Estimation of Rain Height from Rain Rate using Regression-based Statistical Model: Application to SeaWinds on ADEOS-II

by

Bhaskar Natarajakumar

B.E., Electronics and Communication Engineering,

College of Engineering, Guindy – Anna University

Chennai, India, 2001

Submitted to the Department of Electrical Engineering and Computer Science and the Faculty of the Graduate School of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Science.

Thesis Committee

_____

Dr. Glenn Prescott (Chair)

_____

Dr. Richard K. Moore

_____

Dr. David Braaten

Date of thesis defense: 28th January 2004

This thesis is dedicated to my grand father

S. Thirunavukkarasu

Your love and support are still with me, even if you are not.

Abstract

The SeaWinds scatterometer that flew on the ADEOS II satellite and continues on the QuikSCAT satellite determines ocean-surface wind vector by measuring radar backscatter from the ocean. These radar cross-section measurements are hindered by the presence of rain over seas. The deleterious effects of rain are attenuation of the surface signal and the addition of backscatter from rain to the received signal. AMSR, the scanning radiometer on ADEOS II, provided rain-rate estimates which can be used to determine attenuation rate and rain-backscatter intensity. If one knows the rain-cell height, one can estimate the total attenuation and rain backscatter for each beam and correct for the effects of rain. Unfortunately, no direct measurement of rain height was available on ADEOS II.

The objective of this thesis is to determine a method to estimate rain-cell height for different seasons and oceanic regions. Rain height from the Tropical Rainfall Measurement Mission (TRMM) precipitation radar (PR) was used in these studies. An initial investigation on using climatological rain-height estimates showed that they were crude and constrained only to tropical regions. Also, an attempt to relate rain height to sea-surface temperature failed.

TRMM rain rate and rain height showed good correlation over various regions, seasons and rain types (stratiform and convective). Due to the scatter of the measured data, numerous regression schemes were attempted to relate rain rate and rain height. Most of the schemes work well at lower rain rates, but they behave poorly at higher rain rates. A *Log-Linear Combined* regression technique using a linear-log regression for low rain rates and a linear regression for high rates provides a consistent relation over all rain rates. The seasonal effects on slopes and intercepts of the regression lines were also analyzed. Finally, the thesis proposes a *regression-based statistical model* to predict slopes and intercepts for the regressions lines, which can be used to estimate rain height given a rain rate. The estimated rain-cell height is necessary to correct for the rain effects in SeaWinds scatterometer measurements, since it is required to determine the total backscatter from the rain volume.

Acknowledgements

I would like to express my sincere gratitude to my research advisor Professor Emeritus Dr. Richard K. Moore for his support and encouragement throughout this research work. He has devoted so much time and effort to teaching me both in this research and in writing it that my labors will never be able to match. I am highly indebted to him for funding me throughout my graduate studies. I have learnt a lot of valuable lessons from his advice, which will be of great help to shape my career. Thank you sir!

I would like to extend my sincere thanks to Dr. Glenn Prescott and Dr. David Braaten for serving in my committee. I acknowledge the support and help of my team member Vivek. J. Kurisunkal and research associate Dr. Victor Andrade. Special thanks to my parents, T. Natarajakumar and Sivagami, my grandfather and my brother for their endless love and support. To them, I owe all the good things in my life. I feel blessed to have had great teachers in high school, who taught me simple lessons of discipline and hard work to be successful in life. I thank all my friends in Lawrence for having made my stay a truly enjoyable and memorable one. I take this opportunity to thank two of my good friends, Girish and Nandish, for their help and moral support during my studies at the University of Kansas. Last but never the least I thank the almighty for all things that have come my way in life.

Table of Contents

List of Figures

List of Tables

1 Introduction

1.1 Need for Wind Data

Wind stress is the signal largest source of momentum to the upper ocean, and winds drive oceanic motions on scales ranging from surface waves to basin-wave current systems. Winds over the oceans regulate the crucial coupling between atmosphere and ocean that establishes and maintains global and regional climate.

Knowledge of wind velocity over the ocean is of critical importance for understanding and predicting many oceanographic, meteorological, and climate phenomena. Measurements of the surface wind velocity can be assimilated into regional and global numerical weather prediction systems, thereby extending and improving our ability to predict future weather patterns on many scales [1]. Other applications of wind vector measurements include the study of unusual climatological phenomena such as El Niño and hurricane monitoring [2].

Most ship-borne ocean surface wind vector measurements are geographically and phenomenologically biased. Ship reports of wind velocity are also notoriously

inaccurate owing to untrained observers, poor instrumentation, badly placed anemometers, contamination owing to ship motion, and data transcription and transmission errors [3]. Only a satellite-borne instrument can acquire wind data with global coverage, high spatial resolution, and frequent sampling [4]. Both satellite altimeters and multi-channel microwave radiometers can be used to estimate all-weather wind speed; however, these instruments do not measure wind direction. Wind direction is a crucial input to calculating air-sea momentum fluxes needed to understand atmospheric dynamics.

The wind vector on the ocean-surface can be determined from the measurement of radar backscatter of the ocean by utilizing a space-borne scatterometer [5]. Satellite-borne scatterometers are the only remote sensing systems able to provide accurate, frequent, high-resolution measurements of ocean surface wind speed and direction in both clear-sky and cloudy conditions [4]. The next section gives a brief description on the principle of scatterometry and how the ocean backscatter is used to measure wind vector.

1.2 Scatterometry for Wind vector measurements

The scatterometer is a microwave radar sensor that measures the scattering or reflective property of surfaces and/or volumes. Early radar observations of ships and aircraft over oceans were corrupted by "sea clutter" (noise), the backscatter for the surface. Radar response was first related to wind in the 1960's [6-9], based on measurements from aircraft, ships, and shore stations. The first space-borne scatterometer flew as part of the Skylab missions in 1973 and 1974, demonstrating that space-borne scatterometer were indeed feasible [10].

Space-borne scatterometer transmit microwave pulses to the ocean surface and measure the backscattered power received at the instrument. Since the atmospheric motions themselves do not substantially affect the radiation emitted and received by the radar, the scatterometers use an indirect technique to measure the wind vector over the ocean. Wind stress over the ocean generates waves, leading to a rough sea surface. Changes in the wind velocity cause changes in the surface roughness, which in turn modify the radar cross-section and hence the magnitude of backscattered power. The empirical relation that describes the correlation between backscattered power and ocean near-surface wind conditions are given in [11].

Scatterometer measure the backscattered power, allowing estimation of the normalized radar cross-section (s^o) of the sea surface. To extract the wind velocity from these measurements, the relationship between radar cross section and the near-surface winds must be known. The radar cross section s^o is calculated using the basic radar equation:

$$s^o = \frac{(4\pi)^3 R^4 L P_s}{P_t G^2 \lambda^2 A} \qquad \textbf{(1.1)}$$

where R is the slant range to the surface, P_t is the transmitted power, P_s is the received backscattered power, L represents known system losses, G is antenna gain, A is the effective area of illumination, and λ is the wavelength of the transmitted radiation. From each illumination location on the earth, the total received power is the sum of the backscattered power P_s and a contribution P_n resulting from instrument noise and the natural emissivity (at that frequency) from the earth-atmosphere system, Unless the signal-to-noise ratio (SNR) is very large, the noise power P_n must be estimated and subtracted from the total received power $P_{(s+n)}$ to determine P_s accurately; the radar equation can then be used to calculate s^o.

The relationship between s^o and near-surface wind velocity is known as the *geophysical model function.* A scatterometer model function describes the variation of s^o with wind speed, wind direction, and angle of incidence [12]. The geophysical model function can be written most generally as,

$$s^o = f(|U|, f, ...; q, f, pol) \qquad \textbf{(1.2)}$$

where $|U|$ is wind speed, f is the azimuth angle between the incident radiation and the wind vector, q is the incidence angle measured in the vertical plane, f and pol are the frequency and polarization respectively of the incident radiation.

Estimation of wind velocity from a s^o measurement involves inversion of the model function given by (1.2). The model function defines the locus of wind speed and directions corresponding to each s^o measurement. In the case of noise-free measurements and perfect knowledge of the model functions, intersections of these loci define possible wind velocity solutions consistent with all s^o measurements. If only two collocated s^o measurements from different radar geometries are available, up to four intersections of the loci exist, owing to the near-symmetric oscillatory nature of the model function with respect to direction, as shown in Figure 1-1 [4].

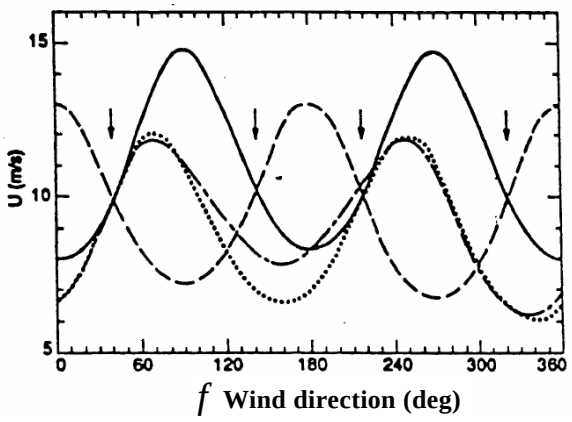


Figure 1-1 Loci of possible vector winds associated with co-located noise-free s^o measurements obtained from antennas pointed at various azimuth angles. Heavy solid line: antenna angle at 90° (v-pol); dashed line; angle at 90° (v-pol); light solid line: angle of 25° (v-pol); dotted line: angle of 25° (h-pol). Arrows indicate solutions obtained using only the antenna at 0° and 90° [4].

Each of these intersections represents a possible wind velocity solution, and is denoted in the scatterometer literature as an *ambiguity.* Additional measurements from different geometries reduce the number of intersections and allow estimation of a unique wind velocity in the idealized case. Hence, at least two, and preferably more, measurements of normalized radar cross-section of the same location, from different viewing angles must be obtained in the space-borne scatterometer. These multiple, near-simultaneous s^o measurements combined with the geophysical model function, will allow estimating the near-surface wind speed and direction over the ice-free oceans.

1.3 Spaceborne Scatterometer

1.3.1 Brief History

The first satellite scatterometer program after the Skylab experiment was the NASA (National Aeronautics and Space Administration) SEASAT scatterometer (SASS), a fan-beam instrument, operating at 14.6 GHz with two beams on each side of the flight track. The European Space Agency's ERS-1/2 Advanced Microwave Instrument (AMI) operating at 5.3 GHz, includes scatterometer modes with 3 fan-beam antennas and acquires s^o measurements in a single, 500 km wide swath (one side of flight track only) [13]. The second scatterometer launched flown by NASA was the

NSCAT (NASA scatterometer), which used six fan-beam antennas, three on each side to obtain multiple azimuth looks. The design was an improvement over SASS because of the extra beam and over AMI because looking on both sides increased the swath width, allowing it to cover 90% of the Earth every two days. Figure 1-2 shows the history and future of scatterometry missions.

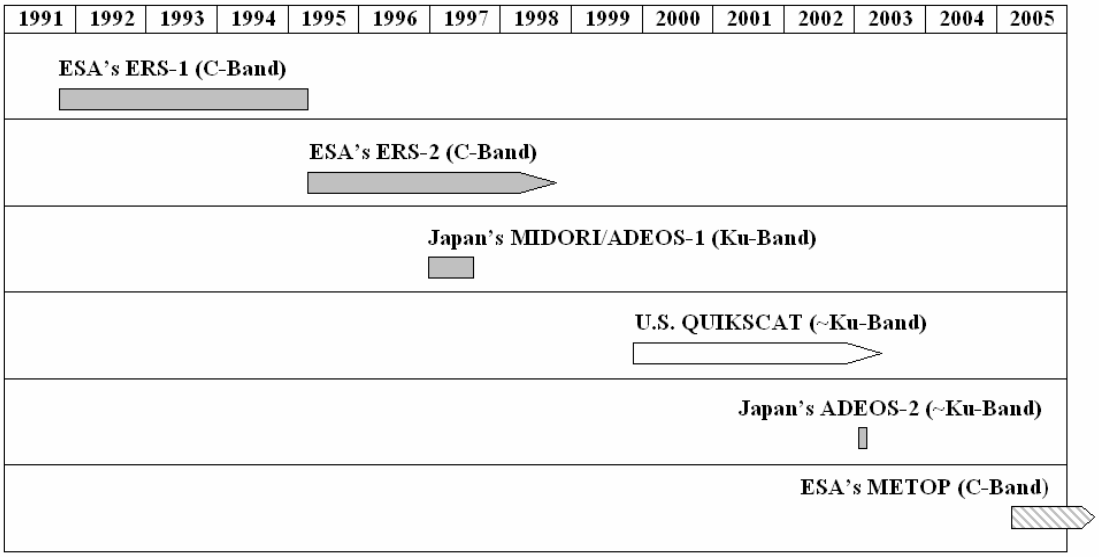


Figure 1-2 Scatterometer Spacecraft – Timeline

1.3.2 SeaWinds Mission

The third and most recent NASA scatterometer design is the Ku-band SeaWinds, operating on both the QuikSCAT satellite and the Japan's ADEOS II (also known as Midori II) satellite [14]. ADEOS II was originally the only planned SeaWinds mission. The unexpected short life of NSCAT and delay in the launch of ADEOS II prompted the development of QuikSCAT as a *quick recovery mission*. The design of the SeaWinds instrument is fundamentally different from fan-beam satellite

scatterometers. Rather than use multiple fan-beam antennas to create the measurement swath, it has only one antenna, a pencil-beam type, which it mechanically rotates about nadir. The combination of rotation and inclination angle to the antenna gives it the advantage of a significant larger swath (see Figure 1-3) and no gap in the nadir region. The rotating antenna has two feeds, creating two beams, an inner and outer, which provide four different azimuth-looks for ground locations [15, 16].

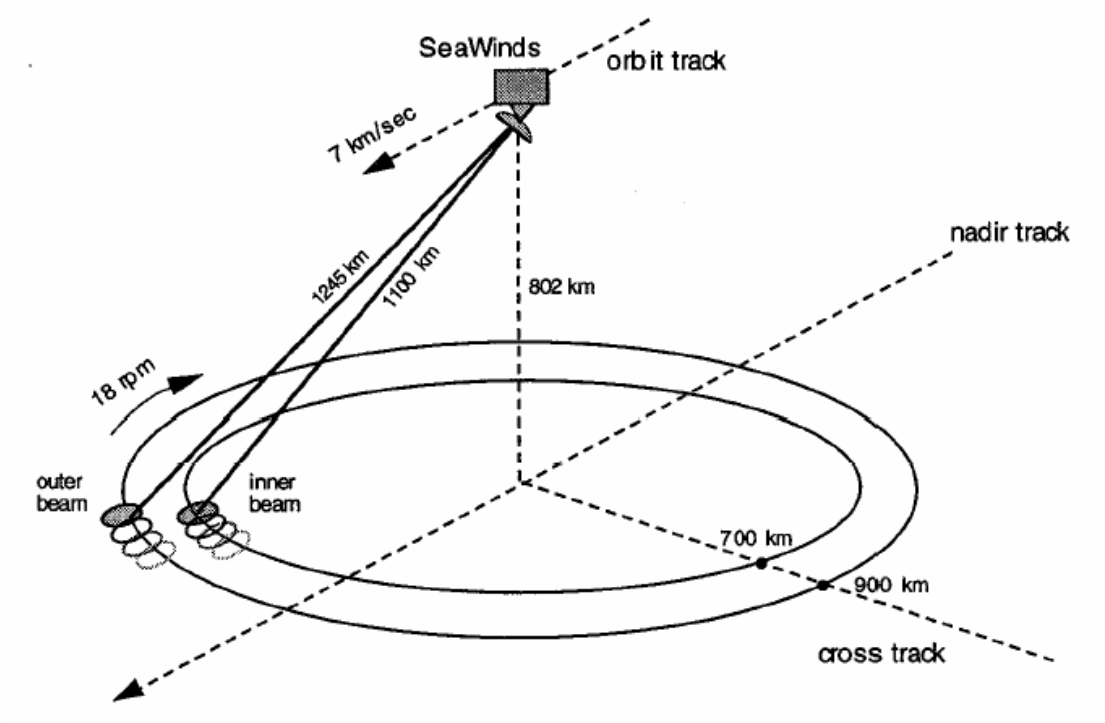


Figure 1-3 Coverage geometry for SeaWinds [15, 16]

The functionality of the SeaWinds instrument and the associated data processing are beyond the scope of this thesis. In the next section, the effect of rain on SeaWinds data is discussed in detail.

1.4 Rain effect on SeaWinds Data

1.4.1 Rain types

Precipitation has two main classifications: convective and stratiform [17]. Convective precipitation regions are generally identified with intermittently strong vertical velocities, high rain rates (>5 mm/hr) and small intense cells (~1-10 km horizontal dimensions). Stratiform precipitation areas are characterized by small vertical velocities, low rain rates (<5 mm/hr) and wide-spread (~100-km horizontal dimension) [18]. If they are high enough to contain frozen particles, they are bounded on the top by a layer of melting ice particles, the "bright band".

