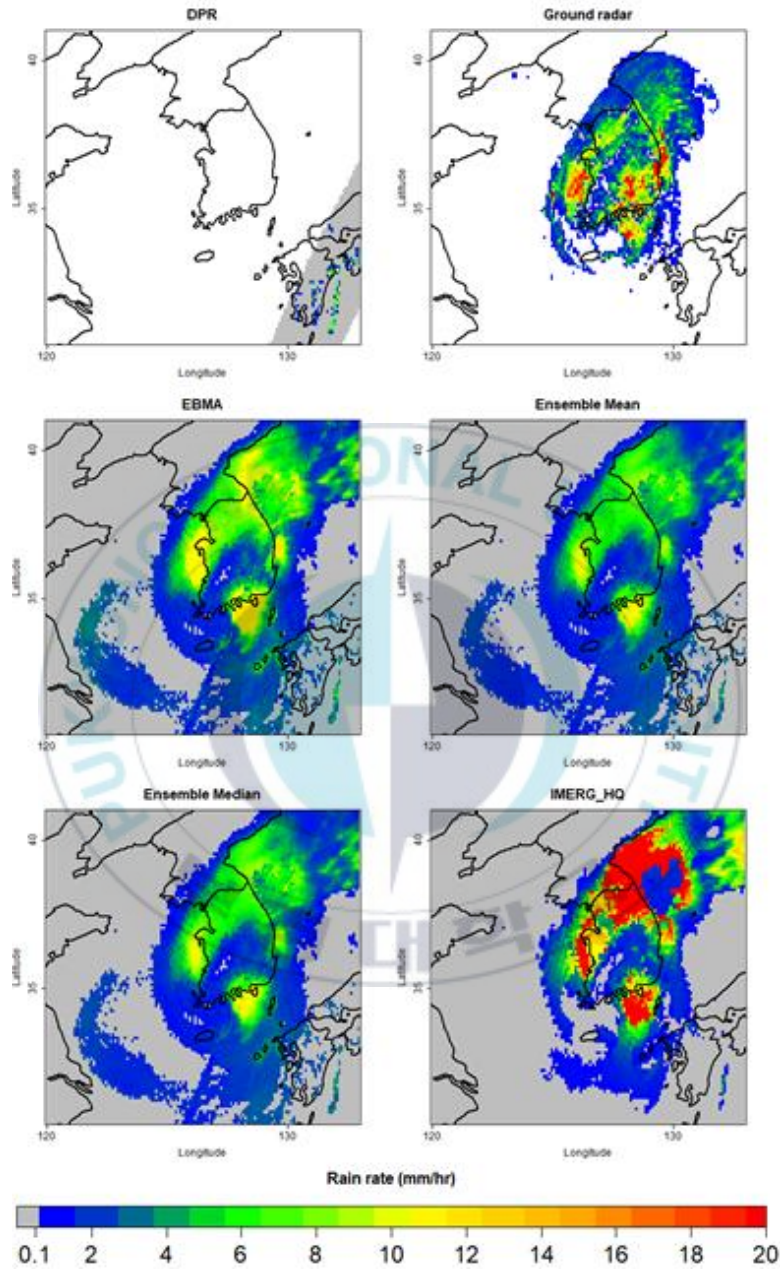
Figure 5.15. The DPR(a), Ground radar(b), blending rain rate maps(EBMA(c), ensemble mean(d), ensemble median(e), IMERG(f)) at 22:00 UTC on 05, October, 2018
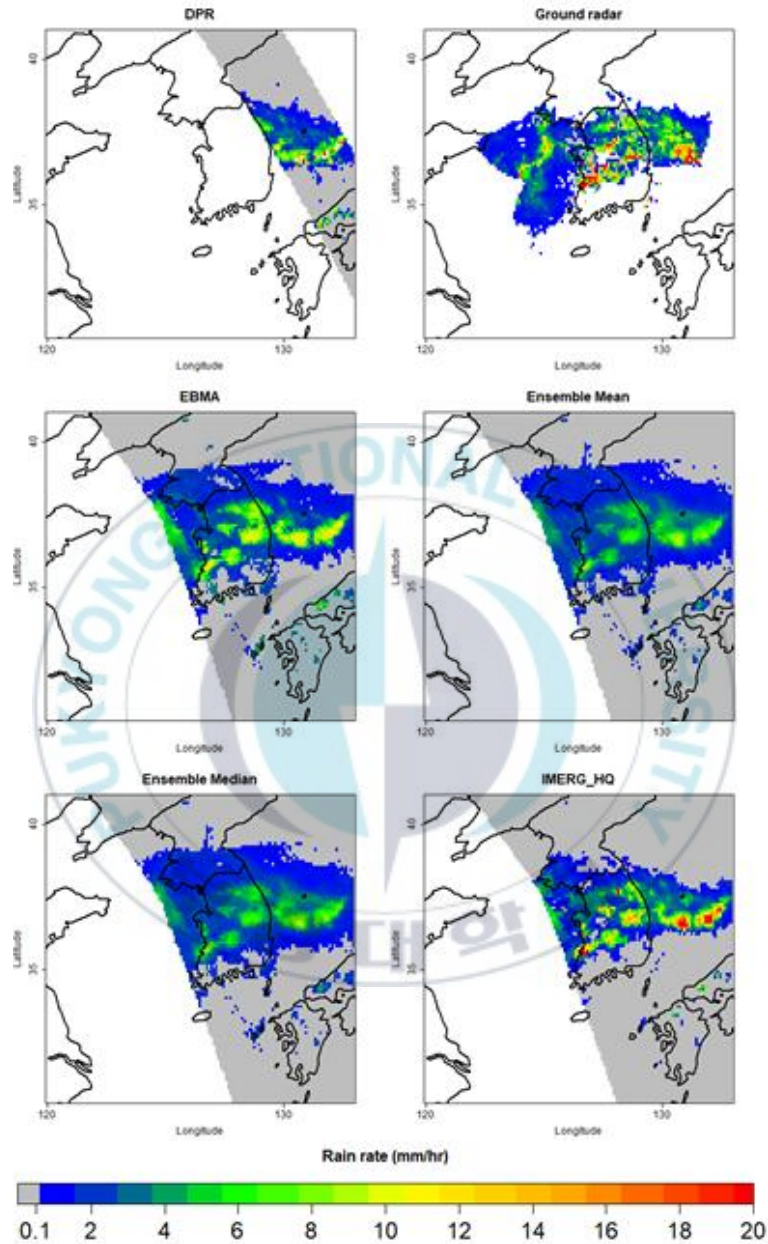
Figure 5.16. The DPR(a), Ground radar(b), blending rain rate maps(EBMA(c), ensemble mean(d), ensemble median(e), IMERG(f)) at 08:30 UTC on 4, July, 2016

