Error Analysis of TRMM, WFD and APHRODITE datasets using Triple Collocation

P. Rathore
Error Analysis of TRMM, WFD and APHRODITE datasets using Triple Collocation

By

P. Rathore

in partial fulfilment of the requirements for the degree of

Master of Science
in Water Management

at the Delft University of Technology,
to be defended publicly on Wednesday October 29, 2014 at 02:30 PM.

Supervisor: Prof. dr. ir. Nick van de Giesen
Thesis committee: Dr. ir. Stephan de Roode, TU Delft
Ir. Martine Rutten, TU Delft

This thesis is confidential and cannot be made public until October 29, 2014.

An electronic version of this thesis is available at http://repository.tudelft.nl/.
Preface

This thesis marks the culmination of my Masters course in Water Management at the Faculty of Civil Engineering, Delft University of Technology. The research has been performed under the supervision of Prof. dr. ir. Nick van de Giesen and Dr. ir. Stephan de Roode with the guidance of Ir. Martine Rutten.

First of all, I would like to express my gratitude to the Delft Research-based Initiative and the Delft University of Technology for funding my masters with a scholarship. Without the scholarship, I would have missed an excellent opportunity to study in Delft. All the work for this research has been carried out at TU Delft with 8 weeks field studies in Myanmar. I am deeply indebted to my mentors, Dr. Nick van de Giesen and Dr. Stephan de Roode, for giving me the freedom to explore the topic and come up with my own work plan, but also supporting and guiding me whenever needed. I also extend my gratitude to Martine, who helped me throughout my work and also guided me whenever I hit a roadblock during my research. The research has been exhilarating for me. Special thanks go to the Mr Rob Roebeling, Scientist at KNMI and Dr. Saleh Abdalla from ECMWF for helping me in getting a better understanding of the Triple Collocation technique.

Moreover, I also thank my close friends in Delft, Avigyan, Naresh, Aabhas, Shivaang, Dirk, Damien, Frans-Willem and others from the student board and Dispuut Watermanagement who made my stay in Delft fun, wonderful and extra-ordinary. Special thanks to my family and friends in India, whose never-ending love and support, helped me reach so far in my life.

Last but not the least, I would like to mention water, the source of life on earth which inspired and motivated me to study this beautiful subject and help make the world better by whatever small contribution I make through this research.

Pradeep Rathore
Delft,
The Netherlands
The use of global precipitation datasets such as TRMM, WFD etc. for data scarce regions is gaining popularity since they provide forcing input for hydrological models. They make up for the lack of ground based data or the poor quality of whatever is available in many parts of the world. Using these datasets would be perfect if they were free of errors. Unfortunately, this is not the case. The geo-spatial data obtained from satellites or reanalysis products are not direct measures of precipitation. They are derived from atmospheric parameters such as cloud depth, brightness temperature etc. (Huffman 2007). The conversion of these to precipitation is done using complex algorithms. Efforts are made to calibrate this data but still errors sneak in. Similarly the interpolated gauge data like APHRODITE also has errors because of the inability of interpolation techniques to capture the high spatio-temporal variability in Precipitation. Hence the error estimation of these datasets remains a big problem.

Lack of ground based data ensures there is no reference to check these global datasets against. In this research, Triple collocation technique is applied to 3 datasets namely APHRODITE, TRMM and WFD for the river basins in Myanmar. The technique gives an estimate of the residual errors in the 3 datasets (with uncorrelated errors) without using any ground measurements or true values (R. A. Roebeling 2012). This is the first time it has been used to estimate errors in Precipitation datasets on a daily scale. Though the errors are not absolute, the results give an insight into the relative quality of these datasets. The errors have been calculated in space and time. Hence both temporal and spatial error characteristics are analysed. The study period is from 1998-2001.

The results obtained show that for TRMM and WFD, the errors are concentrated and of higher magnitude. For APHRODITE, the errors are more evenly distributed in space. All three datasets showed high errors in the central dry parts and the delta region. Overall, APHRODITE seems to show lowest error values in space.

The temporal error characteristics were also different for the 3 datasets. WFD showed highest average and maximum errors. TRMM had some very high error peaks but was in general better than WFD. Looking at the maximum and Average errors, APHRODITE seems to be the best of the three. WFD also shows some error peaks at the onset and end of Monsoon season. This shows its inability to estimate the localized pre and post monsoon storms.

The assumption of uncorrelated errors was also verified post analysis. Errors for 2 locations, Bago and Yangon were used to make scatter plots. No strong correlation is visible in the scatter plots reinforcing the assumption that the errors are uncorrelated.

The research shows that it is possible to make qualitative and quantitative inferences about the errors in the global precipitation datasets in the absence of in-situ measurements. Based on this research, it is concluded that overall, APHRODITE is the best of the 3 datasets. The possibility of a merged dataset formed by combining these 3 based on the error patterns observed in this study should be explored further.
Contents

Preface .............................................................................................................................. 3
Summary ............................................................................................................................ 5
List of Figures ................................................................................................................... 9
List of Tables ................................................................................................................... 11
List of Acronyms and Symbols ........................................................................................ 13
1. Chapter 1 Introduction ............................................................................................... 15
   1.1 Background .......................................................................................................... 15
   1.2 Research Questions and Outline of Thesis ............................................................ 16
2. Chapter 2 Study Area and Data .................................................................................. 19
   2.1 Introduction .......................................................................................................... 19
   2.2 Study Area .......................................................................................................... 19
      2.2.1 Climate and demography .............................................................................. 20
      2.2.2 Topography and Land-use ........................................................................... 21
   2.3 Data ................................................................................................................... 22
      2.3.1 Tropical Rainfall Runoff Mission (TRMM) ................................................. 23
      2.3.2 Watch Forcing Data (WFD) ....................................................................... 24
      2.3.3 Asian Precipitation - Highly-Resolved Observational Data Integration Towards Evaluation (APHRODITE) .................................................. 26
3. Chapter 3 Methods .................................................................................................... 29
   3.1 Error Structure .................................................................................................... 29
      3.1.1 Definitions .................................................................................................... 29
   3.2 Errors and Calibration Coefficients ..................................................................... 30
   3.3 Triple Collocation ............................................................................................... 31
      3.3.1 Theory ........................................................................................................ 32
      3.3.2 Procedure ................................................................................................... 34
4. Chapter 4 Results and Discussions .......................................................................... 35
   4.1 Seasonal Errors in Precipitation .......................................................................... 35
      4.1.1 Spatial trends ............................................................................................... 36
      4.1.2 Temporal Trends ......................................................................................... 39
      4.1.3 Comparison with station data ..................................................................... 41
5. Chapter 5 Conclusions and Recommendations ....................................................... 47
6. Bibliography ............................................................................................................... 49
List of Figures