1.4.2 Rain effects

It is well known that microwave signals are scattered and absorbed by raindrops located within the signal path [19]. The amount of signal alteration depends on the frequency, polarization and angle of incidence of the microwave signal in addition to the amount of rain present in the path and the speed of the local wind field [20]. The higher frequency of the SeaWinds instrument's Ku-band pulse (~14GHz) suffers greater rain contamination than that of the ERS scatterometer's C band (5.3 GHz) pulse. The presence of rain affects the wind vector measurements in the three following ways.

1. The rain attenuates the radar signal as it travels to and from the earth's surface. This reduces the measured s^{o}.
2. The radar signal is scattered by the volume of raindrops. Some of this scattered power returns to the instrument. This increases the measured s^{o}.
3. The roughness of the sea surface is increased because of splashing due to raindrops. This increases the measured s^{o}. An early study of rain effects on radar scattering from water surfaces can be found in [21] and [22] also describes similar effects.

The energy backscattered by the rain can be a significant portion of the total backscattered power measured by the radar; indeed, for high rain rates and low wind speeds it can dominate the signal. The effects of rain on QuikSCAT winds were found to be severe for low and moderate wind speeds (< 10m/s). The wind retrieval processing does not account for these effects, and thus, the retrieved wind vectors will contain errors.

1.4.3 Case Study

A case study on rain in QuikSCAT scatterometer reveals that both cross-track vectors (presence of wind vectors turned perpendicular to the satellite track) and higher winds are found in the very heavy rain cell [23]. Figure 1-4 shows wind vectors over the tropical storm Olga on March 2000 collocated with Special Sensor

Microwave/Imager (SSM/I) rain rate values. The cross-track vectors can be found over the precipitation region (dark gray) toward the west of the storm center.

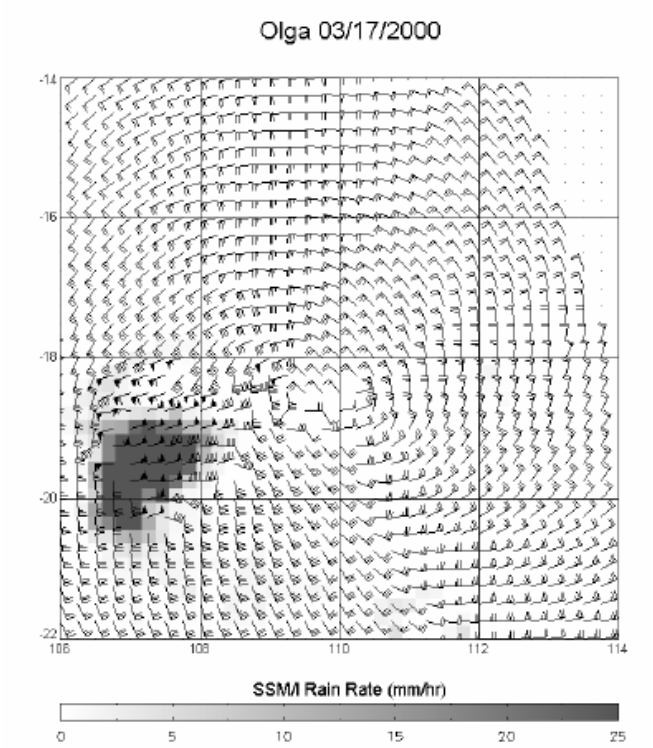


Figure 1-4 QuikSCAT wind vectors, shown as wind barbs in knots, are drawn over collocated F14 SSM/I rain rates for tropical cyclone Olga. Higher wind speeds and cross track directions occur in and near the higher rain rates [23].

Since the early stages of the QuikSCAT Calibration/Validation effort, precipitation has been clearly identified as a significant source of contamination. This can be particularly troublesome, as many ocean meteorological events of prime interest, namely tropical storms (like the one discussed above) and hurricanes, possess abundant precipitation. Various rain flags have been applied to the QuikSCAT winds including the multidimensional histogram (MUDH) technique used in the Jet Propulsion Lab (JPL) and National Oceanographic Atmospheric Administration

(NOAA) products [24], the normalized objective function (NOF) technique developed by Remote Sensing System (also reported in the JPL wind product) [25] and others [26]. These flags were tuned using SSM/I rain rates. Most of the effort by the research community is directed toward flagging rain-contaminated cells and only a few attempts were made to model the effect of rain and correct wind vector estimates [20, 27].

1.5 Need of Rain Height Estimation in Correction

1.5.1 Correction for Atmospheric Attenuation

The Radar System and Remote Sensing Laboratory (RSL) at University of Kansas is developing an algorithm to correct for the rain effect in the SeaWinds scatterometer signal. This work began originally in connection with the Earth Observing System (EOS) STIKSCAT program. The attenuation of the surface signal due to precipitation can be corrected using radiometric brightness temperature. A radiometer observing the same region as the scatterometer can provide estimates of total attenuation because the observed brightness temperature depends on attenuation through the rain and cloud. The Advanced Microwave Scanning Radiometer (AMSR), which flew along with SeaWinds scatterometer in the ADEOS II satellite,

could be used for this purpose, as they had overlapping coverage. Figure 1-5 shows the rain effects on SeaWinds scatterometer and AMSR radiometer signals.

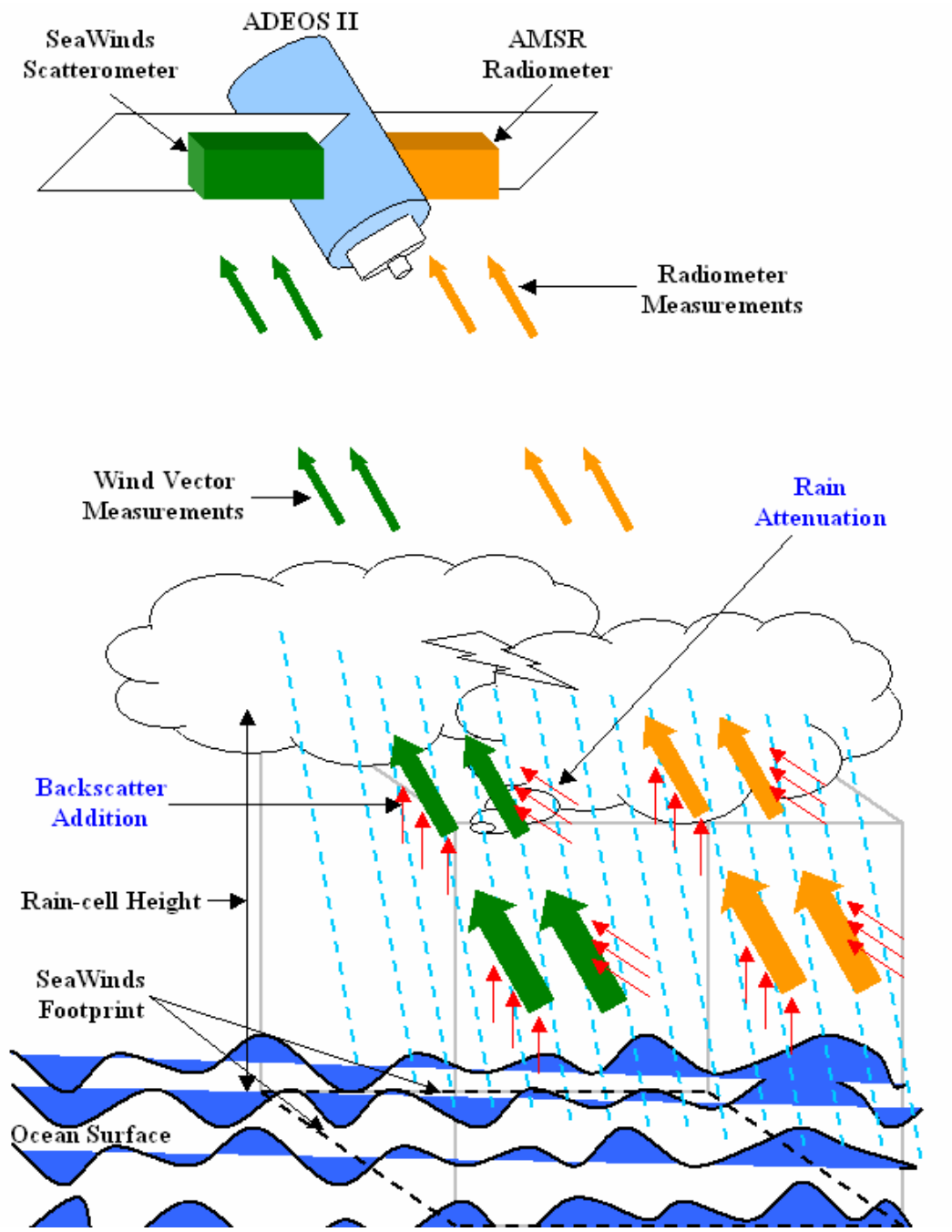


Figure 1-5 Effect of Rain on Scatterometer and Radiometer data

A look-up table based method was devised to determine the radiometer-based attenuation for the SeaWinds Scatterometer [28]. Simulation results on the improvement of scatterometer measurements using radiometer-based atmospheric attenuation correction are given in [29-30]. Another alternative for attenuation correction is to determine the attenuation rate (K) using the AMSR rain rates (RR) with an empirical relation valid for Ku-band frequencies. However, this approach needs rain height to estimate total attenuation suffered by the scatterometer signal. The Rain Height (RH) or Storm Height (SH) (these terms are used interchangeably throughout this thesis) is the height of the effective storm top. The height of the rain cell is also an important input in correcting for volume scatter from rain, which is discussed in the next sub-section. No direct measurement of rain heights was available on the ADEOS-II satellite.

1.5.2 Correction for Rain Volume Scattering

As discussed in the section 1.4, the most deleterious effect of the rain on the scatterometer signal is the backscatter from the rain volume (precipitation echo). The precipitation echo (the volume integral of backscatter) can be obtained from the rain-backscatter intensity and rain volume. The rain-backscatter intensity can be obtained from the AMSR rain rate using empirical relationships for different rain types (stratiform or convective). Rain cell height is required to estimate the volume of the rain cell above each SeaWinds footprint.

With the knowledge of rain cell height, one can estimate total attenuation and rain backscatter for each scatterometer beam and correct for rain effect in the scatterometer signal. However, the effective rain cell height was not available as an output from any of the instruments in ADEOS II satellite; hence the estimation of rain heights is essential.

1.6 Motivation for the thesis

A methodology to estimate rain cell height over oceans forms a motivation to this thesis. This estimated rain cell height should be valid over different seasons and regions of the globe. A study of rain-cell height statistics was carried out using the Tropical Rainfall Measurement Mission (TRMM) Precipitation Radar (PR) rain-height products. But, this climatological mean height provided a very crude estimate of rain heights and is constrained only to tropical regions because of the orbit of TRMM. Different parameters available in the ADEOS II satellite were selected and the consistency of their relation to rain heights was analyzed. A strong relationship between rain height and rain rate was identified and reported in [31]. This thesis focuses on analyzing different regression schemes to evaluate the mathematical relation between rain rate and rain height and developing a universal statistical-model to predict rain height, given the rain rate over various season and oceanic regions.

Chapter 2 gives an overview of the various methodologies adopted for rain height estimation. Chapter 3 provides an extensive analysis on various regression schemes that we tried to relate rain rate and rain heights. Chapter 4 elaborates on the seasonal effects on the slopes and intercepts statistics for the selected regression scheme and how it is accounted using Fourier analysis and the final 'regression-based statistical model' and its use in rain height estimation. A detailed description of the simulation approach to validate the rain height estimated by the model is given in Chapter 5.

2 Rain Height Estimation

2.1 What is Rain Height?

Rain height as used here is the height of the top of the rain column above the mean sea level. Rain height is important for several meteorological and climatological applications, and it is also useful in the design of satellite communication systems affected by rain attenuation. Cloud height is also useful for atmospheric research [32]. Cloud heights and types reveal the thermodynamic and hydrodynamic structure of the atmosphere. Since clouds extend above the region where rain drops form, rain height is less than cloud height. Cloud height, of course, may be measured when no rain is present.

2.2 Rain height measurement using TRMM PR

Cloud top height can be estimated from various weather-satellite sensors using stereo images [33-35]. It may also be estimated using infrared (IR) images, since the temperature of the top of the cloud is related to its height [36]. The details of cloud-height estimation are beyond the scope of this thesis.

The first spaceborne precipitation radar (PR), an active microwave instrument aboard the Tropical Rainfall Measuring Mission (TRMM) satellite, was designed to measure rain and its vertical structures over the both tropical oceans and continents. TRMM PR directly observes the rain-top height inside cloud [37]. In the standard processing of the PR data, each vertical profile of radar reflectivity is evaluated to detect the surface and the backscatter from hydrometers (mostly rain because the instrument is not sensitive enough to detect the weaker cloud echoes). When at least three consecutive 250-m range bins (750 m) contain statistically significant hydrometer echoes, the highest range bin is flagged and used to describe a 'storm height' above the standard geoid. TRMM PR measures height of the first (highest altitude) echo for the 'rain-certain' case and outputs it in level-2 data set (2A23) as 'stormH'. The level-2 products are instantaneous PR observations. The TRMM Science Data Information System (TSDIS) also produces level-3 (3A25) monthly averaged storm-height products from level-2 products [38]. We use this level-3 storm height product (stormHeightMean) extensively in this study. The details of the TRMM storm height products are available in Appendix A. Figure 2-1 shows, as an example, the global distribution of storm height observed by PR in July 1998.

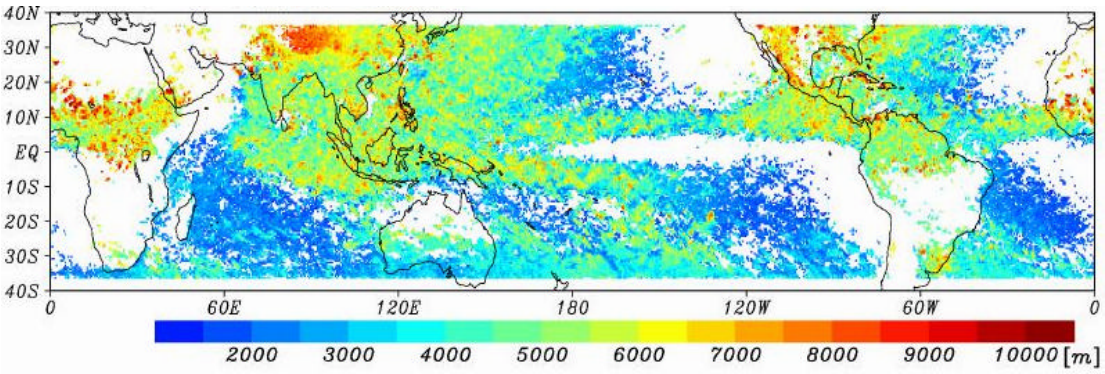


Figure 2-1 TRMM PR Level 3 monthly averaged storm height for July 1998

In the following section, an expression for the total received power by the scatterometer in the presence of rain is derived. Various inputs for the correction algorithm to recover the actual surface scattering coefficient from the total received power are analyzed. Since direct measurement of rain height was not available in ADEOS II satellite, the need for rain height estimation is emphasized.

2.3 Rain-Effects Correction Algorithm

The total power received by the scatterometer in the presence of rain is the sum of the power received from the surface scattering and volume scattering.

The power received from the surface scattering is

$$P_s = s_s^o A_r \qquad \textbf{(2.1)}$$

where s_s^o is the surface scattering coefficient or the normalized radar cross-section (NRCS) and A_r is the surface area that can contribute to the signal.

The power received from volume scatter is

$$P_v = hV \qquad \textbf{(2.2)}$$

where h is the volume backscattering coefficient per unit volume and V is the volume of the rain cell that contributes to the signal. Figure 2-2 shows the geometry for volume calculation.

The volume backscattering coefficient of rain, when drops are small enough relative to a wavelength to permit use of the Rayleigh approximation is [19]

$$h = 10^{-10} \frac{p^5}{l_0^4} |K_w|^2 Z \qquad \textbf{(2.3)}$$

where l_0(cm) is the wavelength, $|K_w|$ is a function of the refractive index of water, and Z (mm^{6}m^{-3}) is the reflectivity factor.