Figure 5.17. The DPR(a), Ground radar(b), blending rain rate maps(EBMA(c), ensemble mean(d), ensemble median(e), IMERG(f)) at 19:00 UTC on 5, May, 2018
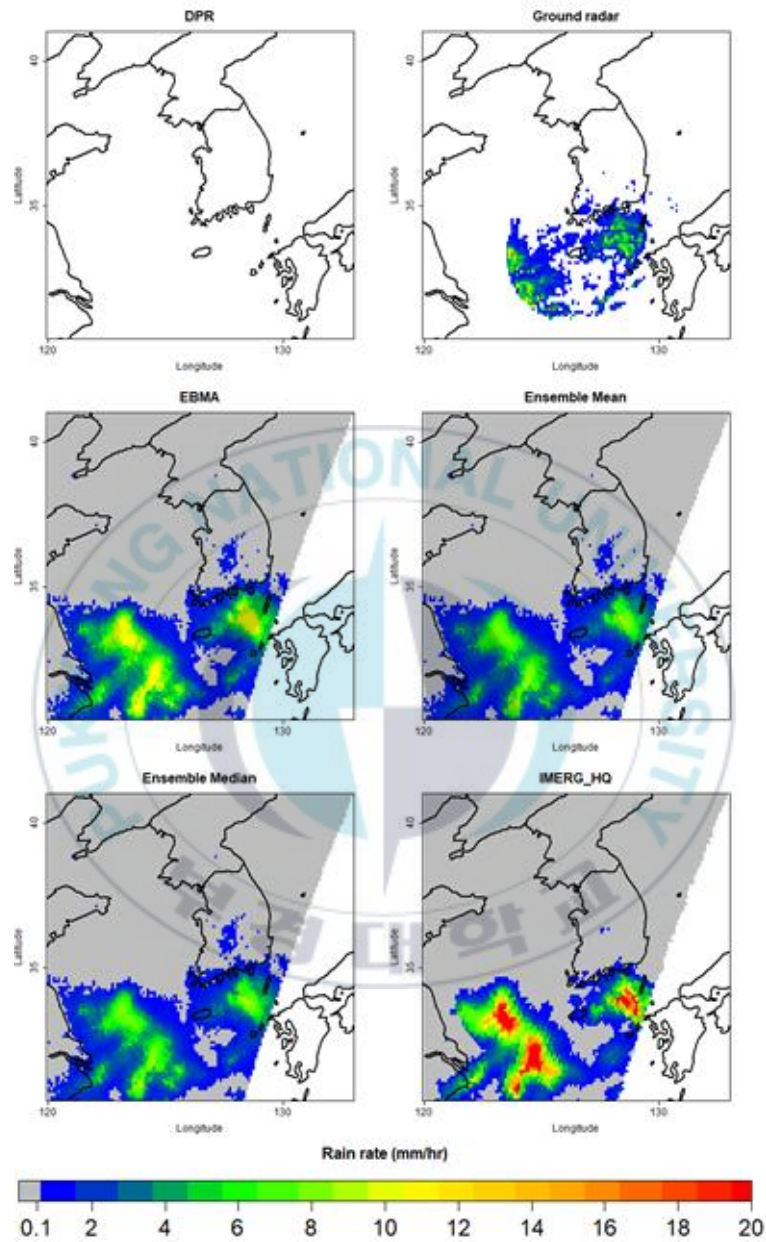
Figure 5.18. The DPR(a), Ground radar(b), blending rain rate maps(EBMA(c), ensemble mean(d), ensemble median(e), IMERG(f)) at 23:00 UTC on 10, July, 2016