Figure 1: Catchment Map with Countries layer .......................................................................................... 19
Figure 2: Southwest Monsoon Advance and Retreat (Hussain 2013) .......................................................... 21
Figure 3: A- Digital Elevation Map; B- Soil Type Map; C- Land Use .............................................................. 22
Figure 4: Precipitation Gauges Used by GPCC .......................................................................................... 32
Figure 5: Average Spatial Precipitation Maps 1998 .................................................................................. 35
Figure 6: Average Spatial Precipitation Maps 1999 .................................................................................. 36
Figure 7: Error Maps for Pre-Monsoon Season 1998 .................................................................................. 36
Figure 8: Error Maps for Monsoon Season 1998 ....................................................................................... 37
Figure 9: Error Maps for Post-Monsoon Season 1998 .............................................................................. 37
Figure 10: Error Maps for Pre-Monsoon Season 1999 ............................................................................. 37
Figure 11: Error Maps for Monsoon Season 1999 .................................................................................... 38
Figure 12: Error Maps for Post-Monsoon Season 1999 .......................................................................... 38
Figure 13: Spatial Mean Error for 2000 ..................................................................................................... 39
Figure 14: Spatial Minimum Error for 2000 ............................................................................................... 40
Figure 15: Spatial Maximum Error for 2000 ............................................................................................. 40
Figure 16: Spatial Standard Deviation for error of 2000 ........................................................................ 40
Figure 17: Comparison with Bago station- 1998 ..................................................................................... 42
Figure 18: Comparison with Yangon station- 1998 .................................................................................. 42
Figure 19: Comparison with Bago station- 1999 ..................................................................................... 43
Figure 20: Comparison with Yangon station- 1999 .................................................................................. 43
Figure 21: Scatter plots for errors at Bago(top) and Yangon(bottom) 1998 ............................................... 44
Figure 22: Scatter plots for errors at Bago(top) and Yangon(bottom) 1999 ............................................... 45
List of Tables

Table 1: Information on Major basins (Thein n.d.) ................................................................. 20
Table 2: Land-use Myanmar ................................................................................................. 22
Table 3: Comparison Statistics with Station Data ................................................................. 41
List of Acronyms and Symbols

WATCH: Water and Global Change
TRMM: Tropical Rainfall Measuring Mission
WFD: WATCH Forcing Data
APHRODITE: Asian Precipitation Highly-Resolved Observational Data Integration Towards Evaluation
TMI: TRMM Microwave Imager
PR: Precipitation Radar
AMSU: Advanced Microwave Sounding Unit
SSM/I: Special Sensor Microwave/Imager
EOS: Earth Observing System
AMSR-E: Advanced Microwave Scanning Radiometer for EOS
GPCP: Global Precipitation Climatology Project
TCI: TRMM Combined Instrument
Tb: Brightness Temperature
NCDC: National Climatic Data Center
CPC: Climate Prediction Center
GPROF: Goddard Profiling Algorithm
CRU: Climate Research Unit
ECMWF: European Centre for Medium Range Weather Forecasting
ERA: ECMWF Re-Analysis
GPCC: Global Precipitation Climatology Center
ITCZ: Inter Tropical Convergence Zone
GTS: Global Telecommunication System
DEM: Digital Elevation Model
GTOPO30: 30 Arc Second Resolution DEM of the World
PRISM: Parameter-elevation Regressions on Independent Slopes Model
Chapter 1

Introduction

1.1 Background
Myanmar experiences floods annually resulting in huge damage to life and property. The official estimated damage in 2008 due to cyclone Nargis was reported to be USD 4.057 billion (Aung 2010), though some sources estimate it as high as USD 10 billion (The Star News 2008). The number of people affected was 2.4 million and the death toll was 84000 (Aung 2010). Nargis was an extreme event but even the yearly floods cause significant financial losses. The Paddy crops lost to floods in Myanmar in 2011 was 1.7 million tonnes (Mizzima.com 2011), the estimated cost of which according to the 2014 prices (Myanmar Times 2014) comes out to be USD 400 million. For a country like Myanmar where the Gross national per capita income (2011) was USD 1126 (UN Data 2012), losses of this magnitude can greatly damage the financial health of the country signifying the need for flood mitigation measures.

One of the best and widely used ways to deal with floods is developing a good flood warning system. This requires a good understanding of the rainfall runoff processes in the region which helps in developing the hydrological model. Developing a good hydrological model can lead to an effective flood warning system for Myanmar. But this task comes with its own challenges. The lack of precipitation and runoff data, and poor quality of whatever is available makes it difficult to calibrate and validate the model (Deltares 2014). One of the solutions is to use the global gridded precipitation datasets like TRMM(Tropical Rainfall Measuring Mission), reanalysis products like Water and Global Change (WATCH) Forcing Data (WFD) and interpolated gridded rain gauge observations like APHRODITE (Asian Precipitation - Highly-Resolved Observational Data Integration Towards Evaluation) as the forcing input.

These datasets also have errors since they are derived indirectly using complex algorithms (Huffman 2007), (G. G. Weedon 2010). Hence to have a good output, it is necessary to estimate the errors in the forcing data. This can be done if reliable precipitation measurements are available that can be used for comparison and calibration. But for Myanmar, this data is not available.

The lack of data for calibration and validation is one of the major problems in hydrological modelling. Satellite observations have resulted in data being available for data scarce regions with a broad spatial coverage and repeat temporal coverage (Lakshmi 2004). But even these datasets are not precise and in absence of ground data to compare them with, the errors in the data are difficult to estimate. Quantitative information on the spatial and temporal error structures in large-scale (regional or global) precipitation data sets is essential for hydrologic and climatic studies (R. A. Roebeling 2012). Geo-spatial precipitation data sets can be from satellites like TRMM, WFD and APHRODITE. In this study, the focus will be on these three datasets. There are differences in precipitation values obtained from different geospatial datasets making it difficult to say which one is better. If run-off data for the modelled catchment is available, the quality of input can be tested by its ability to simulate the runoff. But in places like Myanmar, the study area for this research, where runoff data is sparsely available, it is difficult to use modelled runoff to compare the gridded datasets. This leads the way for using triple collocation as a method of error estimation for these datasets. Triple collocation is a statistical method to determine the relative random errors in measurements if 3 measurements for the same entity are
available. It is a powerful tool to quantify error structures in large-scale data sets (R. A. Roebeling 2012). The method does not require a true value which makes it really useful for the current study.

Triple collocation has been used for estimation of surface wind speeds (Stoffelen 1998), Ocean wave heights (PETER A. E. M. JANSSEN 2006), soil moisture (K. Scipal 2008) etc., but has not been used much for precipitation datasets. A study was done for Europe (R. A. Roebeling 2012) but the spatial errors were estimated on a seasonal scale while the temporal errors were on a decadal scale. For hydrologic models, the precipitation data is required at daily or sub-daily time-scale. There have been no significant studies on application of triple collocation for precipitation data on a daily time-scale till now. This research uses triple collocation for estimation of errors in precipitation datasets at a daily time scale for the first time. The algorithm used to apply this technique makes it possible to use it even on a sub daily time scale.

The motivation of this project is to check the quality of precipitation data products by estimating the relative errors in TRMM, WFD and APHRODITE datasets. This will help in obtaining the best gridded precipitation dataset for Myanmar.

The temporal analysis is carried out for two data cities in Myanmar, Yangon and Bago. The reason for selecting these two points is the availability of daily precipitation data for these stations from 1967 onwards. The spatial analysis is carried out over the major river basins of Myanmar like Irrawaddy, Salween and Sittaung. It is expected that APHRODITE data will show best correlation with the observed data as it uses this observed data as an input though it will be interesting to see the correlations between TRMM, WFD and observed precipitation.

1.2 Research Questions and Outline of Thesis

The aim of this study is to estimate the random errors in geospatial precipitation datasets obtained via satellites, reanalysis or interpolation of in situ-data for major river basins in Myanmar. For the purpose of this study, the datasets used are TRMM, WFD and APHRODITE. The following research questions are formulated as a part of this thesis:

1. What are the spatial error characteristics of the three datasets mentioned above for Myanmar?
2. What are the temporal error characteristics of the three datasets mentioned above for Myanmar?
3. Is the variability in the error characteristics same or different for these datasets over different temporal scales, i.e. daily, monthly and annual/seasonal?
4. How can the knowledge of errors in these datasets help in selecting the best dataset for input forcing in a Hydrologic Model, e.g. best dataset for a flood prediction model can be different than one for drought prediction?