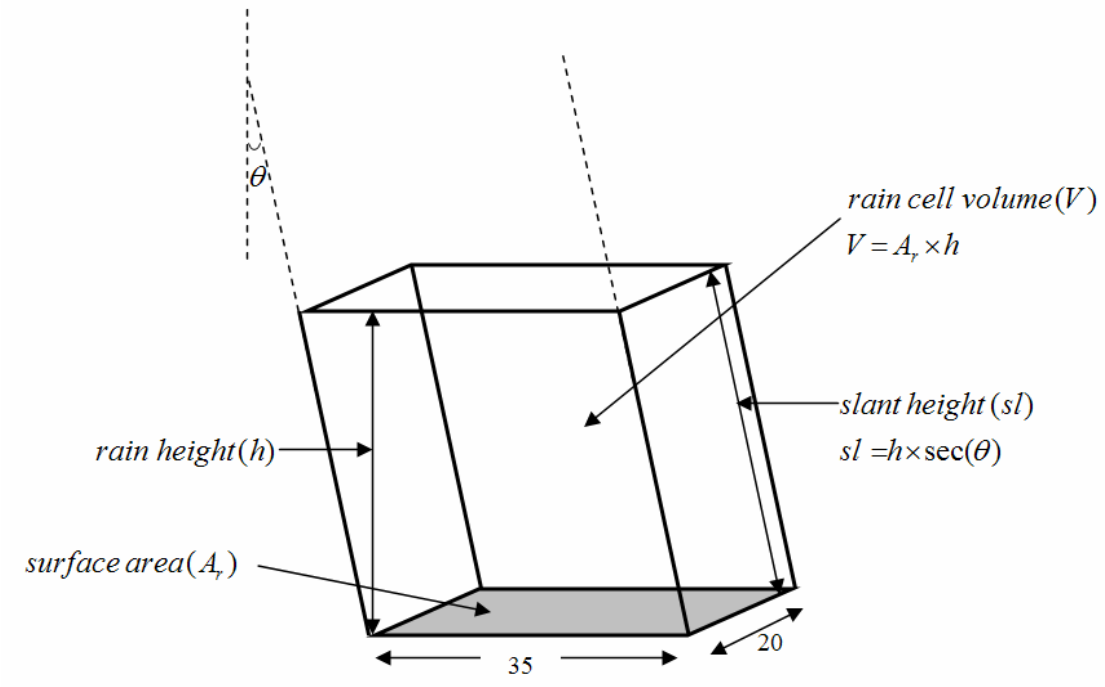


Figure 2-2 Rain cell geometry over SeaWinds footprint (simplified from elliptical footprint)

For the SeaWinds scatterometer, the backscattered power from a given footprint arrives simultaneously for the rain volume and the surface scattering. The backscattered power is also attenuated as a function of the distance through the rain cell volume (slant height 'sl'). Therefore the power received by the scatterometer is given as,

$$P_r = s_r^o A_r = P_s t^2 + P_v x$$

$$s_r^o A_r = s_s^o A_r t^2 + hVx \qquad \textbf{(2.4)}$$

$$s_r^o = s_s^o t^2 + hhx$$

where s_r^o is the *measured* equivalent scattering coefficient and t is the one-way transmissivity of the atmosphere with a value smaller than one:

$$t = e^{-2K_{er}sl} \qquad \textbf{(2.5)}$$

where K_{er} is the attenuation constant and sl is the slant height calculated as $sl = h \times \sec(q)$. The incidence angle is q.

h is the volume backscatter coefficient given by

$$h = 10^{-18} \frac{p^5}{l^4} |k_w|^2 Z \qquad \textbf{(2.6)}$$

The factor $|k_w|^2$ in (2.6) is a function of the refractive index for water and varies from 0.89 to 0.93 over a 0 to 20^{o} C range and a 1 to 10 cm wavelength range; l is the wavelength expressed in meters; Z is the reflectivity factor evaluated using a common

empirical *Z-R* relation, and is expressed in (mm^6 / m^3) x 10^{-18}. The parameter x in (2.4) is a factor to account for attenuation of the rain echo and is defined as [39]:

$$x = \frac{1}{sl}\int_0^{sl} e^{-2K_{er}r}\,dr \qquad \textbf{(2.7)}$$

Equation (2.4) completely models the rain effects in the total received power. The correction algorithm tries to retrieve surface scattering coefficient (s_s^o) from the total received power. Hence the corrected normalized radar cross-section (NRCS) is derived by inverting the equation (2.4), noting that $V=A_r h$:

$$s_{corrected}^{o} = \frac{s_r^o - (h \times h \times x)}{t^2} \qquad \textbf{(2.8)}$$

The basis for the rain-effect correction algorithm is shown in equation (2.8). The correction of surface scattering coefficient (s_s^o) from rain effects requires the following inputs:

i. reflectivity factor (Z)

ii. attenuation constant (K_{er})

iii. rain height (h)

The reflectivity factor and attenuation constant can be obtained using empirical relationships with rain rate. The availability of rain rate from AMSR radiometer in ADEOS II made this approach possible for ADEOS-II, and presumably will do so for future missions. Rain height is the only parameter required to complete the rain-

effect correction algorithm. Since direct measurement of rain height was not available on the ADEOS II satellite (and presumably will not be on other scatterometer satellites), estimation of rain height using other available parameters or statistics is necessary.

Figure 2-3 shows the various inputs necessary to complete the rain-effect correction algorithm. The rain-height parameter can be estimated using one of the following methods:

i. Estimation from ocean rain height statistics (essentially climatology)
ii. Estimation from sea-surface temperature
iii. Estimation from rain rate

The detailed description of these rain height estimation methods, data-processing procedure and feasibility study are given in sections 2.4, 2.6, and 2.7.

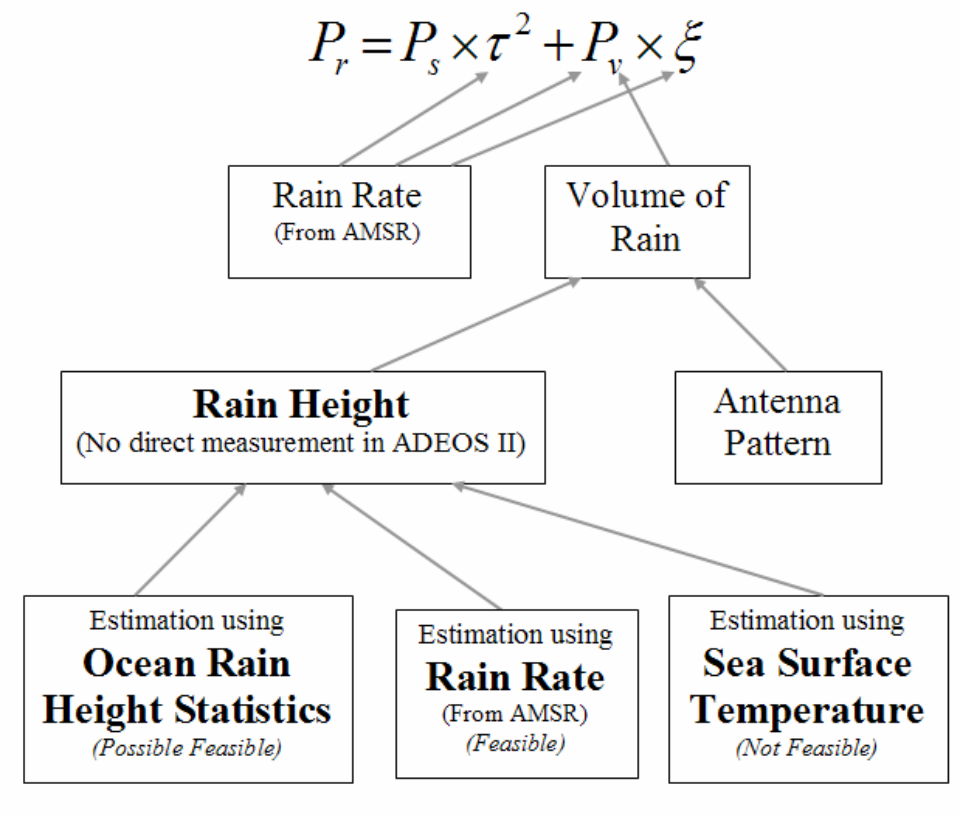


Figure 2-3 Methodologies for rain height estimation

2.4 Estimation using Climatological Rain Height

The first approach, based on ocean rain-height statistics, requires mean-rain-height tables for every month in all the regions of the ocean. Conventional statistics from the weather services do not provide this information. The TRMM PR provides rain-height data throughout the region between 35°N and 35°S latitude starting in January 1998. This period includes El Niño, La Niña and neutral conditions in the South Pacific, so even though it is short, it covers a good range of ocean characteristics. In

this study we use monthly averaged level-3 rain- height data rather than the massive instantaneous level-2 rain height data. We studied the various trends in level-3 rain height data. A similar study on TRMM rain-height statistics is discussed in [40]. The focus of our study is to determine the mean rain height over all oceanic regions, which can be used in the rain-effect correction algorithm. In this section, the analysis of TRMM rain-height statistics in the Pacific Ocean is reported.

2.4.1 Data processing

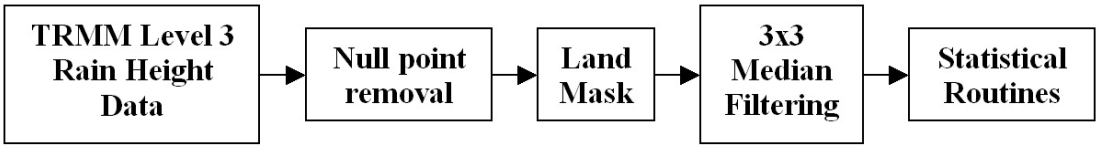


Figure 2-4 Data processing involved in TRMM Level 3 rain height statistical study

The TRMM level-3 monthly averaged rain heights are provided at spacing of 0.5° in latitude and longitude (refer to Appendix A). This data set includes many points with null values, indicating no rain during the month. These null points are removed before processing the data. Some unrealistic heights (some are even greater than 9 km) are ignored in the processing. We used a land mask which includes major continents and islands. Islands in major groups are combined to single areas and some widely dispersed small islands are ignored in this land mask that is used to exclude the land areas from our analysis We used a 3x3 median filter to remove spurious points, mostly those that have excessive heights. The results were then

analyzed using various statistical routines to understand the trends in rain-height data. For simplification in data manipulation, we used only the data sets for January and July months for stratiform and convective rain types were analyzed separately.

2.5 Pacific Rain Height Statistics

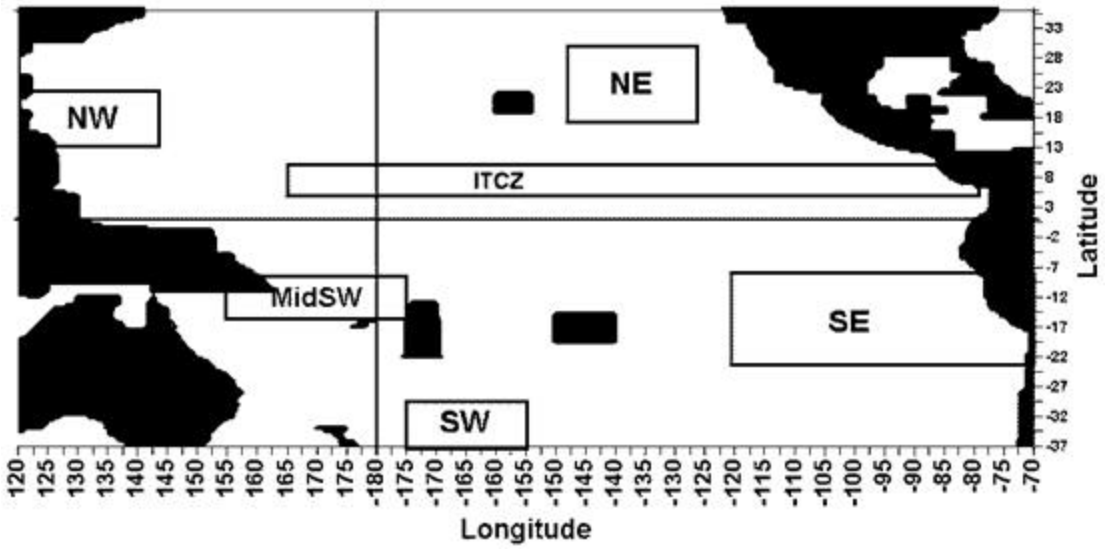


Figure 2-5 Pacific Ocean common regions. Black areas not on continents are island groups. The regions are relatively small because different months showed different boundaries for obviously uniform regions.

The Pacific oceanic region is divided into six homogenous rain height regions to account for variation in regional ocean climate. The common regions were obtained by comparing monthly rain-height maps. These common regions allow month-to-month comparison of rain-height distribution. Figure 2-5 shows the selected regions for the Pacific Ocean. After processing the raw rain-height data, histogram plots were generated for each selected region. Figure 2-6 shows the rain height histograms

for a few regions. Each histogram shows data from the months July 1998, July 1999, and July 2000. The smooth curves are based on 6th-order polynomial regression. A 3-point moving average window was used to smooth the fits.

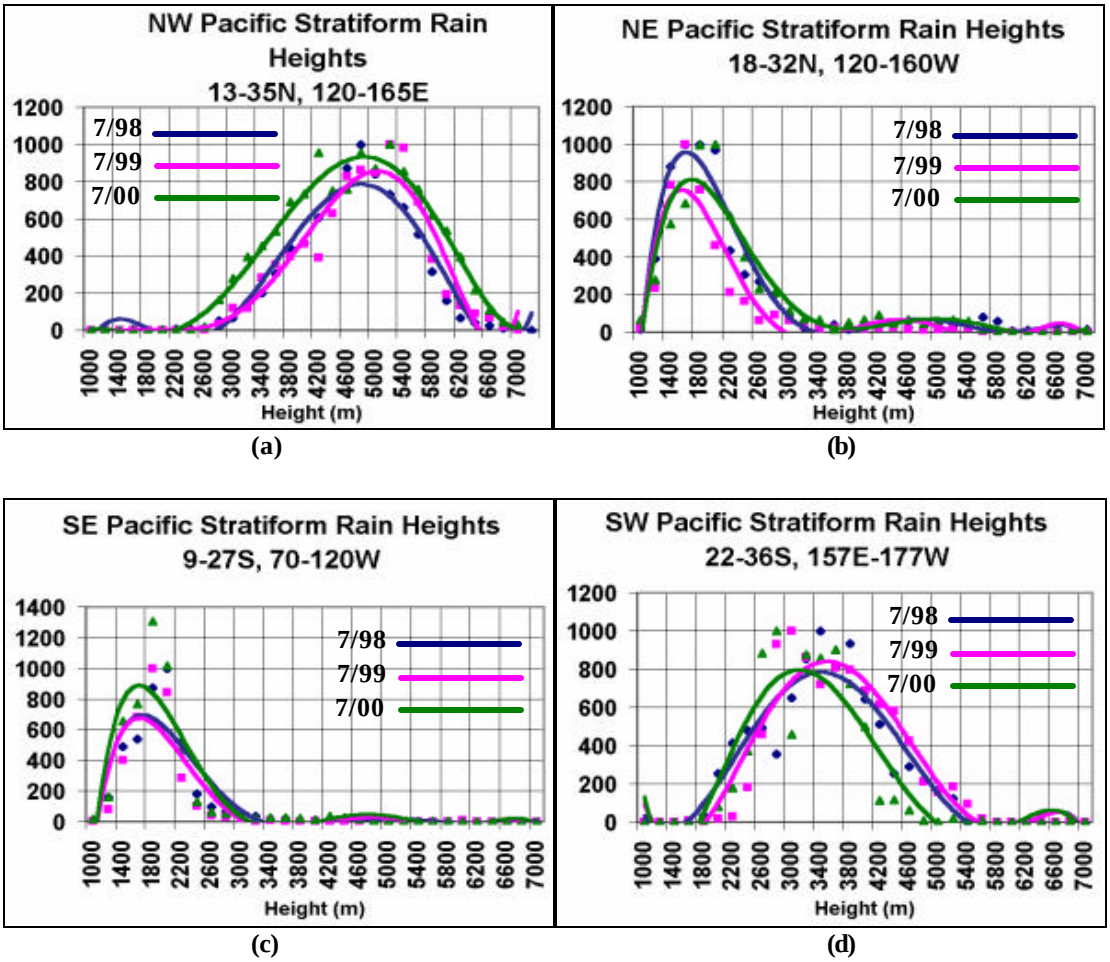


Figure 2-6 Rain height histogram for the selected common regions
(a) NW Pacific, (b) NE Pacific, (c) SE Pacific, (d) SW Pacific.

Similar histograms were generated for all regions, seasons and rain types. To investigate the usability of mean rain-height values in the rain-effect correction algorithm, the mean rain heights were estimated for each common region and tabulated. Figure 2-7 shows the mean rain-height values of the common regions over

the Pacific Ocean for stratiform rain. Similar mean rain-height values for convective rain types are shown in Figure 2-8.