CHAPTER 6 CONCLUSION AND FURTHER STUDIES

The purpose of this study is to improve rainfall product derived from PMW sensors and to blend the improved rainfall data in a reasonable way.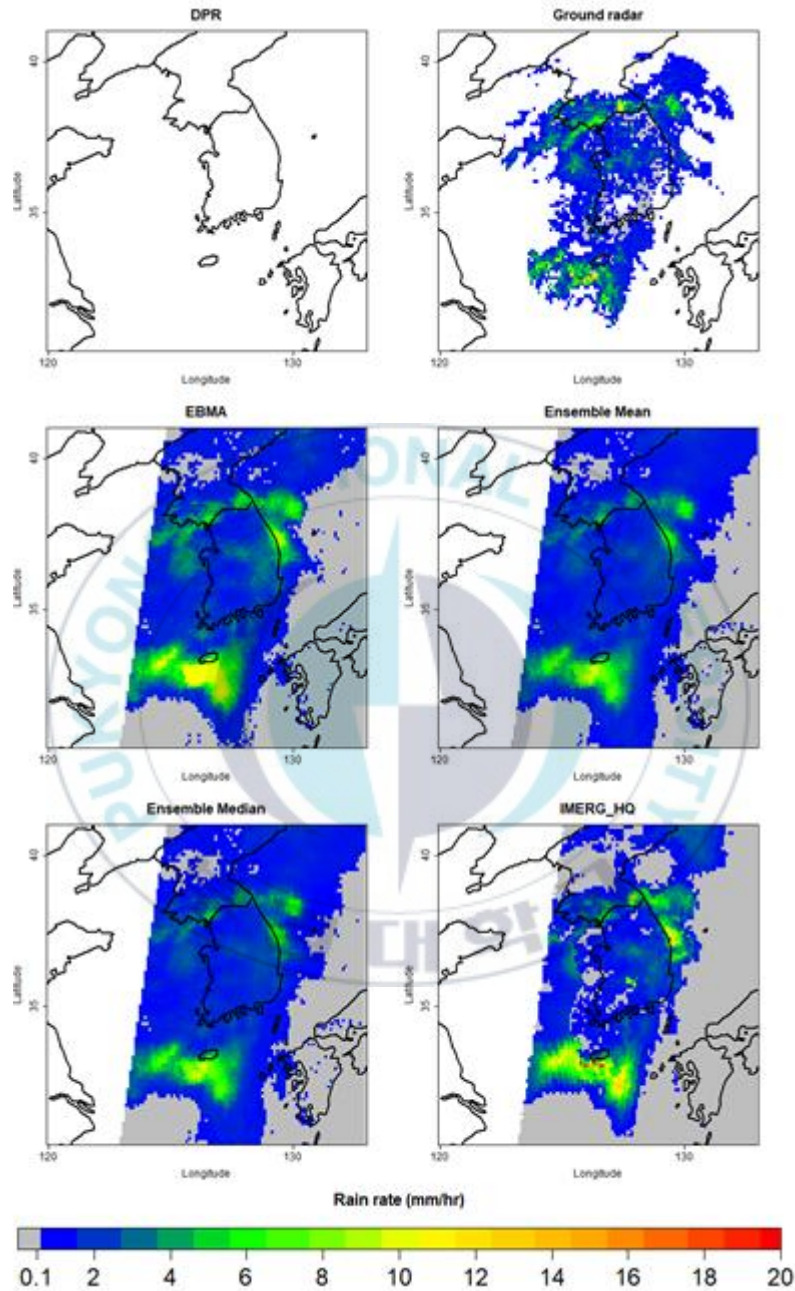 The target sensors were PMW sensor, and DNN was applied as a method for the improvement of rainfall. We developed a model that combined two models to properly apply DNN to rainfall based on satellite. The first model is for determining whether or not to rainfall, and the second is for retrieving the rain rate. Both models were individually optimized on sensors(GMI, SSMIS, MHS) through iterative process. The rain discrimination DNN classification model showed that the improvement of the accuracy up to 6.0% compared to the rainfall data retrieved from the GPROF algorithm. In the variable importance, IWP and BT at the frequency of 187GHz showed high importance in most models. BT at 187 GHz has low transmittance to moisture in the atmosphere, so it can predict the presence of raindrops. The IWP is an index calculated by BT of 150GHz and means the amount of ice and liquid in rain clouds. As a result, it can be confirmed that the amount of ice and water in the cloud is an important factor in determining whether or not there is rainfall.