The remaining chapters of this thesis address the above questions. A brief overview is provided below.

Chapter 2(Study Area and Data) discusses the area of study and the data used in this research in detail. The chapter includes the details about the locations used for the temporal analysis of the datasets. It also focuses on the datasets as well as the algorithms used to generate the data, the spatial and temporal resolution and some preliminary sources of errors based on the algorithms used and the methods employed to capture the precipitation. A brief introduction about the history of these datasets is also included in this chapter along with reasons behind the choice of data for this study.
Chapter 3 (Methods) discusses the triple collocation technique in detail. It also provides the background and derivations for the equations used in this method for estimating collocated errors. An introduction into error structures is included which is necessary to understand the collocation technique. Finally the algorithm used to implement the technique is discussed in a step by step manner.

Chapter 4 (Results and Discussions) addresses the research questions mentioned earlier. This chapter first presents the results of the analysis described in chapter 3 and then tries to provide logical explanations for the results obtained. It also presents discussion on the anomalies giving more insight into the results and forming basis for conclusions.

Chapter 5 (Conclusions and Recommendations) answers the research question and lists some important conclusions resulting from this thesis and recommendations for future work.
Chapter 2
Study Area and Data

2.1 Introduction
In this chapter, the focus area for this research and the data used is discussed in detail. The study area is discussed first and information about the climate, topography, land-use, demography etc. is provided. The reasons for selecting this study area are also discussed. This is important in order to understand the results. Meaningful conclusions can be drawn by analyzing the results in conjunction with the information available about the study area, the data used and how they influence the results. After that, the datasets used in this research are discussed. The section includes the source of data, the acquisition/retrieval methods, corrections employed and some background information about the data. This shall provide insight into the reasons for similarities or differences in results that are produced.

2.2 Study Area
The study area for this research mostly lies in Myanmar with some part in China and Tibet as can be seen in the map-Figure 1.

Myanmar is situated in Southeast Asia between North Latitude 19°32' to 28°31' and East Longitude 92°29’ to 101° 10’. Its total land area is 676,577 km2 (Asian Development Bank 2013). It is expected that the results of this research will be useful in developing a hydrologic routing model for Myanmar that uses remote sensing data as input. For a hydrologic model, it is recommended to work on a catchment scale rather than a country scale since it is easier to close the water balance and check the results. Hence the catchments of all major rivers in Myanmar have been merged together to arrive at the total area under study. Most Myanmar Rivers flow within the national boundaries except Salween which originates in Tibet, flows through China before entering Myanmar from where it enters Thailand and finally ends in Andaman Sea. Hence the part of Salween River in China and Tibet is also a part of study area in this research.

Four major river basins included here namely Irrawaddy (locally known as Ayeyarwady), Chindwin (merges into Irrawaddy), Salween (pronounced as Thanlwin) and Sittaung. The Irrawaddy and Chindwin are often mentioned together as Irrawaddy-Chindwin basin or simply Irrawaddy basin. These 4 basins cover nearly all of Myanmar. Except Sittaung, all other rivers start in the mountainous northern part of Myanmar and flow towards south to the sea. Due to favourable climatic condition, the total surface water potential of
Myanmar is about 1000 km³/year from eight major river basins (Thein n.d.). More information on Major basins is provided in Table 1.

### Table 1: Information on Major basins (Thein n.d.)

<table>
<thead>
<tr>
<th>Basin Number</th>
<th>Basin Name</th>
<th>Drainage Area (1000 x km²)</th>
<th>Surface Water (km³)</th>
<th>Groundwater (km³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Chindwin</td>
<td>115.3</td>
<td>141.293</td>
<td>57.578</td>
</tr>
<tr>
<td>2</td>
<td>Upper Irrawaddy</td>
<td>193.3</td>
<td>227.920</td>
<td>92.599</td>
</tr>
<tr>
<td>3</td>
<td>Lower Irrawaddy</td>
<td>95.6</td>
<td>85.800</td>
<td>153.259</td>
</tr>
<tr>
<td>4</td>
<td>Sittaung</td>
<td>48.1</td>
<td>81.148</td>
<td>28.402</td>
</tr>
<tr>
<td>5</td>
<td>Rakhine State</td>
<td>58.3</td>
<td>139.245</td>
<td>41.774</td>
</tr>
<tr>
<td>6</td>
<td>Taninthari Division</td>
<td>40.6</td>
<td>130.927</td>
<td>39.278</td>
</tr>
<tr>
<td>7</td>
<td>Salween</td>
<td>158.0</td>
<td>257.918</td>
<td>74.779</td>
</tr>
<tr>
<td>8</td>
<td>Mekong</td>
<td>28.6</td>
<td>17.634</td>
<td>7.054</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td><strong>737.8</strong></td>
<td><strong>1081.885</strong></td>
<td><strong>494.713</strong></td>
</tr>
</tbody>
</table>

Various other aspects such as climate, topography etc. of the region have been discussed further in the next sections.

#### 2.2.1 Climate and demography

The climate of Myanmar can be described as tropical monsoon climate. It is characterized by strong monsoon influence, has considerable amount of sun, a high rate of rainfall and high humidity. It has 3 distinct seasons: The cold and dry winter from November to January/February (average monthly temp. between 20-24 °C), then the hot and dry season from February to April (average monthly temp. between 30-35 °C) which is followed by the wet season from May and October (Average monthly temp. 25-30 °C) (World Weather and Climate Information n.d.). Myanmar receives its annual rains mainly from the southwest monsoons from mid-May to mid-October. 90% of the annual rainfall in different regions of Myanmar is monsoonal (Thein). The spatial variability of precipitation is high with coastal areas receiving over 5000 mm of annual precipitation while the north western dry zone receives less than 1000 mm of precipitation annually (Thein). The delta region also receives precipitation over 2500 mm annually (Yangon – 2700 mm) (World Weather and Climate Information n.d.).

Monsoon climate is characterized by seasonal reversal of winds. The easterlies from equator change direction seemingly due to Coriolis effect (Netting 2003) and start moving towards west as they move to higher(northern hemisphere) or lower(southern hemisphere) latitudes. Hence in Myanmar, the monsoon enters from the Bay of Bengal from Southwest Direction as shown in Figure 2. Due to preceding hot summer, there is a low pressure on land compared to the sea. Hence the moisture laden winds move from ocean to land causing precipitation. If a mountain or highland lies in the path of these winds, then the windward side gets higher precipitation while the leeward side remains relatively dry. This can be the reason for low precipitation in the west central part of Myanmar. The southwest monsoon winds cannot cross the hills on the western coast of Myanmar. The branch entering from the delta region in the south meets the highlands in the east causing precipitation. The west central part is dry because it is on the leeward side. The monsoon is a complex phenomenon and there may be many more contributing factors for this spatial precipitation pattern.
This variability in precipitation becomes even more important due to the agrarian economy of the region. In 2010, the agriculture sector accounted for about 36% of GDP, down from 57% in 2001 (Asian Development Bank 2013). Because of this dependence on agriculture, availability of water is important not only in the right quantity but at the right time which emphasizes the need for more reliable models and hence more reliable data. The most important and common crop cultivated here is rice. Because of the high water demand of paddy and the country’s dependence on it, irrigation is very important in Myanmar. While many reservoirs and irrigation schemes have been built in the recent few years, several have not been fully developed. Tertiary and field-level canals do not exist or are in poor condition.