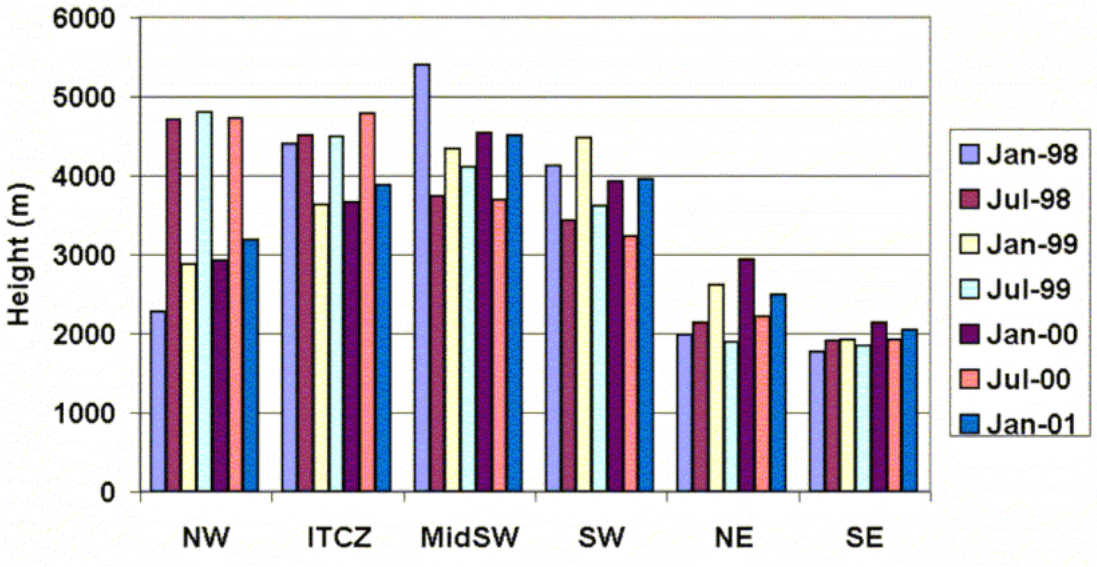


Figure 2-7 Mean pacific stratiform rain heights for the selected common regions

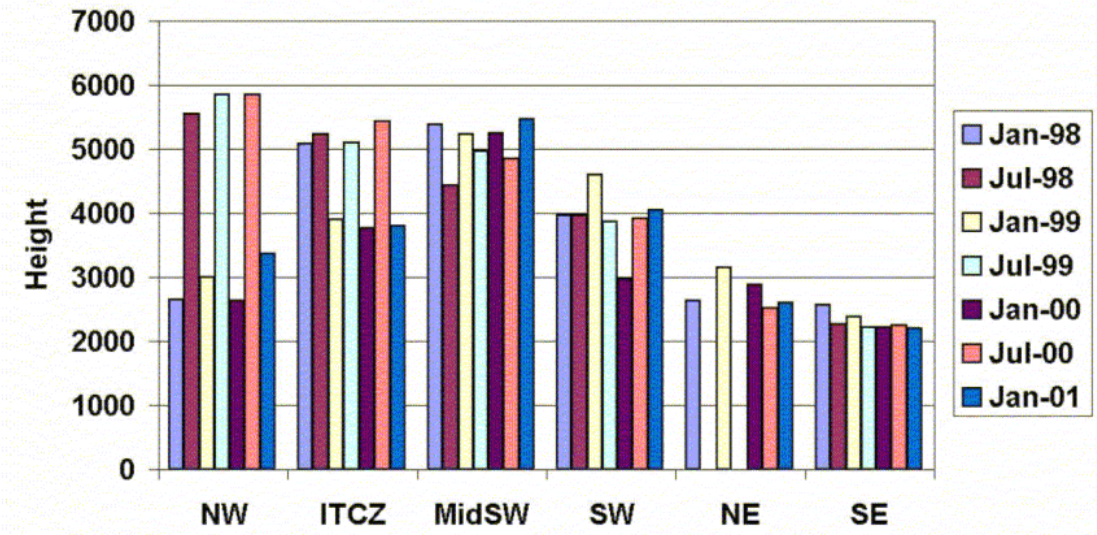


Figure 2-8 Mean pacific convective rain heights for the selected common regions

The graphs of mean rain-height values show variation in different months and common regions. For a given region, the standard deviations are typically less than 2 km. Hence the mean rain-height values for a given common region and month can be used in the rain-effect correction algorithm. Note the seasonal effects.

A major drawback of this approach is the limited latitude coverage of TRMM (± 35°). Hence, mean rain-height tables cannot be developed for higher latitudes. In addition, this approach provides only a statistical measure of rain height. The other approaches can give a better measure of the rain height, if a strong relationship exists between storm height and the parameter (either sea-surface temperature or rain rate) used to estimate the rain height.

2.6 Estimation using Sea-surface Temperature

The second approach tried was to use the measurement of sea-surface temperature (SST) to estimate the rain height. Researchers have shown a strong relationship between seasonal precipitation patterns and sea-surface temperature over ocean [41]. Various independent research studies also illustrate the association of sea-surface temperature with the regional monsoon system [42, 43]. This strong relationship provides a motivation to compare the sea-surface temperature and rain-height data and study the feasibility of estimating rain height using SST. Global sea-surface-temperature products have been produced operationally from National Oceanic and

Atmospheric Administration (NOAA) satellite Advanced Very High Resolution Radiometer (AVHRR) data since 1979 [44-45]. These sea -surface-temperature data were compared with TRMM level-3 rain height data in this study. The details of the products are given in Appendix A. NOAA satellite measurements of SST are available from the Physical Oceanography Archive Center of the Jet Propulsion Laboratory (JPL). The JPL products of monthly global (2048x1024 pixels) 18 kilometer gridded multi-channel sea-surface temperature for January 1998 are shown in Figure 2-9.

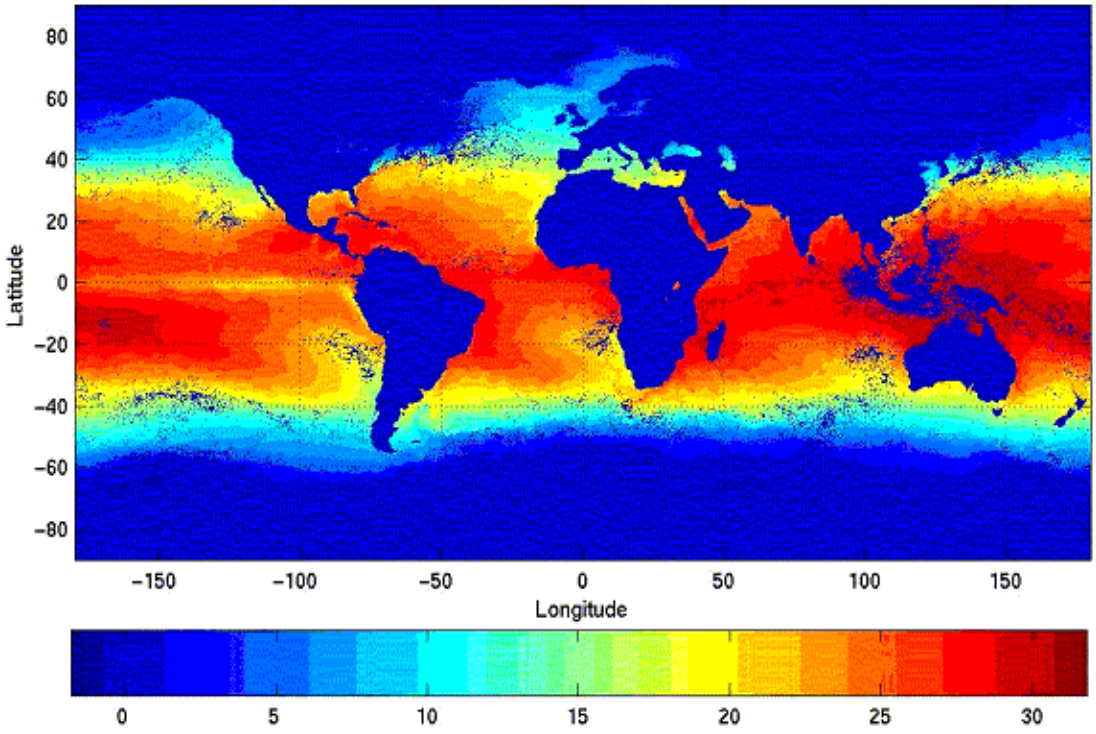


Figure 2-9 Multi-channel sea-surface temperature data derived from the daytime NOAA AVHRR for January 1998. Scales represent temperature in C°

2.6.1 Data Processing

The resolutions of AVHRR SST and TRMM level-3 rain-height data are different. Also the TRMM rain height products are available only from 35°N to 35°S latitude. Hence the sea-surface temperature data were trimmed and interpolated to have one-to-one point correspondence with the rain-height data. The land and null points in both data sets were removed. To account for climatological variations in the different oceanic regions, the entire ocean was divided into 8 regions. The selected regions are Indian north, Indian south, Atlantic north, Atlantic southeast, Atlantic southwest, Pacific northeast, Pacific northwest, Pacific south. In the Pacific Ocean during January, one must separate the inter-tropical convergence zone (ITCZ) region and the equatorial region just south of it from other parts of the sea because of the different conditions that prevail in these special areas. The sea-surface temperature and rain-height data from each region were correlated and analyzed using various statistical routines. The results of these statistical routines help to understand the relation between sea-surface temperature and rain height.

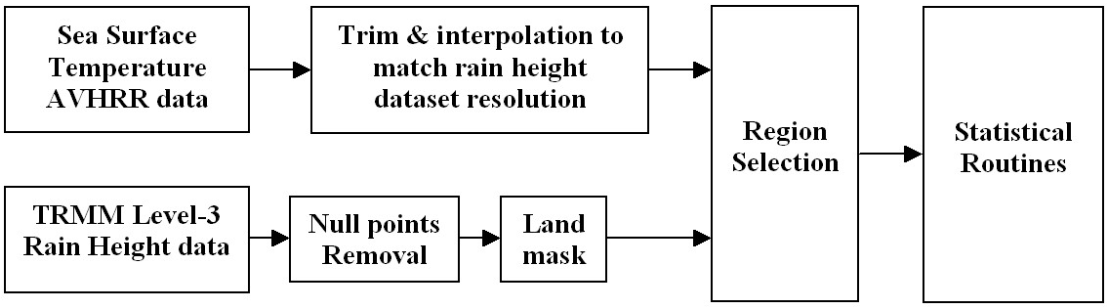


Figure 2-10 Data processing involved in the study of sea-surface temperature and rain height relationship

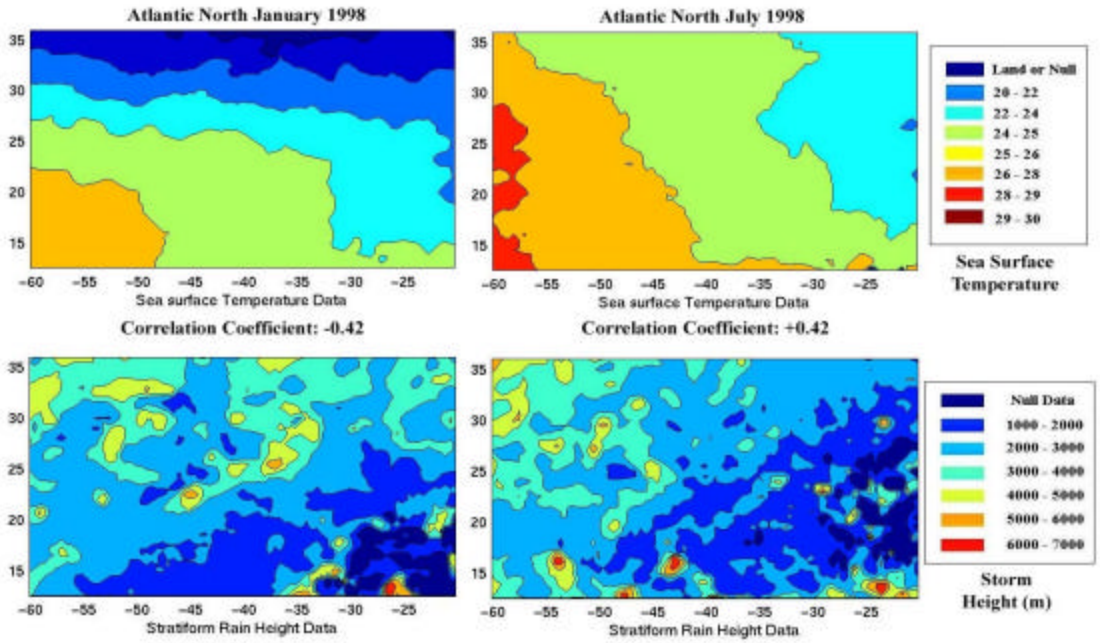
2.6.2 Relation between SST and RH

The processed SST and rain height data from each region were correlated and their correlation coefficients were tabulated. Table 2-1 shows the correlation coefficient values generated for all the eight regions and four different months.

Region	Jan 1998	Jul 1998	Jan 1999	Jul 1999
Indian North	0.57	0.51	0.67	0.35
Indian South	0.45	0.55	0.72	0.58
Atlantic North	-0.41	0.42	-0.35	-0.35
Atlantic South - Eastern	0.66	0.29	0.73	0.50
Atlantic South - Western	-0.00	0.08	0.39	0.51
Pacific North - Eastern	-0.70	0.21	-0.38	-0.48
Pacific North - Western	-0.65	-0.06	0.36	0.02
Pacific South	0.55	0.24	0.60	0.58

Table 2-1 Relationship between sea-surface temperature and rain height illustrated using correlation coefficient generated for various regions and months

The correlation coefficient values from the Table 2-1 show large variation and absence of consistent values over different seasons for the same region. To illustrate this relationship between SST and rain height, the data plots of Atlantic northern region are shown for two different months in Figure 2-11. For January 1998 the correlation coefficient is -0.42 and for July 1998 the correlation coefficient is +0.42.



(a) **(b)**

Figure 2-11 SST and Stratiform storm height data from Atlantic Northern region (a) for January 1998 (b) for July 1998.

We concluded that this approach of estimating rain height using sea-surface temperature is not promising because the correlations between sea-surface temperature and rain height differ widely in the northern and southern hemispheres and in the eastern and western parts of the oceanic basins. Although some regions show high positive correlations with SST, others sometimes correlate negatively, so the SST based rain-height estimation approach cannot be used throughout the world's oceans.

2.7 Estimation using Rain Rate

The three-dimensional structure of rain has been studied for more than half a century. Ground-based radars have been used to characterize the structure of rain for small to meso-scales for decades. More recently airborne and spaceborne sensors have provided information for large to global scales. Researchers have developed models to associate vertical rain-structure statistics with rain rate using regional ocean data [46]. The availability of AMSR rain rate in ADEOS-II satellite would have helped in estimation of rain height, and one can expect similar rain-rate measurements in future wind-vector scatterometer satellites. However, one can only estimate rain height using rain rate if a strong relation exists between them over all regions and seasons. In this section, we investigate the statistical relations between TRMM rain rate and TRMM rain height and discuss the feasibility of estimating rain height using rain rate.

For our study we compared TRMM 'path-averaged' rain rate and TRMM rain height. The details of the 'path-averaged' rain rates are available in Appendix A. TRMM also provides rain-rate products at fixed altitudes separated by 2 km, but the path-averaged rain-rate product is more appropriate since the attenuation and backscatter throughout the rain columns affect the radar behavior. To account for varying meteorological conditions, the ocean is subdivided into eight regions as shown in Figure 2-12.

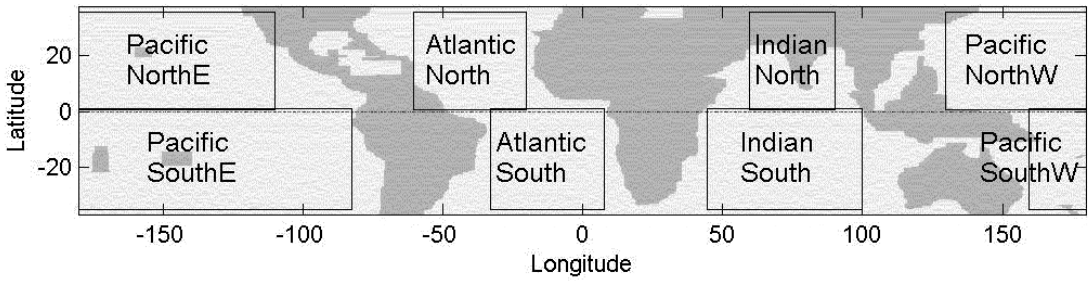


Figure 2-12 Selected regions for RR vs. RH analysis

2.7.1 Data processing

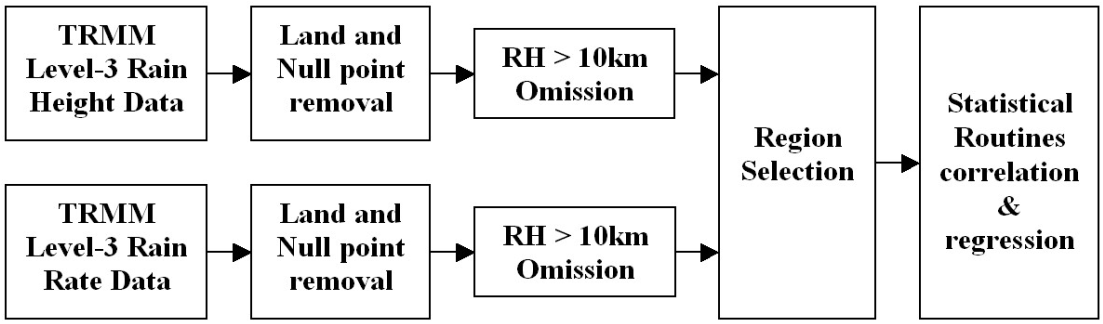


Figure 2-13 Block diagram of data processing steps in rain rate and rain height statistical analysis

The TRMM Science Data Information System (TSDIS) provides TRMM level-3 rain-height and rain-rate products separately for stratiform and convective rain. TSDIS also uses an “other" category for rain data that cannot be classified as either stratiform or convective. We ignore the "other" rain in this analysis. The initial data processing includes masking of land and null points. Extraneous rain height points over ocean can be removed using median filtering, but this median filtering operation affects the one-to-one correspondence of rain rate and rain height data. Hence all points above a threshold rain height (10 km) are omitted. The results of statistical

analysis for stratiform rain type are given in the subsection 2.7.2 and for convective rain type in 2.7.3.