In the rain rate retrieval DNN model, the structure of two hidden layers with 300 nodes for GMI and three hidden layers with 200 nodes for was selected as the optimal hidden unit configuration. And for MHS, the structure having three hidden layers with 100 nodes was selected. The rain rate retrieval DNN model also showed an improvement of MAE up to 17.8% compared to product of the GPROF algorithm. In addition, the CC was also improved from 0.673 to 0.726 (based on GMI), and it was confirmed that it was more consistent with the reference data than GPROF's.

The difference in uncertainty in the overlapping area was considered by using the EBMA blending technique, which is generally used for the ensemble of model data, to create a reasonable blending map. The PDF for the satellite-based rain rate is composed of a combination of logistic regression and gamma distribution. In order to apply EBMA to satellite data, improvement of the technique was performed based on the methodology of previous research. As improvements and changes, first, Bias correction using ECDF was performed on individual blending members. Next, we discovered that the exponent used for logistic regression and gamma distribution is unnecessary. Lastly, by selecting the appropriate percentile and calculating the rain rate, it was possible to generate a blending data more consistent with the reference data. The resulting

blending map covers all the observation areas of the members data, and discontinuities are not seen in the overlapping area. In addition, it was confirmed that overestimated improvements were improved in the overlapping area compared to IMERG data.

DNN was able to produce about 17.8% improved data compared to GPROF. The IMERG data is a blending data, and the accuracy is lower than that of the individual sensor GMI, but it was confirmed that the improvement of accuracy of 14.9% is achieved once again by applying a reasonable weight through EBMA. When used with DNN and EBMA, improving the accuracy of 30.4% (compared with GMI) was confirmed in areas where observations overlap. The error of rain rate showed an average improvement of 0.4-0.5mm/hr compared to IMERG. Although this seems to be a small difference numerically, it is considered that the standard for determining the presence or absence of rainfall is 0.1 mm/hr and 0.2-0.5 mm/hr is a range of drizzle rain, which is considered to be a significant improvement.

As the results, the possibility of improvement in rain rate using DNN and the effect of double improvement of accuracy by appropriate weight distribution and weighted average using EBMA along with the expansion of coverage were confirmed.

REFERENCE

Ali, I., F. Greifeneder, J. Stamenkovic, M. Neumann, and C. Notarnicol, 2015. Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data, *Remote Sensing*, 7(12), pp. 16398-16421.

Agostinelli, F., M. Hoffman, P. Sadowski, and P. Baldi, 2014. Learning activation functions to improve deep neural networks. *arXiv preprint arXiv*:1412.6830.

Aitkenhead, J.J., and I.H. Aalders, 2011. Automating land cover mapping of Scotland using expert system and knowledge integration methods, *Remote Sensing of Environment*, 115, pp. 1285-1295.

Bauer, P. 2001. Including a melting layer in microwave radiative transfer simulation for clouds. *Atmospheric research*, 57(1), pp. 9-30.

Bauer, P., E. Moreau, and S. Di Michele, 2005, Hydrometeor retrieval accuracy using microwave window and sounding channel observations. *Journal of Applied Meteorology*, 44(7), 1016-1032.

Behrangi, A., K.-L. Hsu, B. Imam, S. Sorooshian, and R. J. Kuligowski, 2009 Evaluating the utility of multispectral information in delineating the areal extent of precipitation. *Journal of Hydrometeorology*, 10, 684-700.

Bennartz, R., and G. W. Petty, 2001. The sensitivity of microwave remote sensing observations of precipitation to ice particle size distributions. *Journal of Applied Meteorology*, 40(3), pp. 345-364.

Berg, W., and R. Chase, 1992. Determination of mean rainfall from the Special Sensor Microwave/Imager (SSM/I) using a mixed lognormal distribution. *Journal of Atmospheric and Oceanic Technology*, 9(2), pp. 129-141.

Blended Rain Rate - Algorithm Description [Website] (2020, Mar 22) https://www.ospo.noaa.gov/Products/bRR/Algo.html

Brown, M.E., D.J., Lary, A., Vrieling, D., Stathakis, D., and Mussa, H., 2008. Neural networks as a tool for constructing continuous NDVI time series from AVHRR and MODIS, *International Journal of Remote Sensing*, 29(24), pp. 7141-7158.

Chang, A. T. C., L. S. Chiu, C. Kummerow, J. Meng, and T. T. Wilheit, 1999. First results of the TRMM Microwave Imager (TMI) monthly oceanic rain rate: Comparison with SSM/I. *Geophysical research letters*, 26(15), pp. 2379-2382.