### 2.2.2 Topography and Land-use

Topography and Land-use are very important to understanding the precipitation patterns in a region since the physical processes causing precipitation are affected by atmospheric conditions but sometimes also by topography. If a longer time span is considered, Precipitation may also be affected by Land use due to moisture recycling (H. Kunstmann 2007).

Topographically, the study area can be divided into five regions. They are the northern and western mountains, the eastern plateau (Shan plateau), the central basin and coastal strip. The country has varied features ranging from mountains at an elevation of more than 6000m in the north to delta in the south. In between, there are plateaus with heights of around 2000m and the fertile floodplains of Irrawaddy River. The variation in topography can be observed from the Digital Elevation map of Myanmar Figure 3A.
Along with different soil types and elevations Figure 3B, the land-use in Myanmar is also varied. The percentage of paved area is very less and forests and woodlands account for 49% of the total land area. Also the net sown area is also significant (Table 2).

Table 2: Land-use Myanmar

<table>
<thead>
<tr>
<th>Type of Land</th>
<th>Area (1000 Acres)</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reserved Forests</td>
<td>41752</td>
<td>25</td>
</tr>
<tr>
<td>Current Fallow</td>
<td>597</td>
<td>0</td>
</tr>
<tr>
<td>Net Sown Area</td>
<td>29591</td>
<td>18</td>
</tr>
<tr>
<td>Occupied Area</td>
<td>30188</td>
<td>18</td>
</tr>
<tr>
<td>Culturable Waste</td>
<td>13861</td>
<td>8</td>
</tr>
<tr>
<td>Other woodland</td>
<td>40166</td>
<td>24</td>
</tr>
<tr>
<td>Others</td>
<td>11031</td>
<td>7</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>167186</strong></td>
<td><strong>100</strong></td>
</tr>
</tbody>
</table>

2.3 Data

Three geospatial precipitation datasets are used in this research namely TRMM, APHRODITE and WFD. While TRMM and WFD are Global datasets, APHRODITE is a dataset for Asia region. The data acquisition method for all three datasets is different. While TRMM depends on the TRMM Microwave Imager (TMI) and Precipitation radar mounted on it for estimating Precipitation, WFD is a reanalysis product which uses atmospheric models to estimate rainfall. Perhaps the simplest of the lot is APHRODITE which collects gauge data from all of Asia and interpolates it. The interpolation technique used is different.
from interpolating the precipitation recorded from the gauge. For each dataset, the daily spatial maps were obtained and then the study area was clipped using a mask of the same. The overlap period between the 3 datasets is 1998 -2001. Hence this is the analysis period used in this research. Each measurement technique has its own limitations. Hence a combination of these is expected to provide better estimates. The best estimate will be the measurement combined with ground truth information though it is seldom available. TRMM, WFD and APHRODITE are discussed in the following sections. The algorithms mentioned in this section are derived from (Huffman 2007) for TRMM, (Yatagai 2012) for APHRODITE and (G. Weedon 2011) for WFD.

2.3.1 Tropical Rainfall Runoff Mission (TRMM)

TRMM was the first satellite mission dedicated to measuring tropical Rainfall. It was launched in 1997 with an initial estimated functioning life of at least 12 years, i.e. till 2009. Though, TRMM, carrying a Microwave imager, infrared scanner and Precipitation radar onboard is still working and sending Precipitation data. The satellite mission is mainly designed for rainfall but its derived products also use other satellite and ground data (Huffman 2007).

The 3B42 merged rainfall product used in this study relies on a combination of microwave, infrared and radar information from TRMM, other microwave sensors (AMSU-B, SSM/I and AMSR-E), infrared data from geostationary satellites and ground data merged in the Global Precipitation Climatology Project (GPCP) (Winsemius 2009). This was done as a way of calibrating the data by using different data sources.

3B42 Algorithm (Huffman 2007)

TRMM gridded precipitation estimates are produced on a 3 hour temporal resolution and 0.25 degree by 0.25 degree spatial resolution in a global belt extending from 50 degrees south to 50 degrees north latitude.

The 3B42 estimates are produced in four stages; (1) the microwave precipitation estimates are calibrated and combined, (2) Infrared (IR) precipitation estimates are created using the calibrated microwave precipitation, (3) the microwave and IR estimates are combined, and (4) rescaling to monthly data is applied. Each precipitation field is best interpreted as the precipitation rate effective at the nominal observation time (Huffman 2007). These steps are discussed in further detail here.

1. High Quality (HQ) microwave estimates

All of the available passive microwave data are converted to precipitation estimates prior to use, then each data set is averaged to the 0.25° spatial grid over the time range ±90 minutes from the nominal observation time. All of these estimates are adjusted to a "best" estimate using probability matching of precipitation rate histograms assembled from coincident data. The algorithm takes the TRMM Combined Instrument (TCI; 2B31) as the calibrating data source.

2. Variable Rain Rate (VAR) Infrared (IR) estimates

3B42 uses two different IR data sets for creating the complete record of 3-hourly 0.25° gridded Tb’s. In the period 1 January 1998 to 6 February 2000, 10-km (subsampled), 3-hourly grids of IR data from the National Climatic Data Center (NCDC) GridSat-B1 were used in the 3B42 processing. For the period 1 January 1998 - 03 UTC 16 June 1998 geo-IR data were not available in the Indian Ocean sector, but high zenith angle data from adjacent geo-satellites is generally sufficient for fill-in. For the period from 7
February 2000 onwards, the Climate Prediction Center (CPC) Merged IR is used. For each hour the IR data are averaged to 0.25° resolution and combined into hourly files as ±30 minutes from the nominal time. The amount of imagery delivered to NCDC and CPC varies by satellite operator, but international agreements mandate that full coverage is provided for the 3-hourly synoptic times (00Z, 03Z, ..., 21Z). Histograms of time-space matched HQ precipitation rates and IR Tb’s, each represented on the same 3-hourly 0.25° grid, are accumulated for a month, and then used to create spatially varying calibration coefficients that convert IR Tb’s to precipitation rates. As in the HQ, the calibration interval for the IR is a calendar month, and the resulting adjustments are applied to data for the same calendar month. This choice is intended to keep the dependent and independent data sets for the calibrations as close as possible in time.

3. Combined HQ and VAR estimates

It is frequently quite challenging to combine different estimates of an intermittent field such as precipitation. The process of combining passive microwave estimates is relatively well-behaved because the sensors are quite similar and GPROF is used for most retrieval. This is not the case for the HQ and VAR fields.

Currently a simple approach is taken for combining the HQ and VAR estimates, namely the physically-based HQ estimates are taken "as is" where available, and the remaining grid boxes are filled with VAR estimates. This scheme provides the "best" local estimate, at the expense of a time series that is built from data sets displaying heterogeneous statistics.

4. Rescaling to monthly data

The final step in generating 3B42 is the indirect use of rain gauge data. It is highly advantageous to include rain gauge data in combination data sets. However, experience shows that on any time scale shorter than a month the gauge data are not reported with sufficient density nor reported with consistent observational intervals to warrant direct inclusion in a global algorithm that provides sub-monthly resolution. The authors solved this issue in the GPCP One-Degree Daily combination data set by scaling the short-period estimates to sum to a monthly estimate that includes monthly gauge data (Huffman et al. 2001). Here, we take a similar approach with the 3B42 estimates. All available 3-hourly HQ+VAR estimates are summed over a calendar month to create a monthly multi-satellite (MS) product. The MS and gauge are combined to create a post-real-time, monthly satellite-gauge combination (SG), which is a TRMM product in its own right (3B43). Then the field of SG/MS ratios is computed (with controls) and applied to scale each 3-hourly field in the month.