2.7.2 Relation between RR and RH – Stratiform rain

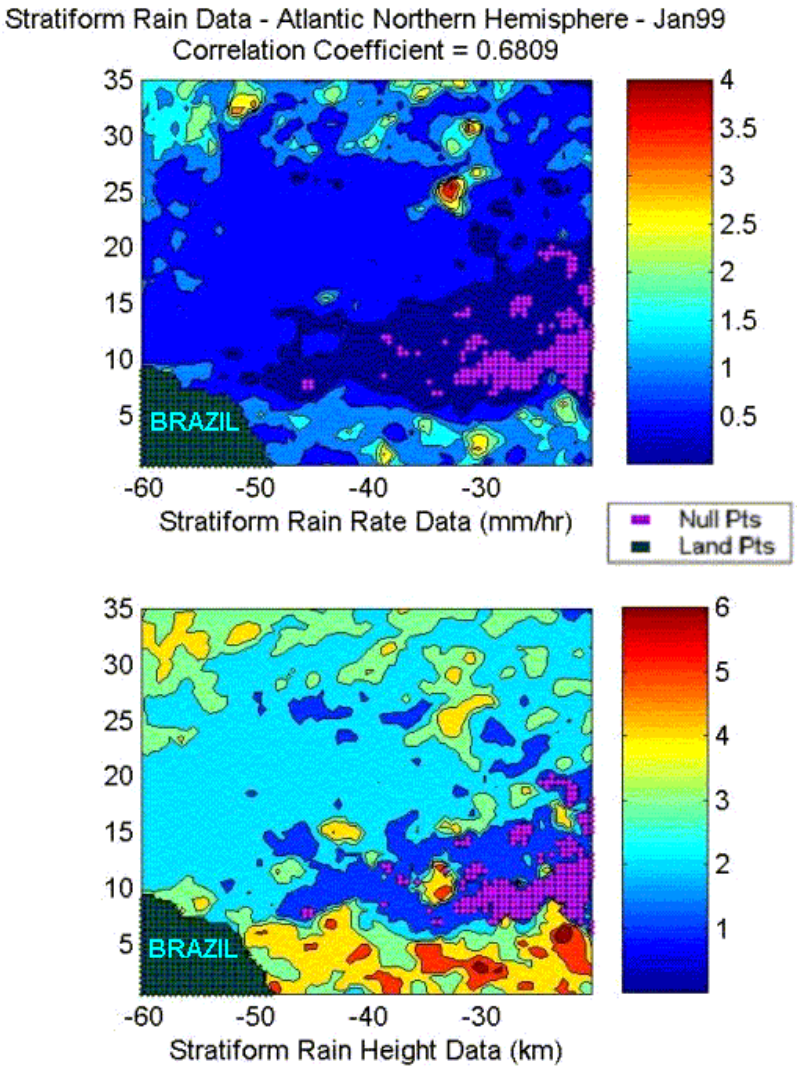


Figure 2-14 Data plots of stratiform rain rate and rain height for Atlantic North region during Jan99. Scales of rain rate data is in mm/hr and rain height data in km.

Figure 2-14 is an example that shows maps for stratiform rain rate and rain height in the north Atlantic region (Jan 99). Note that the 'null' (no rain during the month) points and 'land' points are masked, as we are concerned only with the relationship between rain rate and rain height over the oceans. To analyze the possibility of estimating rain height from rain rate, we correlated rain height with rain rate over different regions on a quarterly basis for 4 years (Jan98 – Jan 02). Figure 2-15 shows the correlation coefficient values for the selected oceanic regions.

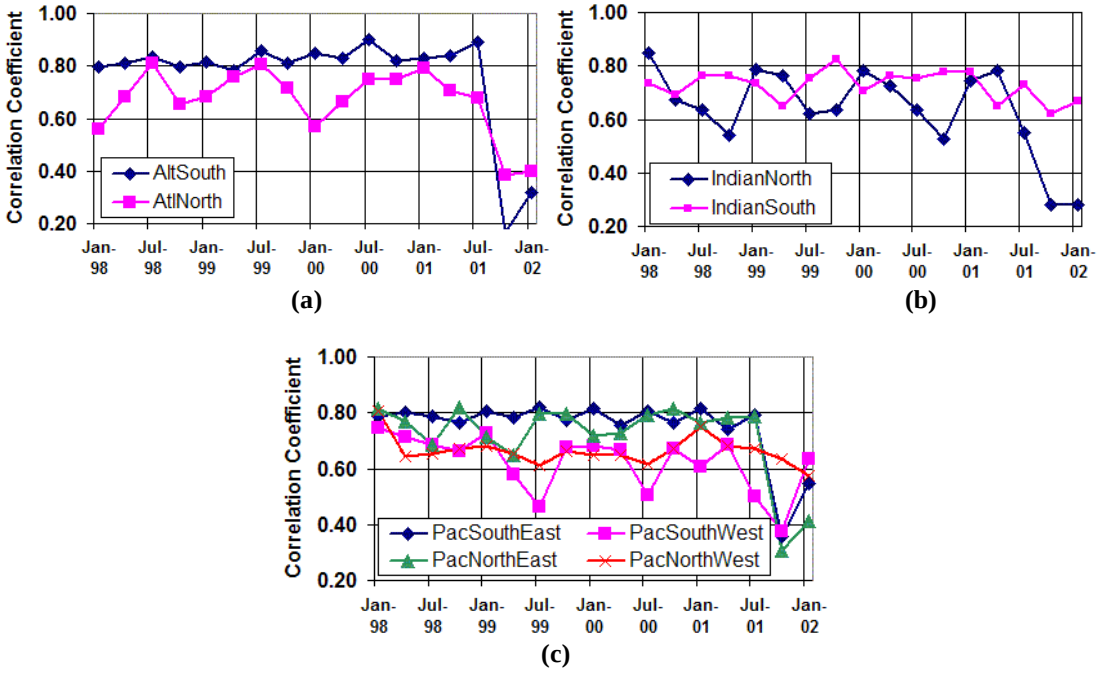


Figure 2-15 Result of correlation between stratiform 'path averaged' rain rate and stratiform rain height over the selected regions (a) for Atlantic (b) for Indian (c) for Pacific Ocean.

The correlation coefficient values are always positive and mostly exceed 0.6 for all regions prior to September 2001. Similar results were obtained by combining data over the entire northern and southern hemispheres. This shows that a generalized statistical model to relate rain rate and rain height can be developed for each

hemisphere. The notable exception is in the data since September 2001, where correlations are quite low in most regions for stratiform rain (note the dip in the correlation values after Sep 2001 in Figure 2-15). This is probably due to the increase in the orbit height of TRMM by 100 km in August 2001. The details of the TRMM orbit change are given in [47, 48]. The report [49] from the NASDA (National Space Development Agency of Japan) PR team shows the impact of orbit change in TRMM PR data, which suggests possible degradation of the observed echo-top height because of reduction in received power. The change of orbit resulted in abnormal presence of a large number of high (physically impossible) rain-height values for light stratiform rain. Figure 1-16 shows this abnormality in stratiform rain data using a mesh plot over the south Atlantic region. The amplitude of the plot represents rain height in km and the color represents rain rate in mm/hr for the months October 2000 (before the orbit change) and October 2001 (after the orbit change).

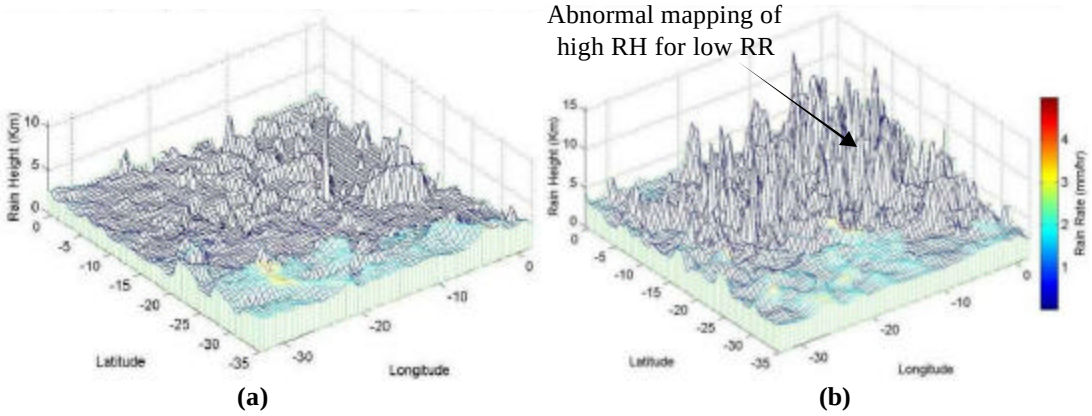


Figure 2-16 Abnormal stratiform rain data showing mapping of high RH for low RR over Atlantic south region (a) for October 2000 – before orbit change and (b) for October 2001 – after orbit change.

This abnormality in stratiform rain data is reported in [50]. This physically impossible set of points greatly reduces the correlation coefficient. The correlation coefficients improve dramatically with the removal of points with rain rate less than 0.5 mm/hr. Figure 2-17 shows improvement in the correlation coefficient values after the removal of points with rain rate less than 0.5 mm/hr. Since these light rain rates (below 0.5 mm/hr) do not significantly degrade the SeaWinds instrument measurement, this abnormal behavior does not affect our analysis. Hence, the stratiform rain data for rain rate less than 0.5 mm/hr have to be removed before analyzing the statistical relation between rain rate and rain height.

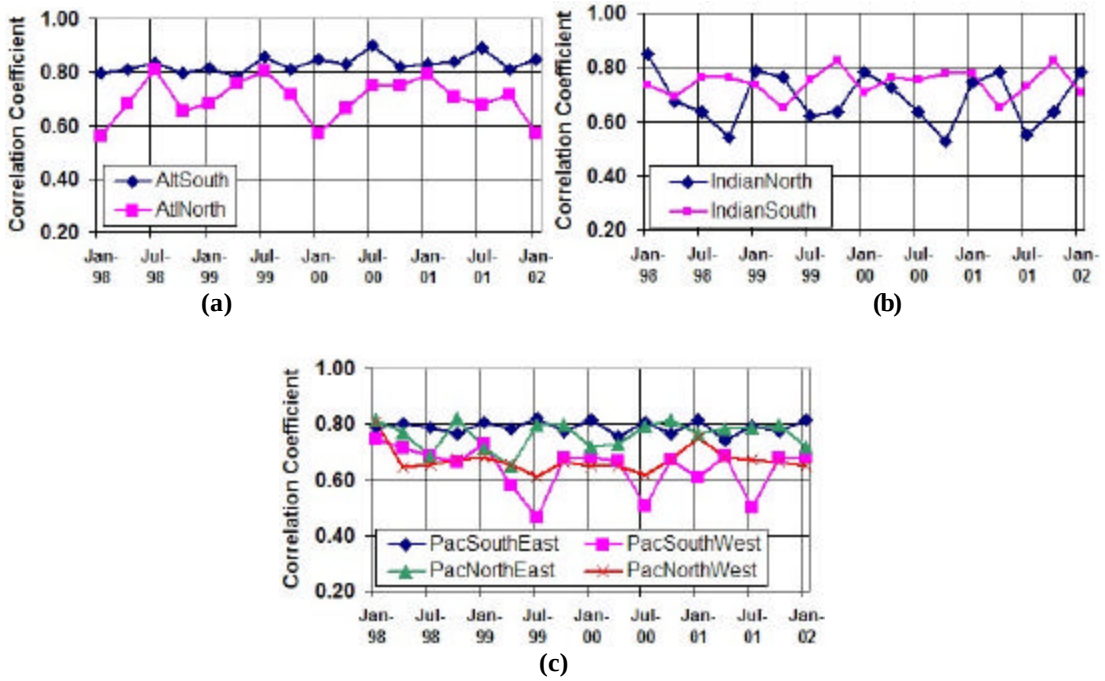


Figure 2-17 Result of correlation between stratiform 'path averaged' rain rate and stratiform rain height over the selected regions with rain rates less than 0.5 mm/hr remove d (a) for Atlantic Ocean (b) for Indian Ocean (c) for Pacific Ocean.

2.7.3 Relation between RR and RH – Convective rain

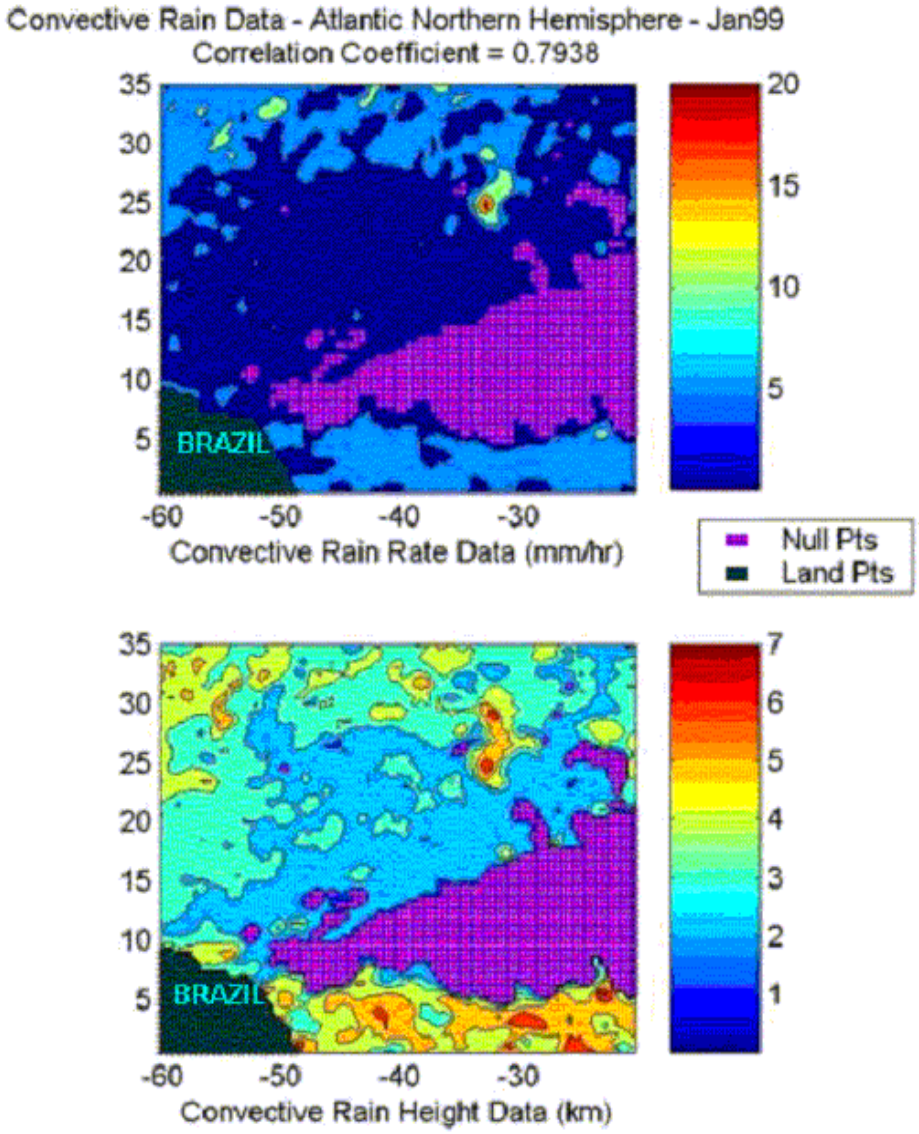


Figure 2-18 Data plots of Convective rain rate and rain height for Atlantic North region during Jan99. Scales of rain rate data is in mm/hr and rain height data in km.

The processed convective rain-rate and rain-height data are shown in Figure 2-18 for the north Atlantic region during January 1999. In general, convective rain rates (stronger signal) are higher than those for stratiform rain. This can be understood by

comparing the rain rate scales in Figure 2-14 (stratiform) and Figure 2-18 (convective). Figure 2-19 shows the correlation between convective rain rate and rain height. The correlation coefficient values are high for all periods with the convective rain, so there is little question about their utility. Presumably this is because the stronger convective signals are not significantly affected by the weak-signal noise that affects the stratiform data.