Di Tomaso, E., F. Romano, and V. Cuomo, 2009, Rainfall estimation from satellite passive microwave observations in the range 89 GHz to 190 GHz. *Journal of Geophysical Research: Atmospheres*, 114(D18).

Erhan, D., Y. Bengio, A. Courville, P.A. Manzagol, and P. Vincent,

2010. Why does unsupervised pre-training help deep learning? *Journal of Machine Learning Research*, 11, pp. 625-660.

Ferraro, R. R. and G. F. Marks, 1995. The development of SSM/I rain-rate retrieval algorithms using ground-based radar measurements. *Journal of Atmospheric and Oceanic Technology*, 12(4), pp. 755-770.

Foody, G.M., 1996. Fuzzy modelling of vegetation from remotely sensed imagery, *Ecological Modeling*, 85, pp. 2-12.

Gedeon, T. D. 1997. Data mining of inputs: analysing magnitude and functional measures. *International Journal of Neural Systems*, 8(02), 209-218.

Gonzalez Vilas, L., E. Spyrakos, and J.M. Torres Palenzuela, 2011. Neural network estimation of chlorophyll a from MERIS full resolution data for the coastal waters of galician rias (NW Spain), *Remote Sensing of Environment*, 115, pp. 524-535.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.

Gottschalck, J., J. Meng, M. Rodell and P. Houser, 2005. Analysis of multiple precipitation products and preliminary assessment of their impact on global land data assimilation system land surface states. *Journal of Hydrometeorology*, 6(5), pp. 573-598.

Grody, N. C. 1991. Classification of snow cover and precipitation

using the Special Sensor Microwave Imager. *Journal of Geophysical Research: Atmospheres*, 96(D4), pp. 7423–7435.

Grumman, N. (2002). Algorithm and data user manual (ADUM) for the Special Sensor Microwave Imager/Sounder (SSMIS). Northrup Gruman Electronic Systems, Azusa, CA, Report, 12621, 69.

Hong, Y., K. L. Hsu, S. Sorooshian and X. Gao, 2004. Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *Journal of Applied Meteorology*, 43(12), pp. 1834–1853.

Hou, A. Y., R. K. Kakar, S. Neeck, A. A. Azarbarzin, C. D. Kummerow, M. Kojima, R. Oki, K. Nakamura and T. Iguchi, 2014. The global precipitation measurement mission. *Bulletin of the American Meteorological Society*, 95(5), pp. 701–722.

Jensen, J.R., F. Qiu, and M. Ji, 1999. Predictive modeling of coniferous forest age using statistical and artificial neural network approaches applied to remote sensing data, *International Journal of Remote Sensing*, 20(14), pp. 2805–2822.

Jiang, D., Yango, X., Clinton, N., and Wang, N. (2004), An artificial neural network model for estimating crop yields using remotely sensed information, *International Journal of Remote Sensing*, 25(9), pp. 1723–1732.

Jiang, S., Ren, L., Hong, Y., Yong, B., Yang, X., Yuan, F., & Ma,

M. (2012). Comprehensive evaluation of multi-satellite precipitation products with a dense rain gauge network and optimally merging their simulated hydrological flows using the Bayesian model averaging method. *Journal of Hydrology*, 452, pp. 213-225.

Ji, C.Y., 2000. Land-use classificatioin of remotely sensed data using kohonen self-organizing feature map neural networks, *Photogrammetric engineering and remote sensing*, 66(12), pp. 1451-1460.

Joyce, R. J. and P. Xie, 2011. Kalman filter-based CMORPH. Journal of Hydrometeorology, 12(6), pp. 1547-1563.

Hsu, K. L., X. Gao, S. Sorooshian and H. V. Gupta, 1997. Precipitation estimation from remotely sensed information using artificial neural networks. *Journal of Applied Meteorology*, 36(9), pp. 1176-1190.

Hsu, K. L., H. V. Gupta, X. Gao and S. Sorooshian, 1999. Estimation of physical variables from multichannel remotely sensed imagery using a neural network: Application to rainfall estimation. *Water Resources Research*, 35(5), pp. 1605-1618.

Huffman, G. J., R. F. Adler, B. Rudolf, U. Schneider, and P. R. Keehn, 1995, Global precipitation estimates based on a technique for combining satellite-based estimates, rain gauge analysis, and NWP model precipitation information. *Journal of*

Climate, 8(5), 1284-1295.

Huffman, G. J., D. T. Bolvin, E. J. Nelkin, D. B. Wolff, R. F. Adler, G. Gu, Y. Hong K. P. Bowman, and E. F. Stocker, 2007, The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of hydrometeorology*, 8(1), 38-55.

Huffman, G. J., and C. Klepp, 2011, Fifth workshop of the international precipitation working group. *Bulletin of the American Meteorological Society*, 92(11), ES54-ES57.