2.3.2 Watch Forcing Data (WFD)

WFD gridded precipitation estimates are produced on a 3 hour temporal resolution and 0.5 degree by 0.5 degree spatial. WFD is a global gridded meteorological forcing data. The variables included are

1. Wind speed at 10 m
2. Air temperature at 2 m
3. Surface pressure
4. Specific Humidity at 2 m
5. Downward long-wave radiation flux
6. Downward short-wave radiation flux
7. Rainfall rate
8. Snowfall rate

WFD does not provide estimates over Antarctica. It uses the same land-sea mask as used by the Climate Research Unit (CRU).

**WFD Algorithm (G. G. Weedon 2010)**

The Watch Forcing Data was derived from the ERA-40 reanalysis product of the European Centre for Medium Range Weather Forecasting (ECMWF). All the variables are available with the ERA-40 product. The precipitation rates estimated from ERA-40 were further corrected for number of wet days, bias, precipitation gauge catch etc.

1. Wet-day correction

   Wet day is defined as the day when precipitation is recorded in a particular grid cell. The ERA-40 product overestimated precipitation especially near the tropics (G. G. Weedon 2010). A wet day correction is applied to a grid cell where the number of wet days exceeds the number of wet days for that location in the CRU (Climate Research Unit) dataset. The precipitation rate for the day with the lowest precipitation is set to zero and the process is repeated till the number of wet days remaining is same as the number of wet days for CRU. It was observed that wet day deficit compared to CRU data was comparatively rare with some instances in small areas outside Greenland, northern Siberia in winter and north east Brazil. Hence no correction was made for the wet day deficit. The other variables remained untouched while making correction for the number of wet days.

2. Bias Correction

   Data from GPCC half degree version 4- full product (monthly totals) was used for bias correction in WFD. For some cases especially islands which are represented by CRU but not are not covered by GPCC, CRU Time Series (TS) 2.1 monthly totals were used. The bias correction was done by simply taking the ratio of WFD monthly total precipitation obtained after wet day correction and the monthly GPCC precipitation. This ratio is then multiplied by each sub daily value to obtain the bias corrected daily values. This method is simple but it also assumed the bias to be non-static. Once the bias is corrected, the precipitation amount is allocated to interim snowfall and rainfall rates in proportions associated with interpolated data.

3. Precipitation gauge catch correction

   The gauge catch correction take into account gauge design, wind induced errors and wetting losses. The correction factor used is called “catch ratio” which are provided as average monthly value for each half degree grid cell. The catch ratio considered here were taken from Adam and Lettenmaier (2003). These ratio represent average precipitation measured in local gauges divided by the catch corrected precipitation (Adam 2003). There were no separate corrections for rainfall and snowfall for earlier products, but they were added later on.

   Some other corrections include the correction for the boundary of Inter Tropical Convergence Zone (ITCZ). WFD creates extremely high precipitation values at those locations. This problem was solved by using a local threshold to clip extreme rates. These extreme rates were decided based on 1980-81 precipitation rates. Hence, it is not the best solution. This may result in errors at these locations.
2.3.3 Asian Precipitation - Highly-Resolved Observational Data Integration Towards Evaluation (APHRODITE)

APHRODITE is a daily gridded precipitation dataset from 1951 to 2007. It was created by collecting and analysing rain gauge observation data across Asia through the activities of the Asian Precipitation Highly Resolved Observational Data Integration Towards Evaluation of Water Resources (APHRODITE) project. The spatial resolution for APHRODITE is 0.25 degrees by 0.25 degrees. The product is based on data collected at 5,000–12,000 stations. This is 2.3–4.5 times the data made available through the Global Telecommunication System (GTS) network and is used for most daily gridded precipitation products. The data for this project came from APHRODITE’s own collection as well as from different sources like GTS-based data (Global summary of the day) and data precompiled by other organizations.

**APHRODITE Algorithm (Yatagai 2012)**

APHRODITE at first seems to be the simplest of the 3 datasets w.r.t. the algorithm since it is prepared as an interpolation of gauge data. But this is not the case. A lot of methods for checking, maintaining and improving the quality of were used in while making APHRODITE. Only after the quality control the interpolation is done to get the final data. The data generation involves the following steps:

1. Quality Control (QC)

The conversion of collected data to a common data format was the biggest task, but it also allowed the errors in data handling to be checked manually. Only after 2007 is the QC completely automated. In this step, the data is checked for initial errors such as shifted columns, invalid dates, incorrect metadata etc. The second task is to check for locations. Some gauges have incorrect or no location. Hence it is necessary to find out the actual location of the gauge. The locations were compared with the national boundaries and the elevation was compared with the global 30 Arc-second elevation dataset (GTOPO30). The third and most important step is to make a black list and remove spurious stations’ data. This is done by conducting tests on single station results as well as multiple station results. After this step, wrong or erroneous data is discarded and the data left is usable.

2. Adopting Climatology

For generating spatial data, the easiest way is to adopt a climatology already defined. For the data used in this research, the climatology adopted is WorldClim (Hijimans 2005) because it provides data throughout the domain. Other methods like the mountain mapper (MM) do not work well for areas with no rain gauge data. Parameter-elevation Regressions on Independent Slopes Model (PRISM) (Daly 1994) was used for generating APHRODITE dataset. After 2007, the team replaced it with WorldClim. This leads to the final product generation which involves seven steps:

I. Daily (except for GTS oriented) data that have undergone QC are summed as monthly values.

II. The monthly data, including those obtained in step 1, are gathered and the averaged value is calculated if the station has recorded data for more than 5 yr.

III. The world climatology is prepared at 0.05° resolution

IV. The ratio of the data obtained in step 2 (station climatology) to the data obtained in step 3 is taken for each month.
V. The ratio in step 4 is interpolated using Sheremap at a resolution of 0.05°.
VI. The interpolated values in step 5 are multiplied with the world climatology prepared in step 3.
VII. The first six components of the fast Fourier transform of the values obtained in step 6 are taken and the daily climatology is obtained.

It should also be known that while preparing APHRODITE, weights are given based on topography. The windward side on a slope gets a higher slope compared to the leeward side.
Chapter 3
Methods

3.1 Error Structure

Understanding the error structure is very important for using Triple collocation. Triple collocation is mostly used in data related to ocean sciences like wave height measurements (PETER A. E. M. JANSSEN 2006), wind velocity (Stoffelen 1998), soil moisture (K. Scipal 2008) etc. It has been used only once on precipitation data (R. A. Roebeling 2012) but it was not applied on a daily timescale. The application of the technique on a daily timescale was a challenge. To overcome this challenge, it becomes even more important to understand the mathematics behind the technique. The discussion below includes a description of the triple collocation error model, derivation of calibration coefficients and the measurement of error variances, the assumptions needed in triple collocation methods and its numerical implementation.

3.1.1 Definitions

$X_i, i=1,2,3,\ldots,N$ be the N measurements of a quantity $X$.

The mean also called the first moment of $X$ is denoted by $M_X$. $<>$ denotes statistical average or expected value of a variable.

$$M_X = <X> = \frac{1}{N}\sum_{i=1}^{N} X_i$$  \hspace{1cm} (3.1)

Now if there is another quantity $Y$ with $N$ measurements then the mixed second moment is denoted by

$$M_{XY} = <XY> = \frac{1}{N}\sum_{i=1}^{N} X_iY_i$$  \hspace{1cm} (3.2)

If $X=Y$, then one ends up with the ordinary second moment $M_{XX}$.