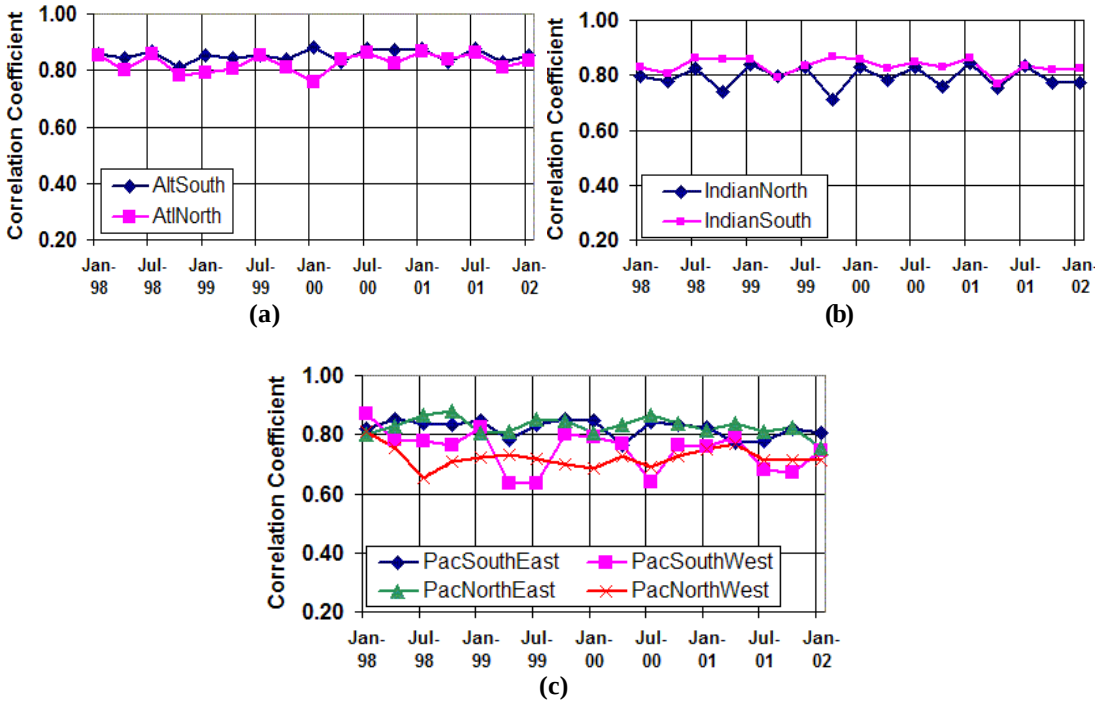


Figure 2-19 Result of correlation between convective 'path averaged' rain rate and convective rain height over the selected regions (a) for Atlantic (b) for Indian (c) for Pacific Ocean.

The strong correlation between rate and height over the ocean demonstrates the feasibility of estimating rain height using rain rate. The next chapter provides details of regression analysis to model the statistical relation between rain rate and rain height over all rain types and regions.

3 Regression Analysis

Regression analysis can also be used to identify the mathematical dependency of one random variable to another random variable [51]. This mathematical relationship (or the model) can be used to estimate the dependent variable for any given instance of the independent random variable. In our study, the mathematical dependence of rain height to rain rate is analyzed using regression analysis. The regression model will help to estimating the unknown rain-height parameter for each rain-rate value obtained from the AMSR radiometer. TRMM level-3 monthly averaged rain-rate and rain-height products are used in this regression analysis and in the development of the mathematical model. The next section provides the various scatter plots to visualize the dependency of rain rate to rain height.

3.1 RR vs. RH Scatter

Plotting individual data points in the xy plane produces a scatter plot. Scatter plots are used to visually identify relationships between the paired data. Figure 3-1 shows the scatter plots for the Atlantic northern hemisphere region for stratiform and convective rain types. From Figure 3-1, one can observe that the stronger convective rain requires a higher rain-rate scale (0-16 mm/hr) than the stratiform rain (0-5

mm/hr). The number of valid data points for stratiform rain in a region is much larger than for the sparsely occurring convective rains for the same region.

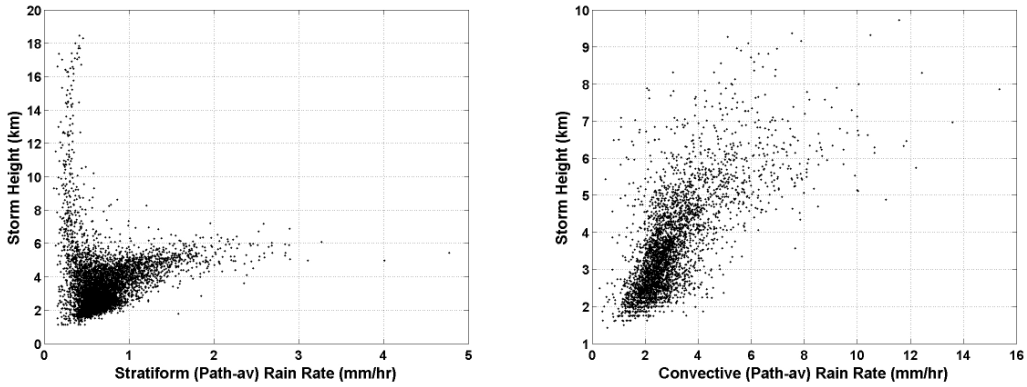


Figure 3-1 RR vs. RH scatter plot of Atlantic northern region data in January 1998 for (a) Stratiform rain and (b) Convective rain

3.1.1 Abnormal data removal

Stratiform rain scatter shows an abnormal spread of rain-height values for lower rain rates. Refer to Figure 3-1 (a), where the rain-height values are spread from 1 km to 18 km for rain rates less than 0.5 mm/hr. This abnormality was greater after the TRMM orbit change in August 2001 (similar effects are discussed in section 2.6). This huge spread of rain-height values for light stratiform rain is not physically possible and it affects the correlation between rain-rate vs. rain-height scatter, reducing it. To remove these spurious points, we initially attempted to trim the data set with the following conditions:

1. Remove data pairs with rain rate less then 0.5 mm/hr
2. Remove data pairs with rain height greater than 10 km

Figure 3-2 shows the scatter plot of stratiform Atlantic northern region data for January 1998 with points corresponding to RR < 0.5 mm/hr and RH > 10 km removed.

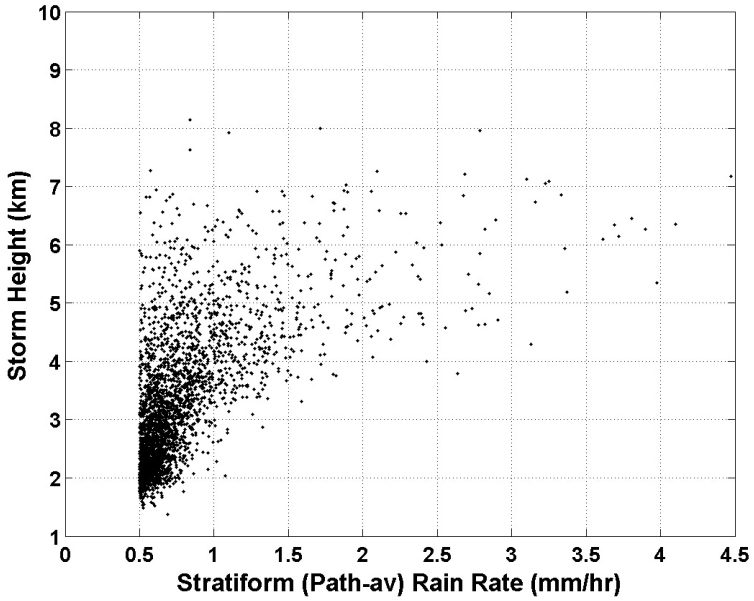


Figure 3-2 RR vs. RH scatter plot of stratiform Atlantic northern region data in January 1998 with points corresponding to RR < 0.5 mm/hr and RH > 10 Km removed

TRMM level-3 high-resolution statistics provide stratiform and convective rain-count products. The details of the rain-count products are given in Appendix A. The rain count represents the number of valid rain observation available in each 0.5 x 0.5 degree grid for a month. The average of these rain observations is presented as the monthly averaged rain-rate and rain-height values in the TRMM level-3 data set. Figure 3-3 shows the histogram of rain count data for the Atlantic northern region for a sample month.

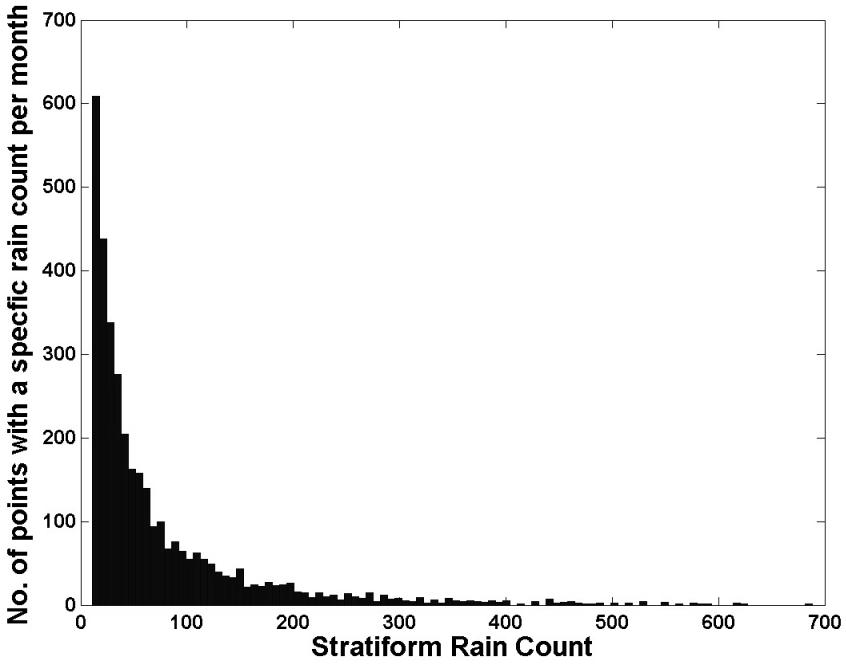
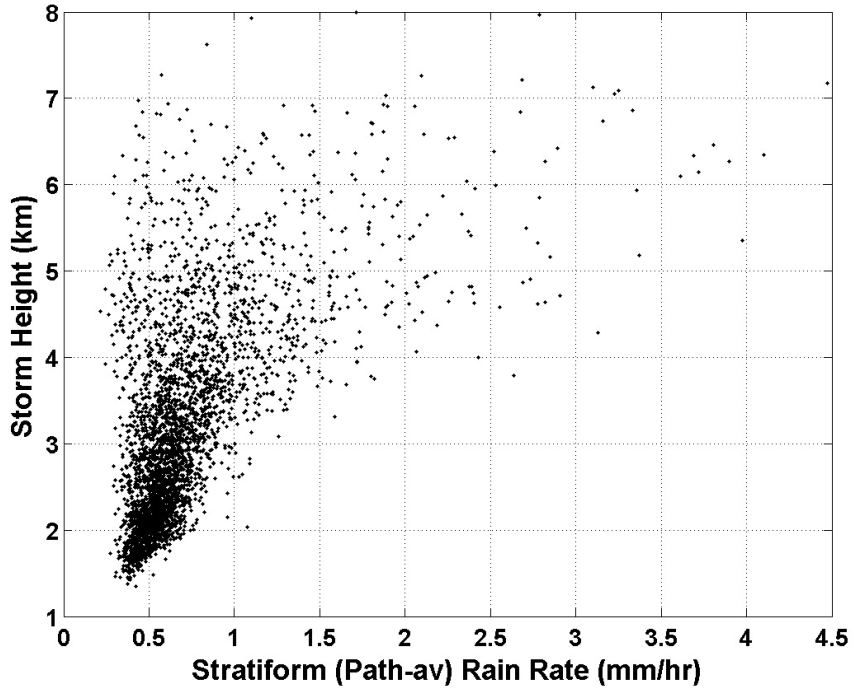
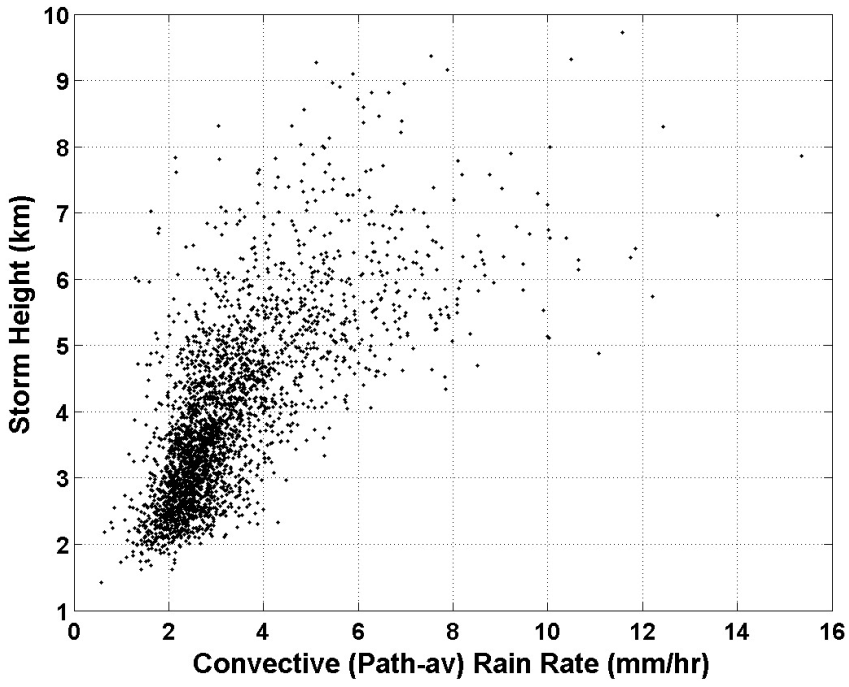


Figure 3-3 Histogram of rain count for stratiform Atlantic northern region (January 1998)

A larger rain count represents a statistically significant monthly averaged rain datum. Hence, this rain-count information can be used to remove the spurious points. Most of the spurious points have rain counts less than 10 for stratiform rain. This rain-count threshold (less than 10) was determined after experimenting with various rain count removal thresholds and their effects on the data set. For convective rain, this abnormal spread of rain-height values for lower rain rates is not present (the TRMM algorithm requires a signal >40 dBZ for convective classification). However, a rain-count removal threshold of less than 2 was set to remove any spurious points present in the convective data. Figure 3-4 shows the scatter plot of Atlantic north for January 1998 with the spurious points removed based on rain count.



(a)



(b)

Figure 3-4 RR vs. RH scatter plot of Atlantic northern region data in January 1998 with spurious point removed based on rain count
(a) Stratiform rain – rain count < 10 (b) Convective rain – rain count < 2

The data-processing steps in rain-rate vs. rain-height regression analysis are shown in Figure 3-5. A similar block diagram of the data processing steps can be found in Figure 2-13. Note the replacement of 'RH > 10km omission' block in Figure 2-13 with the 'Rain count omission' block in the modified data processing steps shown below.