Hu, X., and Q. Weng, 2009. Estimating impervious surfaces from medium spatial resolution imagery using the selforganizing map and multi-layer perceptron neural networks, *Remote Sensing of Environment*, 113, pp. 2089-2102.

Kidd, C., J. Tan, P. E. Kirstetter, and W. A. Petersen, 2018, Validation of the Version 05 Level 2 precipitation products from the GPM Core Observatory and constellation satellite sensors. *Quarterly Journal of the Royal Meteorological Society*, 144, 313-328.

Kim, K., M. Yoon, J. Cho, S. Hong, H. Yoon, H. Mo, and Y. W. Lee, 2016, Blending of satellite SST products using ensemble Bayesian model averaging (EBMA). *Remote Sensing Letters*, 7(9), 827-836.

Kim, K., J. Park, J. Baik, and M. Choi, 2017, Evaluation of topographical and seasonal feature using GPM IMERG and TRMM 3B42 over Far-East Asia. *Atmospheric Research*, 187, 95-105.

Kummerow, C., and L. Giglio, 1995: A method for combining passive microwave and infrared rainfall observations. *Journal of Atmospheric and Oceanic Technology*, 12, 33-45.

Kummerow, C. D., W. S. Olson, and L. Giglio, 1996, A simplified scheme for obtaining precipitation and vertical hydrometeor profiles from passive microwave sensors. *IEEE Transactions on Geoscience and Remote Sensing*, 34(5), 1213-1232.

Kummerow, C., Y. Hong, W. S. Olson, S. Yang, R. F. Adler, J. McCollum, R. Ferraro, G. Petty, D. B. Shin and T. T. Wilheit, 2001. The evolution of the Goddard Profiling Algorithm (GPROF) for rainfall estimation from passive microwave sensors. *Journal of Applied Meteorology*, 40(11), pp. 1801-1820.

Kummerow, C. D., D. L. Randel, M. Kulie, N. Y. Wang, R. Ferraro, S. Joseph Munchak, and V. Petkovic, 2015, The evolution of the Goddard profiling algorithm to a fully parametric scheme. *Journal of Atmospheric and Oceanic Technology*, 32(12), 2265-2280.

LeCun, Y., Y. Bengio, and G. Hinton, 2015. Deep learning. *Nature,* 521(7553), pp. 436-444.

Lee, E. J., J. G. Jhun, and C. K. Park, 2005, Remote connection of the Northeast Asian summer rainfall variation revealed by a newly defined monsoon index, *Journal of Geophysical Research,* 18, 4381-4393

Lee, J., and E. H. Lee, 2018, Evaluation of daily precipitation estimate from integrated MultisatellitE Retrievals for GPM (IMERG) data over South Korea and East Asia. *Atmosphere*, 28(3), 273-289.

Mahesh, C., S. Prakash, V. Sathiyamoorthy and R. M. Gairola, 2011. Artificial neural network based microwave precipitation estimation using scattering index and polarization corrected temperature. *Atmospheric research*, 102(3), pp. 358-364.

Menzel, W. P., 2005. Remote sensing applications with meteorological satellites. *The Solar Spectrum,* 3(10).

Mugnai, A., H. J. Cooper, E. A. Smith, and G. J. Tripoli, 1990, Simulation of microwave brightness temperatures of an evolving hailstorm at SSM/I frequencies. *Bulletin of the American Meteorological Society*, 71(1), 2-13.

Mugnai, A., E. A. Smith and G. J. Tripoli, 1993. Foundation of physical-statistical precipitation retrieval from passive microwave satellite measurements. Part11:Emission source and generalized weighting function properties of a time dependent cloud-radiation model. Journal of Applied Meteorology, 32, pp. 17-39

Nitta, T., 1987, Convective activities in the tropical western Pacific and their impact on the Northern Hemisphere summer circulation. J. *Meteor. Soc. Japan*, 65, 373-390.

Ninomiya, K. (1984). Characteristics of Baiu front as a predominant subtropical front in the summer northern hemisphere. Journal of the Meteorological Society of Japan. Ser. II, 62(6), 880-894.

Ninomiya, K., and T. Akiyama, 1992, Multi-scale features of Baiu, the summer monsoon over Japan and the East Asia. Journal of the Meteorological Society of Japan. Ser. II, 70(1B), 467-495.

Olson, W. S., 1989. Physical retrieval of rainfall rates over the ocean by multispectral microwave radiometry: Application to tropical cyclones. *Journal of Geophysical Research, Atmospheres*, 94(D2), pp. 2267-2280.

Prakash, S., A. K. Mitra, D. S. Pai, and A. AghaKouchak, 2016, From TRMM to GPM: How well can heavy rainfall be detected from space?. *Advances in Water Resources*, 88, 1-7.