The covariance of $X$ & $Y$, $C_{XY}$ is defined as

$$C_{XY} = M_{XY} - M_XM_Y$$  \hspace{1cm} (3.3)

Again when $X=Y$, Eq 4.3 yields the variance of $X$,

$$C_{XX} = \sigma_X^2$$  \hspace{1cm} (3.4)

For triple collocation, we assume system $X$ as reference system which has a calibration coefficient of 1. We take 3 datasets, with $X$ as the reference system where each data value can be represented as:

$$x = t + e_x$$  \hspace{1cm} (3.5)

$$y = b_y + a_y(t + e_y)$$  \hspace{1cm} (3.6)

$$z = b_z + a_z(t + e_z)$$  \hspace{1cm} (3.7)
3.2 Errors and Calibration Coefficients

The first statistical moment of the data is

\[ M_x = \langle t \rangle + \langle e_x \rangle \]  
\[ M_y = b_y + a_y \langle t \rangle + \langle e_y \rangle \]  
\[ M_z = b_z + a_z \langle t \rangle + \langle e_z \rangle \]  

(3.8)
(3.9)
(3.10)

Since the errors are random, the first average of errors \( \langle e_x \rangle \) is zero which means \( M_x = \langle t \rangle \). Eliminating \( \langle t \rangle \) from the others result in:

\[ b_y = M_y - a_y M_x \]  
\[ b_z = M_z - a_z M_x \]  

(3.11)
(3.12)

The ordinary second-order moments are

\[ M_{xx} = \langle t^2 \rangle + \sigma^2_x \]  
\[ M_{yy} = a_y^2 \langle t^2 \rangle + b_y^2 + a_y^2 \sigma^2_y + 2a_y b_y M_x \]  
\[ M_{zz} = a_z^2 \langle t^2 \rangle + b_z^2 + a_z^2 \sigma^2_z + 2a_z b_z M_x \]  

(3.13)
(3.14)
(3.15)

Using the values of \( b_y \) and \( b_z \) obtained earlier and introducing co-variances yields:

\[ C_{xx} = \langle t^2 \rangle - M_x^2 + \sigma^2_x \]  
\[ C_{yy} = a_y^2 \langle t^2 \rangle - M_x^2 + \sigma^2_y \]  
\[ C_{zz} = a_z^2 \langle t^2 \rangle - M_x^2 + \sigma^2_z \]  

(3.16)

The mixed second order moments are

\[ M_{xy} = a_y \langle t^2 \rangle + b_y M_x + a_y r^2 \]  
\[ M_{yz} = a_y a_z \langle t^2 \rangle + a_y b_z M_x + a_z b_y M_x + b_y b_z M_x \]  
\[ M_{zx} = a_y \langle t^2 \rangle + b_z M_x \]  

Where \( r^2 = \langle e_y e_z \rangle \) is the representation error when errors are uncorrelated. In this research though, the errors are assumed uncorrelated which simplifies things.

Again eliminating \( b_y \) and \( b_z \) and introducing co-variances results in

\[ C_{xy} = a_y (\langle t^2 \rangle - M_x^2 + r^2) \]  
\[ C_{yz} = a_y a_z (\langle t^2 \rangle - M_x^2) \]  
\[ C_{xy} = a_y (\langle t^2 \rangle - M_x^2) \]  

Solving these for \( a_y \) and \( a_z \) yields
\[ a_y = \frac{c_{yz}}{c_{zx}} \]  
\[ a_y = \frac{c_{yz}}{c_{zy} - a_y r^2} \]  

Substituting \( a_y \) and \( a_z \) and solving for error variances yields

\[ \sigma_x^2 = c_{xx} - \frac{c_{zx} (c_{zy} - a_y r^2)}{c_{yz}} \]  
\[ \sigma_y^2 = c_{yy} - \frac{c_{zy} (c_{xy} - a_y r^2)}{c_{yx}} \]  
\[ \sigma_z^2 = c_{zz} - \frac{c_{zx} c_{yz}}{c_{xy} - a_y r^2} \]

This is the generalized method for calculating error variances and calibration coefficients for datasets. The actual implementation is a little bit different based on the datasets and assumptions involved. This is explained in the next section.

### 3.3 Triple Collocation

Triple collocation is a statistical technique used to estimate random errors in 3 different datasets without knowing the true value of the measured parameter (Discussed in details in 2.1.1). The choice of three aforementioned datasets for use in this study is governed by the two assumptions for triple collocation:

1. The data is bias corrected
2. The errors are uncorrelated

The first assumption is true because all three datasets are bias corrected before making them available for use to the public.

The second assumption is assumed to be true because of the way the datasets are prepared. While TRMM measures the precipitation using data from TMI (TRMM Microwave Imager) and PR (Precipitation Radar), the EU-WFD uses weather parameters like temperature, wind-velocity etc. to estimate Precipitation and APHRODITE is just the smoothened interpolation of global rain gauge data. Hence these datasets are suitable to be used for triple collocation. TRMM uses GPCC station data for monthly bias correction which may have some overlaps with some APHRODITE stations. But for the purpose of this research, the errors are assumed to be uncorrelated due to few stations in Myanmar. In total, there are 31 precipitation gauges used by GPCC in the catchment under study- Figure 4.
While collocating the three datasets, a recursive algorithm is used to vary the estimated errors in data. A convergence criterion is defined and the iteration is stopped when the data converges.

3.3.1 Theory
Let us take 3 bias corrected datasets X, Y and Z, measuring the same parameter with uncorrelated random errors. Let each data measurement be represented as:

$$W_i = \alpha + \beta T_i + e_i \quad (3.22)$$

Where

- $\alpha = \text{Fixed bias in the measurement system}$
- $\beta = \text{Calibration constant of the measurement system}$
- $T_i = \text{Truth or True value}$
- $e_i = \text{Unbiased random error or errors with a zero mean}$

Also the errors are assumed to be independent of the common signal $t$.

Note that except for $W_i$ all the variables are unknown.

Now for the 3 systems, eq. 3.22 can be written as

$$X_i = \beta x T_i + e_{x_i} \quad (3.23)$$

$$Y_i = \beta y T_i + e_{y_i}$$

$$Z_i = \beta z T_i + e_{z_i}$$
α = 0, since the data is bias corrected

In triple collocation, the calibration constant β is found by iteration. For the time being, eq. 3.23 are divided by β_x, β_y, β_z respectively.

\[ \Rightarrow X_i' = T_i + e'_{x_i} \]
\[ Y_i' = T_i + e'_{y_i} \]
\[ Z_i' = T_i + e'_{z_i} \]

Where:

\[ X' = X/β_x \]

To get rid of \( T_i \), difference between each pair of equations is calculated:

\[ X_i - Y_i = e_{x_i} - e_{y_i} \]  \hspace{1cm} (3.25)
\[ X_i - Z_i = e_{x_i} - e_{z_i} \]  \hspace{1cm} (3.26)
\[ Y_i - Z_i = e_{y_i} - e_{z_i} \]  \hspace{1cm} (3.27)