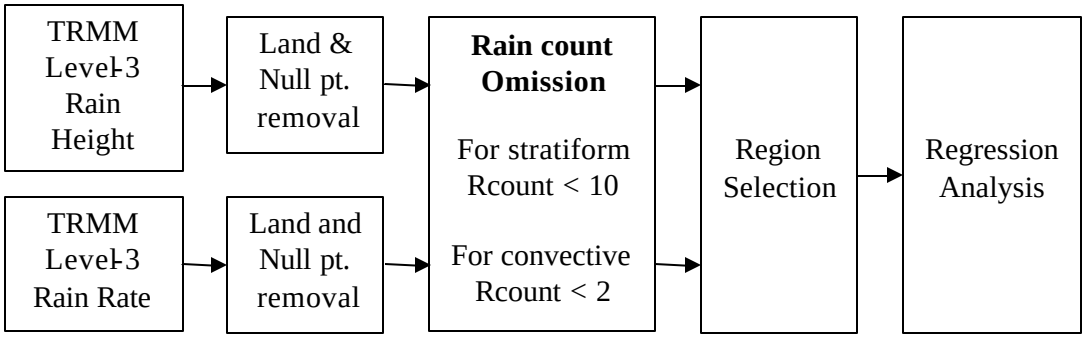
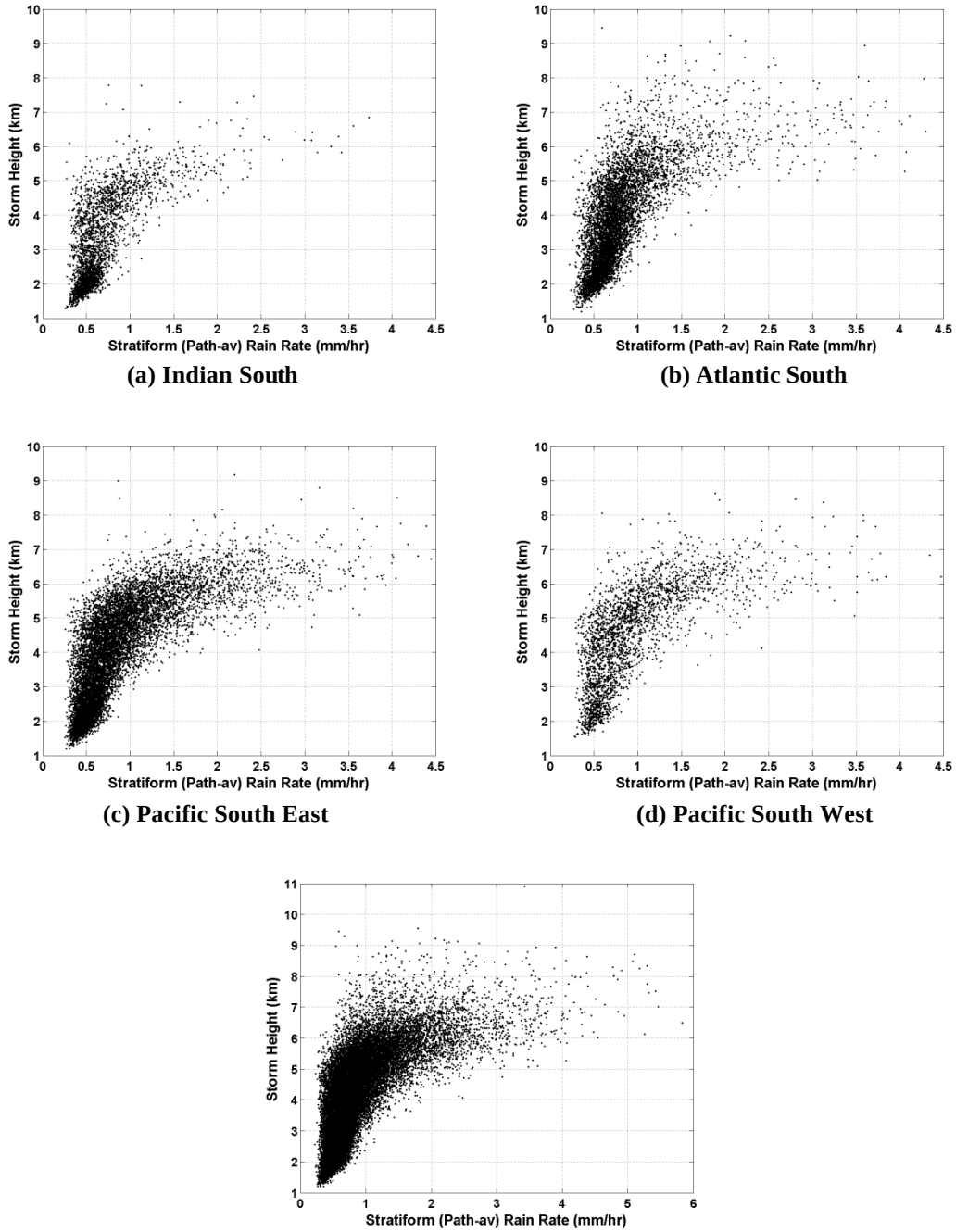


Figure 3-5 Data processing steps involved in rain-rate vs. rain-height regression analysis

3.1.2 Hemispherical and Regional Scatter

Rain-rate vs. rain-height scatter exhibit similar trends for data based on individual regions and the entire hemisphere. To illustrate this, Figure 3-6 shows the scatter plot for stratiform rain in the southern hemisphere and all four regions in it (Indian south, Atlantic south, Pacific southeast, Pacific southwest). Consistent trends in scatter were also observed for all sections of the northern hemisphere and during all seasons. The convective rain data also show similar trends in rain-rate vs. rain-height scatter.



(a) Indian South **(b) Atlantic South**

(c) Pacific South East **(d) Pacific South West**

(e) Southern Hemisphere

Figure 3-6 Comparison of hemispherical and regional RR vs. RH scatter in January 1998 (a) Indian South (b) Atlantic South (c) Pacific South East (d) Pacific South West (e) Southern Hemisphere

3.2 Objective of Regression Methods

The rain-rate vs. rain-height scatter plots exhibit two different trends for different rain-rate regimes. For lighter rain, the RH vs. RR relation is steep. However, for higher rain rates, the scatter almost saturates. Figure 3-7 shows these trends for typical stratiform rain data. The blue strip represents the steeper RR vs. RH scatter and red strip represents the almost flat RR vs. RH scatter. It can also be observed that the scatter is more concentrated for lower rain rates and it is widely spread for higher rain rates. The convective rain data also exhibits similar trends in the rain-rate vs. rain-height scatter. The objective of the regression method is to fit these trends with regression curves and make them consistent over all regions, seasons and rain types.

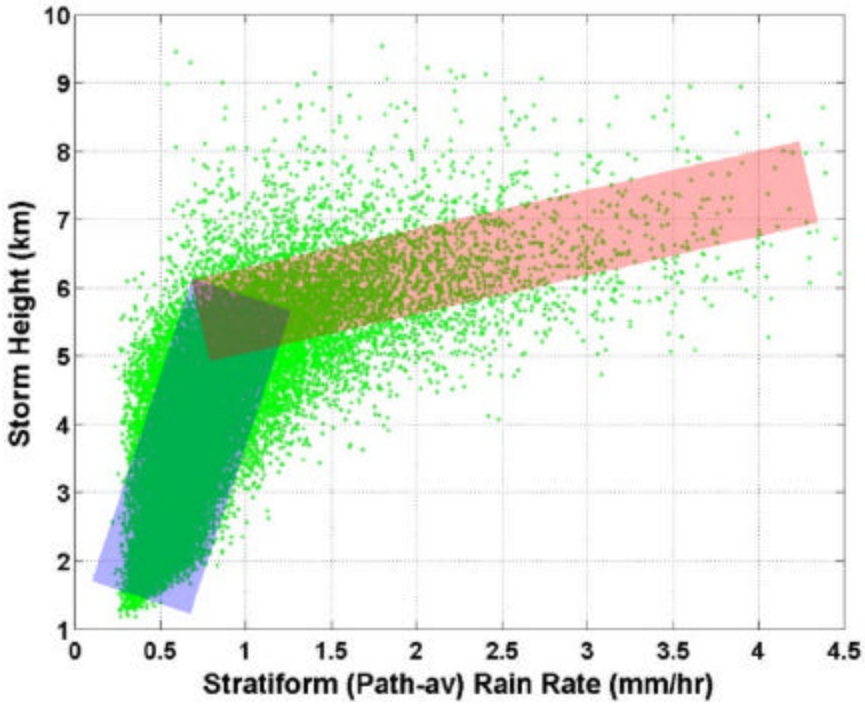


Figure 3-7 Two different trends exhibited in rain-rate vs. rain-height scatter – the blue strip represents steeper relation for lower rain rates and the red strip represents flat relation for higher rain rates

Listed below are a few important aspects that need to be considered in the selection of the appropriate regression scheme:

1. Regression scheme should give a good fit to the scatter, and be consistent over all rain rates.
2. A typical RR vs. RH scatter shows a concentration of large number of scatter points over the lower rain-rate section compared to few widely-spread points for higher rain rates. Hence it is important to get a good fit in the lower rain-rate section.
3. A smooth transition from one trend to another has to be achieved in the regression curves to maintain consistency.
4. A regression scheme should be selected such that the distribution of estimated rain height is consistent with the original rain-height distribution.
5. Identification of a universal regression scheme that applies to various regions, seasons and rain types is important because, for simplicity, one should have a single regression model for the final prediction of storm height from rain rates.
6. The model should allow linear extrapolation of the regression curves to higher rain rates; this is more important than finding the best function that reduces total mean-squared error for the available data.

The next section provides a summary of the all the regression methods investigated and the reasons for selection or rejection of each method in the final regression analysis.

3.3 Regression methods

Standard linear regression schemes have to be modified to satisfy the criterions mentioned in section 3.2. Higher order polynomial regression schemes are omitted in our study, because they are very poor in extending the functional relationship (by extrapolation) beyond the scope of available data. The technical information about the regression methods used in this study can be found in [51-55]. After analyzing various regression schemes, the 'Log-Linear Combined' regression scheme is finally selected as the best regression schemes. This regression scheme fits log (RR) vs. RH in the lower rain-rate section and RR vs. RH in the higher rain-rate section.

3.3.1 Multi-Linear RH vs. RR Regression

3.3.1.1 Linear Regression

The simplest functional relationship between two variables is linear. The most common method for fitting a straight line is the method of least squares. This method calculates the best-fitting line for the observed data by minimizing the sum of the squares of the vertical deviations from each data point to the line. Figure 3-8 shows the single line fit for the RR vs. RH scatter of stratiform Atlantic northern hemisphere data in January 1998. Clearly the single straight line fit is not suited for the multi-trend rain-rate vs. rain-height scatter and it produces a very high standard error of estimate (2.5 km for this example). Also, the extrapolated regression line will result

in unrealistic rain heights for higher rain rates. Hence, multiple regression lines have to be used to achieve a good fit to this type of scatter.

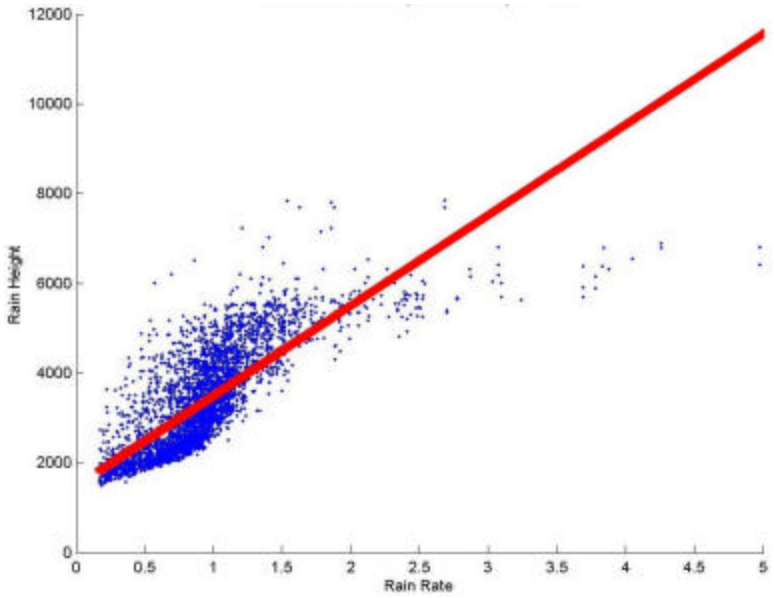


Figure 3-8 Linear least square fit for stratiform Atlantic north data in Jan 1998

3.3.1.2 Bilinear Regression

In bilinear regression, two straight lines, one for higher and another for lower rain-rate sections are regressed to get better fit over all rain rates. A break point must be selected to separate the scatter. After analyzing the trend in numerous scatter plots, a break point rain rate of 1.5 mm/hr was selected for stratiform rain and one of 4mm/hr for convective rain. An objective function of bilinear regression was devised to minimize the sum of the squares of vertical deviation from each point to the line fit for both the rain-rate sections. Additional constraints were introduced in the objective function to make the ends of the regression lines meet. The derivation of the bilinear regression scheme is given in Appendix B. The unknown slopes and

intercepts of the regression lines are found for a given set of scattered data and a break point. Figure 3-9 shows the bilinear regression lines for RR vs. RH scatter of stratiform rain data over the Atlantic northern region in January 1998.

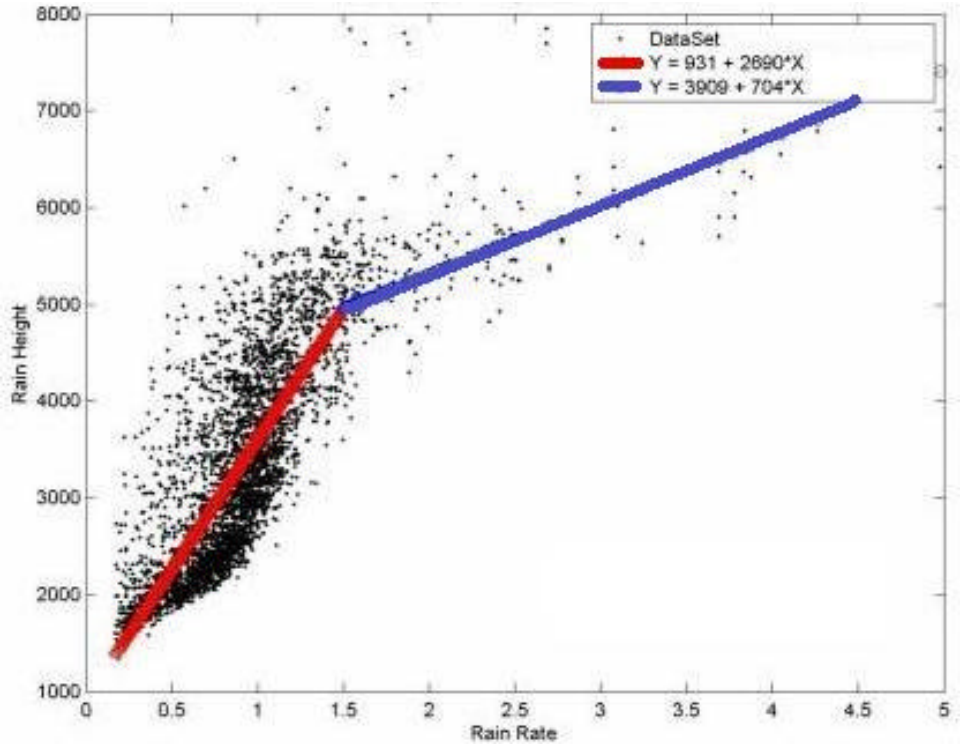


Figure 3-9 Bilinear fit for RR vs. RH scatter of stratiform Atlantic north data in January 1998

Bilinear regression provides a good fit for higher rain-rate sections. However, the fit is very poor for the lower rain-rate section. As discussed in section 3.2, it is very important to achieve good regression fit for lower rain rates, as more points are concentrated there.

3.3.1.3 Trilinear Regression

Some rain-rate vs. rain-height scatter plots exhibit three different trends along the rain rate. This type of scatter is most commonly found in TRMM level-3 2-km stratiform rain-rate vs. rain-height data. Trilinear regression attempts to fit these points with three regression lines, which are joined together to maintain consistency. Figure 3-10 shows the trilinear fit for RR vs. RH scatter of stratiform rain data over the Atlantic northern region in January 1998. As with bilinear regression, trilinear regression lines also fit poorly for the lower rain-rate section. The convective RR vs. RH scatter never shows trilinear trends. Hence the trilinear regression methods were not selected, because we were trying to develop a universal regression scheme that is valid for all seasons, regions and rain types.

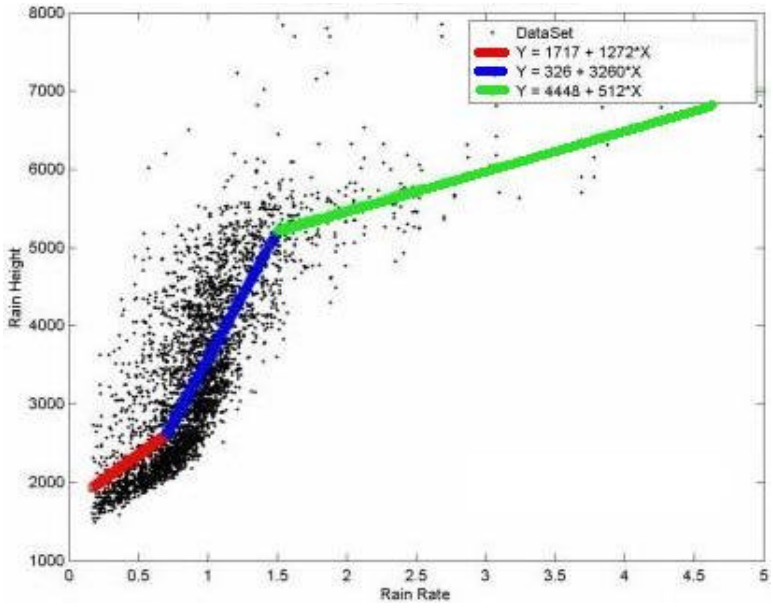


Figure 3-10 Trilinear fit for RR vs. RH scatter of stratiform Atlantic north data in January 1998

3.3.1.4 Orthogonal Regression

Ordinary linear regression was derived on the assumption that the values on the *x* axis are known exactly, with all random variation in the *y* direction. Often the uncertainty in the (*x, y*) data lies with both coordinates. Examples where this is the case includes situations where both x and y are observed quantities having random variations (TRMM rain rate and rain-height parameter are such quantities). In these cases, orthogonal regression can be the best way to determine the linear relationships between *x* and *y*. In orthogonal regressions, the sum of the squared perpendicular distance from the data points to the fit line is minimized. For orthogonal regression to work well both the x and y numerical scales should be similar or they should be adjusted to agree with the variance in the separate directions

Bilinear orthogonal regression should be used to capture the dual trend present in the rain-rate vs. rain-height scatter. This requires the objective function to reduce the squared perpendicular distance between data points and fit line in both of the rain-rate sections and also to make the lines to meet a point. It can be shown that the objective function of the bilinear orthogonal regression with additional constraints cannot be solved in closed form and this scheme was not further analyzed in our regression study. Figure 3-11 shows the two different orthogonal regression fits for the two selected rain-rate sections. The orthogonal regression fails to capture the relatively flat trend in the scatter for higher rain rates.