Park, M. S., M. I. Lee, H. Kim, J. Im, and J. M. Yoo, 2016, Spatial and diurnal variations of storm heights in the East Asia summer monsoon: storm height regimes and large-scale diurnal modulation. *Climate Dynamics*, 46(3-4), 745-763.

Panegrossi, G., D. Casella, P. Sanò, S. Dietrich, A. C. Marra, and

M. Petracca, 2014, CDRD and PNPR passive microwave precipitation retrieval algorithms: Extension to the MSG full disk area within H-SAF. In *Proc. EUMETSAT Meteorol. Satell. Conf.*

Petty, G. W. and W. F. Krajewski, 1996. Satellite estimation of precipitation over land. *Hydrological sciences journal*, 41(4), pp. 433-451.

Pham, V., T. Bluche, C. Kermorvant, and J. Louradour, 2014, September. Dropout improves recurrent neural networks for handwriting recognition. *In Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on (pp. 285-290).* IEEE.

Qin, Y., Z. Chen, Y. Shen, S. Zhang, and R. Shi, 2014, Evaluation of satellite rainfall estimates over the Chinese Mainland. *Remote Sensing*, 6(11), 11649-11672.

Qiu, S., P. Pellegrino, R. Ferraro and L. Zhao, 2005. The improved AMSU rain-rate algorithm and its evaluation for a cool season event in the western United States. *Weather and forecasting*, 20(5), pp. 761-774.

Raftery, A. E., T. Gneiting, F. Balabdaoui and M. Polakowski, 2005. Using Bayesian model averaging to calibrate forecast ensembles. *Monthly weather review*, 133(5), pp. 1155-1174.

Santi, E., S. Pettinato, S. Paloscia, P. Pampaloni, G. Fontanelli, A. Crepaz, and M. Valt, 2014. Monitoring alpine snow using

satellite radiometers and artificial neural networks, *Remote Sensing of Environment*, 144, pp. 179-186.

Sarma, D. K., M. Konwar, S. Sharma, S. Pal, J. Das, U. K. De, and G. Viswanathan, 2008, An artificial-neural-network-based integrated regional model for rain retrieval over land and ocean. *IEEE transactions on geoscience and remote sensing*, 46(6), 1689-1696.

Sanò, P., G. Panegrossi, D. Casella, F. D. Paola, L. Milani, A. Mugnai, M. Petracca, and S. Dietrich, 2015. The Passive microwave Neural network Precipitation Retrieval (PNPR) algorithm for AMSU/MHS observations: description and application to European case studies. *Atmospheric Measurement Techniques*, 8(2).

Sanò, P., G. Panegrossi, D. Casella, A. C. Marra, L. P. D'Adderio, J. F. Rysman, and S. Dietrich, 2018. The passive microwave neural network precipitation retrieval (PNPR) algorithm for the CONICAL scanning Global Microwave Imager (GMI) radiometer. *Remote Sensing*, 10(7), 1122.

Skofronick-Jackson, G., and B. T. Johnson, 2011. Surface and atmospheric contributions to passive microwave brightness temperatures for falling snow events. *Journal of Geophysical Research: Atmospheres*, 116(D2).

Skofronick-Jackson, G., W. A. Petersen, W. Berg, C. Kidd, E. F. Stocker, D. B. Kirschbaum, R. Kakar, S. A. Braun, G. J.

Huffman, T. Iguchi, P. E. Kirstetter, C. Kummerow, R. Meneghini, R. Oki, W. S. Olson, Y. N. Takayabu, K. Furukawa and T. Wilheit, 2017. The Global Precipitation Measurement (GPM) mission for science and society. Bulletin of the *American Meteorological Society*, 98(8), pp. 1679-1695.

Sloughter, J. M. L., A. E. Raftery, T. Gneiting and C. Fraley, 2007. Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Monthly Weather Review*, 135(9), pp. 3209-3220.

Smith, E. A., X. Xiang, A. Mugnai and G. J. Tripoli, 1994. Design of an inversion-based precipitation proflie retrieval algorithm using an explicit cloud model for initial guess microphysics. *Meteorology and Atmospheric Physics*, 54(1-4), pp. 53-78.

Smith, E. A., J. E. Lamm, R. Adler, J. Alishouse, K. Aonashi, E. Barrett, P. Bauer, W. Berg, A. Chang, R. Ferraro, J. Ferriday, S. Goodman, N. Grody, C. Kidd, D. Kniveton, C. Kummerow, G. Liu, F. S. Marzano, A. Mugnai, W. Olson, G. Petty, A. Shibata, R. Spencer, F. Wentz, T. Wilheit, and E. Zipser, 1998, Results of WetNet PIP-2 project, *Journal of the Atmospheric Sciences*, 55, 1483-1536.