Eq. 3.25, 3.26 and 3.27 are multiplied with each other to get the expected values of the squares of the differences (Dropping subscript ‘i’ for clarity):

\[ <(X - Y)(X - Z)> = <(e_x)^2> - <e_x e_y> - <e_x e_z> \]  \hspace{1cm} (3.28)
\[ <(Y - X)(Y - Z)> = <(e_y)^2> - <e_y e_x> - <e_y e_z> \]  \hspace{1cm} (3.29)
\[ <(Z - Y)(Z - X)> = <(e_z)^2> - <e_z e_x> - <e_z e_y> \]  \hspace{1cm} (3.30)

Now all quantities on the LHS are known from measurements. Also, the errors are uncorrelated, i.e. all error co-variances/correlations:

\[ <e_x e_y> = <e_x e_z> = <e_y e_z> = 0 \]  \hspace{1cm} (3.31)

So, it is possible to find the error variances:

\[ <(e_x)^2> = <(X - Y)(X - Z)> \]  \hspace{1cm} (3.32)
\[ <(e_y)^2> = <(Y - Z)(Y - X)> \]  \hspace{1cm} (3.33)
\[ <(e_z)^2> = <(Z - Y)(Z - X)> \]  \hspace{1cm} (3.34)

Now if \( β \) is known, the total error is known. In triple collocation, one of the systems is assumed to be calibrated (i.e. \( β_x = 1 \)). The other two systems are calibrated w.r.t. this system. It does not make any difference which system is assumed to be calibrated as the \( β \) values will change proportionally. Also, neutral regression is required to find \( β_y \) and \( β_z \) (Conventional regression assumes one of the system is error free). This yields:

\[ \beta_y = \frac{-B + (\sqrt{B^2 - 4AC})}{2A} \]  \hspace{1cm} (3.35)

Where
A = γ <X,Y>; \hspace{1cm} B = <X^2> - γ <Y^2>; \hspace{1cm} C = <X,Y>; \hspace{1cm} γ = <e_x^2>/<e_y^2>

β_z can be calculated with the same equation by replacing y with z.

3.3.2 Procedure (PETER A. E. M. JANSSEN 2006)
1. Find three datasets with independent errors
2. Remove the bias from each data set independently
3. Compute variances and co-variances of the data sets
4. Assume β_x = β_y = β_z = 1
5. Solve for squared errors
6. Solve for β_y and β_z using neutral regression
7. Adjust the data for each data set using calibration constants calculated in step 7
8. Repeat starting from step 3 until convergence.
Chapter 4
Results and Discussions

4.1 Seasonal Errors in Precipitation
Average daily error is calculated for three periods in Myanmar namely Pre-monsoon (January to first half of May), Monsoons (May to October) and Post monsoon (late October to December). The seasonal error presented is the mean of daily errors for that duration. The results for all 4 years (1998-2001) are presented in this section and some observed inter-annual, intra-annual trends are discussed. The spatial distribution of average annual precipitation for the 3 datasets is presented for 1998 and 1999 to give an estimated of the precipitation magnitudes Figure 5 and Figure 6. It is evident that on an average, there is more precipitation in the mountains in the north and the delta in the south. Also, the west-central dry zone is clearly visible in these maps. There is not much difference in the maximum average values to derive any conclusions from these maps.

Another noticeable feature is the difference in average precipitation in 1998 and 1999. The difference is remarkable (40-50%) for the high precipitation areas. This is because 1998 was a dry year because of the El-Nino.

* All map colour-bars are in mm/day. The map axis are longitudes (x-axis) and latitudes (y-axis)
4.1.1 Spatial trends

The seasonal error patterns for 1998 are very peculiar. The error is much more evenly distributed in space for APHRODITE compared to TRMM and WFD. This is different from the period 1999-2001 when the difference in spatial error pattern is not that obvious. One possible reason for this can be the occurrence of the great El-Nino in 1997-98. TRMM and WFD being based on observation and reanalysis respectively should be able to account for low precipitation in an El-Nino year. On the other hand APHRODITE is based on interpolation of in-situ measurements. It is possible that the heavy rainfall areas in the northeastern part of India affected the precipitation values in Myanmar and hence the error.

Figure 6: Average Spatial Precipitation Maps 1999

The error trends and the average precipitation trends are similar. In general, the higher the precipitation at a point, the higher is the error. The difference between the error distributions of the three datasets is discussed with focus on recurring trends as well as anomalies. The reasons for trends are discussed and explained wherever possible.

Figure 7: Error Maps for Pre-Monsoon Season 1998
Figure 8: Error Maps for Monsoon Season 1998

Figure 9: Error Maps for Post-Monsoon Season 1998

Figure 10: Error Maps for Pre-Monsoon Season 1999
There are also some peculiar hotspots in average spatial errors. As can be seen from the maps presented, the west central part of Myanmar shows high error magnitudes. This is the dry part of Myanmar with average annual precip. in the order of 600-700 mm. From a hydrological point of view, a good estimate in dry areas is very important due to highly non-linear relation between rainfall and runoff. A small change in precipitation can result in significant change in runoff (Lu Li 2013). Both TRMM and WFD are not good in estimating precipitation in the semi-arid regions. Another reason for this high error is the presence of the ITCZ which passes through this dry part of Myanmar and also crosses the boundary between the easter-central plateau and the mountains up north. It is no coincidence that this whole region shows a high error. The precipitation in ITCZ is highly localized and convective in nature. The TRMM and WFD algorithms are not well equipped to handle this convective precipitation while the APHRODITE does not capture the localized nature of precipitation.

The second hotspot includes the southwestern part. This is the place where the southwest monsoon enters the country and bifurcates into two branches due to the hills along the west coast. Another interesting aspect is the occasional recurring hotspot for WFD in the southern delta region. The delta region has higher errors in some maps for TRMM and APHRODITE as well, but it is more frequent in
WFD. Since the WFD uses the CRU land cover mask, one of the possible reasons for this may be the inability of WFD to capture the effect of sea storms which are very common in the delta region.

Other than these, some general observations can be made. The APHRODITE errors are much more evenly distributed and hence lesser in magnitude. On the other hand, the TRMM and WFD show localized error hotspots with higher magnitudes. The average seasonal error in mountains is highest for APHRODITE and lowest for TRMM.

The errors in daily precipitation show the expected trend. The magnitude and extent increases in the monsoon, and starts decreasing afterwards with minimum values in the dry summers. This can be because of the ability to correctly predict no rain by these datasets. During summer, the precipitation is rare and hence lesser errors. The same trend can be observed on the monthly scale as well.

4.1.2 Temporal Trends

The maximum, minimum and mean error values in space for each day were calculated along with the standard deviation. The trends presented are for year 2000 but other years also showed similar trends.

The average and maximum errors are highest for WFD in general, though TRMM shows few very high peaks. At present, it is difficult to point out the reason for these peaks but it can also be some anomalous result. Looking at the average and maximum precipitation and standard deviations, APHRODITE again seems to be the best dataset for the catchment. The minimum error graph shows higher values for APHRODITE though the magnitude is insignificant. As discussed in the previous section, APHRODITE seems to have much more evenly distributed errors with low peaks.

Another important aspect is the presence of some peaks in WFD during the start and end of monsoon season. This trend was further verified by checking error trends at individual grid cells. In Myanmar, there are storm events before and after monsoon causing localized precipitation. These are similar to the so called mango showers (Pre monsoon showers/Local Storms) in India. The WFD algorithm seems unable to make a good estimate of these storm events. TRMM also shows similar spikes but the magnitude is lesser than WFD. But it must be noted that the error during the monsoons for WFD is lesser compared to TRMM.