Smith, E. A., P. Bauer, F. S. Marzano, C. D. Kummerow, D. McKague, A. Mugnai, and G. Panegrossi, 2002, Intercomparison of microwave radiative transfer models for precipitating clouds. *IEEE transactions on geoscience and*

remote sensing, 40(3), 541-549.

Sorooshian, S., K. L. Hsu, X. Gao, H. V. Gupta, B. Imam, and D. Braithwaite, 2000, Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bulletin of the American Meteorological Society*, 81(9), 2035-2046.

Spencer, R. W., D. W. Martin, B. B. Hinton, and J. A. Weinman, 1983. Satellite microwave radiances correlated with radar rain rates over land. *Nature*, 304(5922), pp. 141-143.

Spencer, R. W., H. M. Goodman, and R. E. Hood, 1989, Precipitation retrieval over land and ocean with the SSM/I: Identification and characteristics of the scattering signal. *Journal of Atmospheric and Oceanic Technology*, 6(2), 254-273.

Srivastava, N., G.E. Hinton, A. Krizhevsky, I. Sutskever, and Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *Journal of machine learning research*, 15(1), pp. 1929-1958.

Stephens, G. L., and C. D. Kummerow, 2007. The remote sensing of clouds and precipitation from space: A review. *Journal of the Atmospheric Sciences*, 64(11), pp. 3742-3765.

Takayabu, Y. N., and K. Hikosaka, 2009, Statistical analysis of oceanic rainfall characteristics in the Baiu season utilizing TRMM PR data. *Journal of the Meteorological Society of Japan. Ser*. II, 87, 339-352.

Tan, J., W. A. Petersen, G. Kirchengast, D. C. Goodrich, and D. B. Wolff, 2018. Evaluation of global precipitation measurement rainfall estimates against three dense gauge networks. *Journal of Hydrometeorology*, 19(3), 517-532.

Tomita, T., T. Yoshikane, and T. Yasunari, 2004, Biennial and lower-frequency variability observed in the early summer climate in the western North Pacific, *Journal of Climate*, 17, 4254-4266,

Weng, F., and N. C. Grody, 2000. Retrieval of ice cloud parameters using a microwave imaging radiometer. *Journal of the atmospheric sciences*, 57(8), pp. 1069-1081.

Weng, F., L. Zhao, R. R. Ferraro, G. Poe, X. Li and N. C. Grody, 2003. Advanced microwave sounding unit cloud and precipitation algorithms. *Radio Science*, 38(4).

Wilheit, T. T., A. T. Chang, and L. S. Chiu, 1991, Retrieval of monthly rainfall indices from microwave radiometric measurements using probability distribution functions. *Journal of Atmospheric and Oceanic Technology*, 8(1), 118-136.

Wilheit, T., R. Adler, S. Avery, E. Barrett, P. Bauer, W. Berg, A. Chang, J. Ferriday, N. Grody, S. Goodman, C. Kidd, D. Kniveton, C. Kummerow, A. Mugnai, W. Olson, G. Petty, A. Shibata and E. Smith, 1994. Algorithms for the retrieval of rainfall from passive microwave measurements. *Remote Sensing Reviews*, 11(1-4), pp. 163-194.

Wu, R., and B.Wang, 2000: Interannual variability of summer monsoon onset over the western North *Pacific and the underlying processes*. J. Climate, 13, 2483-2501.

Xu, W., E. J. Zipser, and C. Liu, 2009, Rainfall characteristics and convective properties of mei-yu precipitation systems over South China, Taiwan, and the South China Sea. Part I: TRMM observations. Monthly weather review, 137(12), 4261-4275.

Xu, W., and E. J. Zipser, 2011, Diurnal variations of precipitation, deep convection, and lightning over and east of the eastern Tibetan Plateau. *Journal of Climate*, 24(2), 448-465.

Yao, Z., W. Li, Y. Zhu, B. Zhao and Y. Chen, 2001. Remote sensing of precipitation on the Tibetan Plateau using the TRMM Microwave Imager. *Journal of Applied Meteorology*, 40(8), pp. 1381-1392.

Yokoyama, C., Y. N. Takayabu, and S. Kanada, 2014, A contrast in precipitation characteristics across the baiu front near Japan. Part I: TRMM PR observation. *Journal of climate*, 27(15), 5872-5890.

Yokoyama, C., Y. N. Takayabu, and T. Horinouchi, 2017, Precipitation characteristics over East Asia in early summer: Effects of the subtropical jet and lower-tropospheric convective instability. *Journal of Climate*, 30(20), 8127-8147.

You, J., X. Li, M. Low, D. Lobell, and S. Ermon, 2017. Deep Gaussian Process for Crop Yield Prediction Based on Remote

Sensing Data. *In AAAI* (pp. 4559-4566).

Zhao, L., and F. Weng, 2002. Retrieval of ice cloud parameters using the Advanced Microwave Sounding Unit. *Journal of Applied Meteorology*, 41(4), pp. 384-395.