![Figure 13: Spatial Mean Error for 2000](image-url)
Figure 14: Spatial Minimum Error for 2000

Figure 15: Spatial Maximum Error for 2000

Figure 16: Spatial Standard Deviation for error of 2000
4.1.3 Comparison with station data

Precipitation data for two major cities Bago and Yangon was made available by the Department of Meteorology and Hydrology, Myanmar. An overview of comparison of station data and the three datasets used in this research is provided in Table 3. The precipitation values for Yangon and Bago were extracted from the all 3 datasets and the difference between daily precipitation of these datasets and gauge measurements was plotted—Figure 17-18. All 3 datasets seem to underestimate the precipitation. TRMM seems to have the closest average values to the gauge data though it also has a high error peaks in monsoon. The Pearson Rank Correlation values are consistent for APHRODITE. For TRMM and WFD, the correlation values are not consistent to derive any conclusions. With the limited data, APHRODITE appears to have the best correlation with the station data, which is not surprising as APHRODITE itself is made derived from interpolated gauge data.

Table 3: Comparison Statistics with Station Data

<table>
<thead>
<tr>
<th></th>
<th>APHRODITE</th>
<th>TRMM</th>
<th>WFD</th>
<th>Gauge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bago</td>
<td>6.8336</td>
<td>6.5531</td>
<td>5.6502</td>
<td>6.5192</td>
</tr>
<tr>
<td>Yangon</td>
<td>4.8708</td>
<td>5.4718</td>
<td>4.9444</td>
<td>6.6923</td>
</tr>
<tr>
<td><strong>Avg. Ann. Prec. 1999</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yangon</td>
<td>7.5775</td>
<td>8.8291</td>
<td>9.8175</td>
<td>9.6731</td>
</tr>
<tr>
<td><strong>Pearson RC 1998</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bago</td>
<td>0.2570</td>
<td>0.3384</td>
<td>0.0636</td>
<td>-</td>
</tr>
<tr>
<td>Yangon</td>
<td>0.0629</td>
<td>-0.0591</td>
<td>0.147</td>
<td>-</td>
</tr>
<tr>
<td><strong>Pearson RC 1999</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bago</td>
<td>0.2454</td>
<td>0.066</td>
<td>-0.005</td>
<td>-</td>
</tr>
<tr>
<td>Yangon</td>
<td>0.2849</td>
<td>0.2563</td>
<td>0.0264</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure 17: Comparison with Bago station - 1998

Figure 18: Comparison with Yangon station - 1998
These comparison plots assume the gauge data is the best representative for that area in general. It can be seen that the higher peaks are common for all three datasets. The interesting part is when the
difference is different for the 3 datasets. In these cases, APHRODITE is by far one with the lowest values. The observations from the triple collocation analysis are also validated with these comparison plots. It can be clearly seen that WFD has some high peaks in the beginning and end of monsoon season. This is because it’s unable to predict the localized showers which are common in that period in a monsoon climate. Another important aspect is that TRMM errors are more or less equally distributed on positive and negative sides in comparison to WFD whose negative errors outweighs positive ones. WFD in general underestimates precipitation in Tropical Climates which is also evident in this comparison.

Another important thing is the odd spikes in every plot for each of the dataset. The positive spikes may be because of some missing data in for the gauge or spurious measurement. The negative spikes are for very high precipitations measured by the gauge. This can again be due to spurious data or some strong localized precipitation. Since the geo-spatial datasets interpolate precipitation over a larger area, the precipitation values are lower since high localized precipitation is compensated by lower precipitation values at other locations in the same grid cell. The cell size is approximately 25Kms which is quite big to assume uniform precipitation pattern.

The assumption that the errors are uncorrelated was also validated on the errors for Bago and Yangon. The scatter plots for the daily errors for Yangon and Bago can be seen in Figure 21 and Figure 22. From the scatter plots, it can be concluded that there is not very strong correlation between the errors of the 3 datasets at Yangon and Bago. Even when the scatter plots show some correlation, the values are not similar. For 1999, the scatter plots for APHRODITE and WFD at Bago show some correlation, but the higher values of WFD are correlated with lower values of APHRODITE. None of the plots show a dense scatter plot near 45 degrees which will mean a positive correlation.

![Figure 21: Scatter plots for errors at Bago(top) and Yangon(bottom) 1998](image)
Figure 22: Scatter plots for errors at Bago (top) and Yangon (bottom) 1999
Chapter 5
Conclusions and Recommendations

The research provided a way to apply triple collocation on daily or sub-daily time scales for precipitation data. This has great implications for the hydrological modeling in data scarce regions around the world. Using the approach presented in this work, it is possible to make a comparative error study between different geo-spatial datasets with uncorrelated errors. Using the error maps thus produced, preliminary conclusions can be drawn about the quality of each dataset relative to the other two in space and time. For the case of Myanmar, the study suggests that there is a lot of spatial and temporal variation in the errors for TRMM, WFD and APHRODITE. Also, different time scales provide insight into the pros and cons of each data set. The errors can be region specific or time specific. To sum it up, the following conclusions can be drawn from this research:

1. What are the spatial error characteristics of the three datasets mentioned above for Myanmar?
   - The spatial errors are more distributed for APHRODITE but have higher peaks for WFD and TRMM.
   - All three datasets are poor in estimating precipitation in the west-central dry zone and the north-eastern Plateau. These datasets provide poor estimates for precipitation at the boundary of ITCZ.
   - The precipitation errors at the entry point of monsoon in southwest Myanmar are also high for all 3 datasets while WFD also shows higher errors in the Delta region.
   - The precipitation in the mountains is best estimated by WFD.

2. What are the temporal error characteristics of the three datasets mentioned above for Myanmar?
   - The error magnitude in time is least for APHRODITE. WFD and TRMM show higher maximum and average error values in time.
   - The estimation for pre and post monsoon showers for TRMM and WFD needs improvement.

3. Is the variability in the error characteristics same or different for these datasets over different temporal scales, i.e. daily, monthly and annual/seasonal?
   - The error is maximum during monsoons and minimum during the dry summer.
   - There are differences in the annual error patterns based on the climatic events during that year (e.g. El-Nino 1998).
   - Different temporal scales show different error trends.

4. How can the knowledge of errors in these datasets help in selecting the best dataset for input forcing in a Hydrologic Model, e.g. best dataset for a flood prediction model can be different than one for drought prediction?
   - Based on this research, it is difficult to say which dataset is best for flood or drought prediction individually.
• APHRODITE dataset seems to be the best option in general as the average precipitation values are closer to the gauge data and also the errors are much more evenly distributed.
• The best option would be to combine the 3 datasets based on their spatial and temporal error variability.

Based on the results obtained in this research, the following recommendations are made:

1. For hydrological modeling, APHRODITE should be preferred over TRMM and WFD as it shows the least varied error trends.
2. For the dry zone, a better network of gauge stations is required to supplement these datasets as none of the datasets have a good estimate of precipitation in the dry zone.
3. TRMM and WFD need to be improved to better estimate localized precipitation events and precipitation near the ITCZ boundary.
4. The best option for a hydrological model is the combination of all three datasets. Based on the best dataset for each geographical location and time step, the preparation of a weighted combined dataset should be looked into.
5. More research is needed for the monsoon in Myanmar as literature on this phenomenon is scarce. This is necessary for better understanding and estimating the precipitation in Myanmar.
6. Representation errors should be included and the results be compared to this analysis.
Bibliography


Yatagai, Akiyo, Kenji Kamiguchi, Osamu Arakawa, Atsushi Hamada, Natsuko Yasutomi, and Akio Kitoh. “APHRODITE: Constructing a long-term daily gridded precipitation dataset for Asia based on a