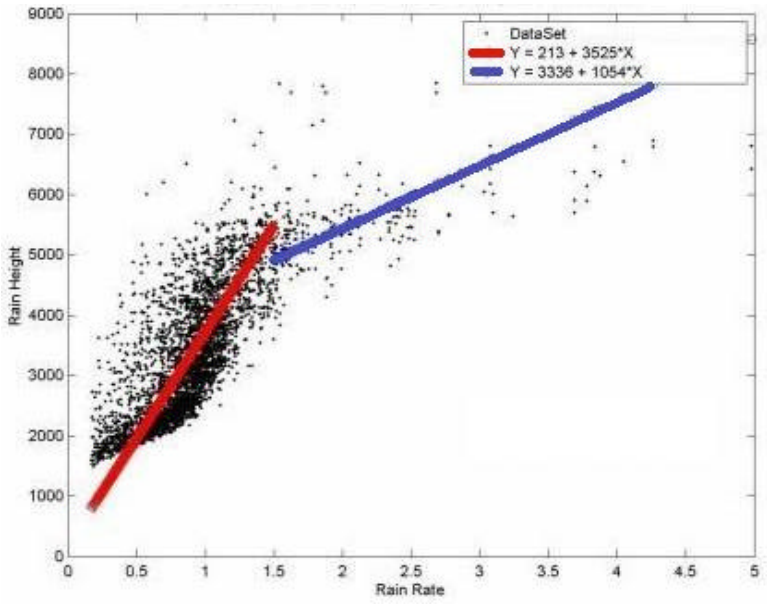


Figure 3-11 Orthogonal regression fits for rain-rate vs. rain-height scatter

3.3.2 Weighted Regression

The TRMM level-3 high-resolution statistics provide stratiform and convective *rain-count* products. The rain count is the number of valid rain observations used to produce the monthly average for a grid point. A (RR, RH) data pair with very high rain count represents a statistically significant pair. Figure 3-12 shows the stratiform rain-rate vs. rain-height scatter over the Atlantic north region. The data points in the scatter are color coded and weighted with corresponding rain count values. Typically, the regression schemes should provide lesser weight to the numerous dark blue points (low rain count) found near lower rain rates than the dark red points (high rain count). Rain-count data should be used in the regression schemes to provide a better fit to the RR vs. RH scatter. Appendix B derives the weighted bilinear

regression, which is a modification to the bilinear regression schemes with the rain-count information used in the objective function.

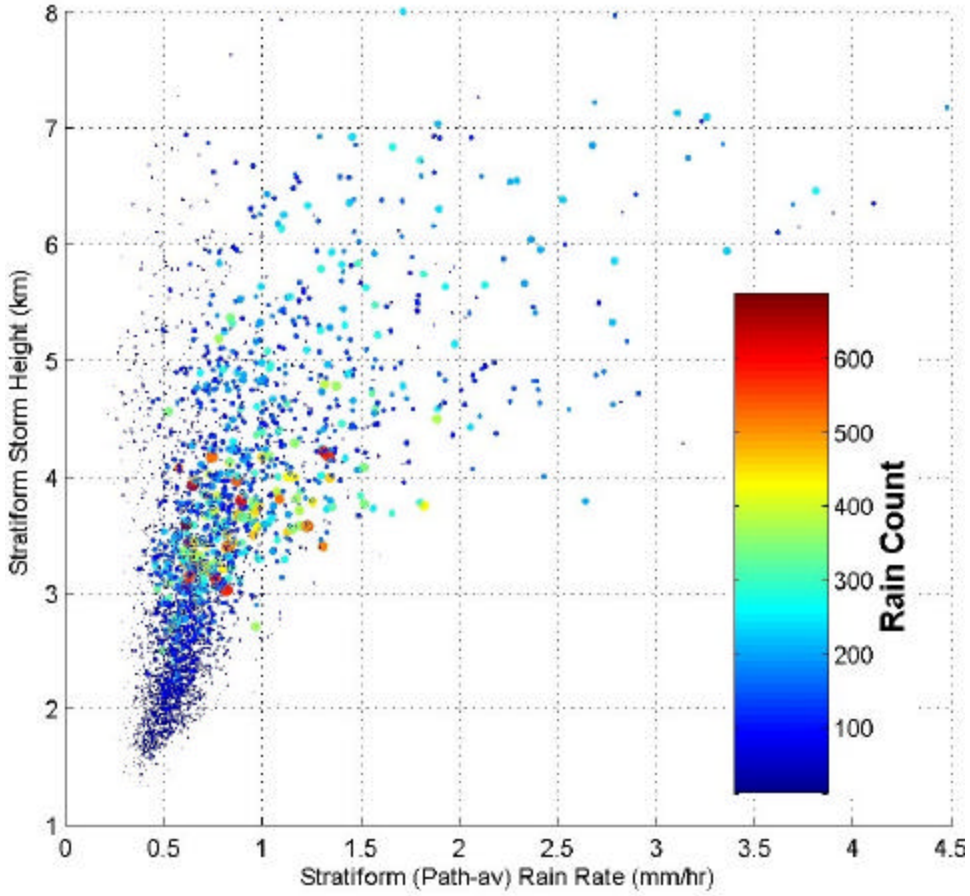


Figure 3-12 Stratiform Atlantic northern region rain-rate vs. rain-height scatter. The rain count data is used to weight and color-code the scatter points.

3.3.3 Log (RR) vs. RH Regression

In this regression scheme, we attempted to fit the dual trend present in rain-rate vs. rain-height scatter by using the logarithm of rain rate in the linear regression instead of rain rate itself. Figure 3-13 shows the log (rain rate) vs. rain-height scatter of Atlantic north region in January 1998. This scatter is fit with linear and orthogonal

single regression lines. This regression schemes failed to fit many convective scatter cases, producing very high standard error of estimate.

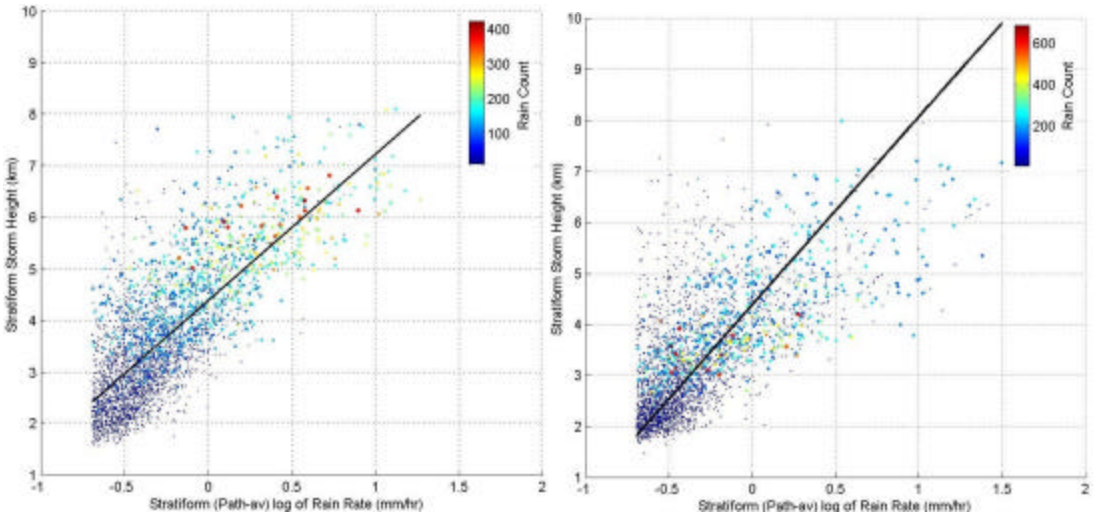


Figure 3-13 Log (rain rate) vs. rain-height scatter of Atlantic north region in January 1998 with two kinds of regression fits (a) Linear regression (b) Orthogonal regression.

3.3.4 Combined Linear – RR vs. RH (for lower RR) and RH vs. RR (for higher RR) Regression

In bilinear regression, the two lines tend to influence each other, because of the additional constraints applied to make the lines to meet at a point. Figure 3-14 shows two different bilinear regression fits. The first uses rain rate as an independent variable and rain height as a dependent variable (RR vs. RH regression fit shown in blue dotted line). The second does the opposite by using RH as the independent variable (RH vs. RR regression fit shown as red dotted line). From the figure, the influence of the upper regression line on the lower line for RR vs. RH bilinear regression scheme (blue dotted line) can be easily identified. The figure also shows a

linear RR vs. RH regression for rain rate less than 1.5 mm/hr (blue solid line) and linear RH vs. RR regression fit based on data for rain rate greater than 1.5 mm/hr (red solid line).

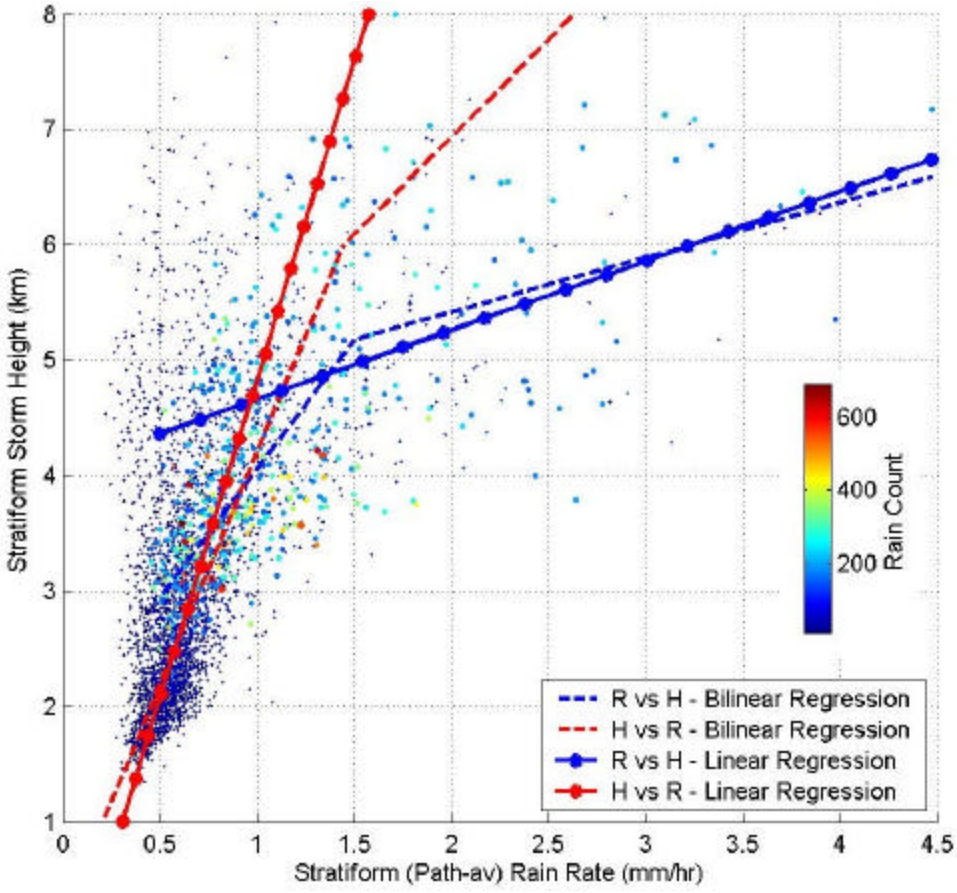


Figure 3-14 Comparison of Bilinear RR vs. RH (Blue dashed), Bilinear RH vs. RR (Red dashed) and Linear RR vs. RH (Blue Solid) and Linear RH vs. RR (Red Solid) regression fit in Stratiform Atlantic northern region scatter for January 1998 data.

Figure 3-14 shows the two linear regression lines extended to all rain-rate values. The linear RR vs. RH regression line and linear RH vs. RR regression line seem to fit the data for lower and higher rain-rate sections respectively. These linear regression lines forms the 'Combined Linear – RR vs. RH (for lower RR) and RH vs. RR (for

higher RR)' regression scheme. The break points for 'combined linear' regression fits are determined by extending the regression lines. A rain-rate threshold (1.5 mm/hr for this case) is selected to separate the data into lower and higher rain-rate sections, which are regressed with two different linear regression fits. The concept of 'Combined Linear' regression is shown in Figure 3-15.

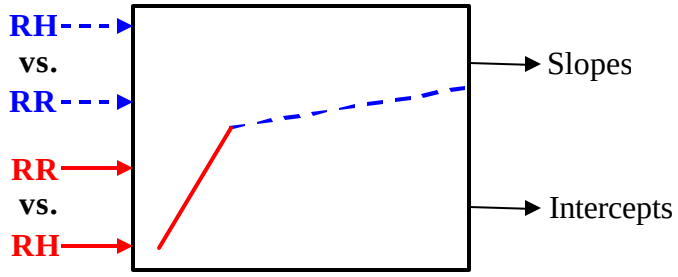


Figure 3-15 Conceptual figure of 'Combined Linear – RR vs. RH (for lower RR) and RH vs. RR (for higher RR)' regression scheme

The combined linear regression schemes always have sharp transitions between the lower and higher regression lines. The nature of estimated rain height, which is obtained from rain rate using the combined linear regression, is affected by this sharp transition. Figure 3-16 compares the histogram of measured rain height and the estimated rain height obtained from TRMM rain-rate data for stratiform Atlantic north region in January 1998. The sharp transition in regression lines shows as a cluster of estimated rain-height values near 5 km in the figure. Another drawback of the combined linear regression scheme is the failure of the RR vs. RH regression line to capture the almost logarithmic trend exhibited by the RR vs. RH scatter for lower rain-rate sections.

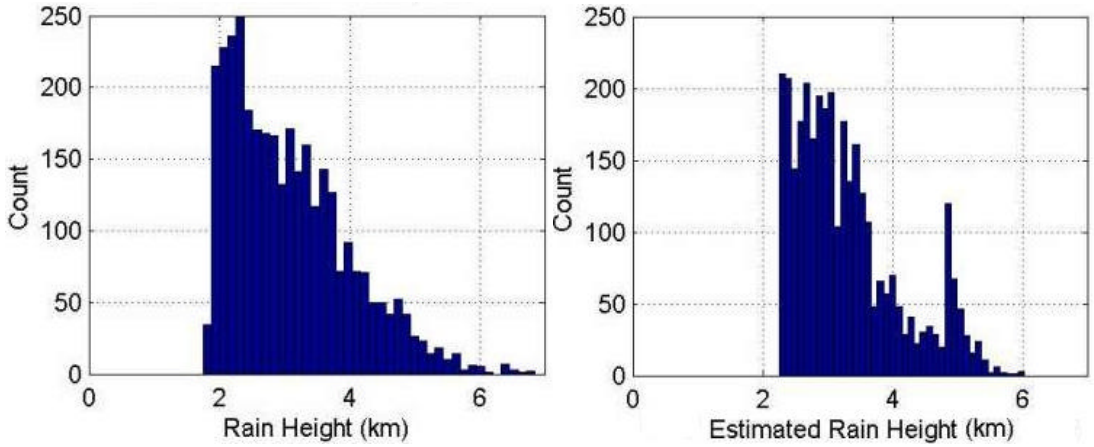


Figure 3-16 Histogram of measured rain height (a) and estimated rain height (b) from rain rate using combined linear regression scheme

3.3.5 Log-Linear Combined – Log (RR) vs. RH (for lower RR) and RR vs. RH (for higher RR)

The 'Log-Linear Combined' regression scheme is a modification to the 'Combined Linear' regression schemes discussed in the previous section. In the lower rain-rate section, a straight-line regression is made between log (rain rate) and rain height to capture the logarithmic trend. The higher rain-rate scatter is fit with a linear regression between rain rate and rain height. Figure 3-17 shows the conceptual figure of 'Log-Linear Combined' regression scheme.

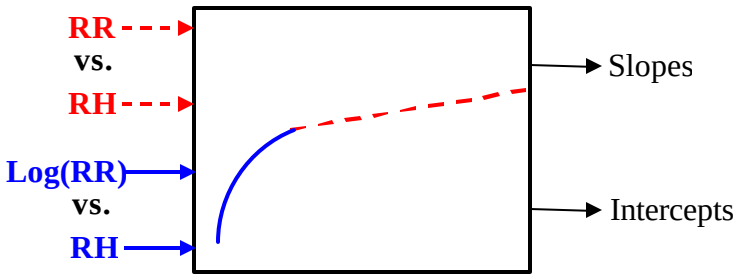
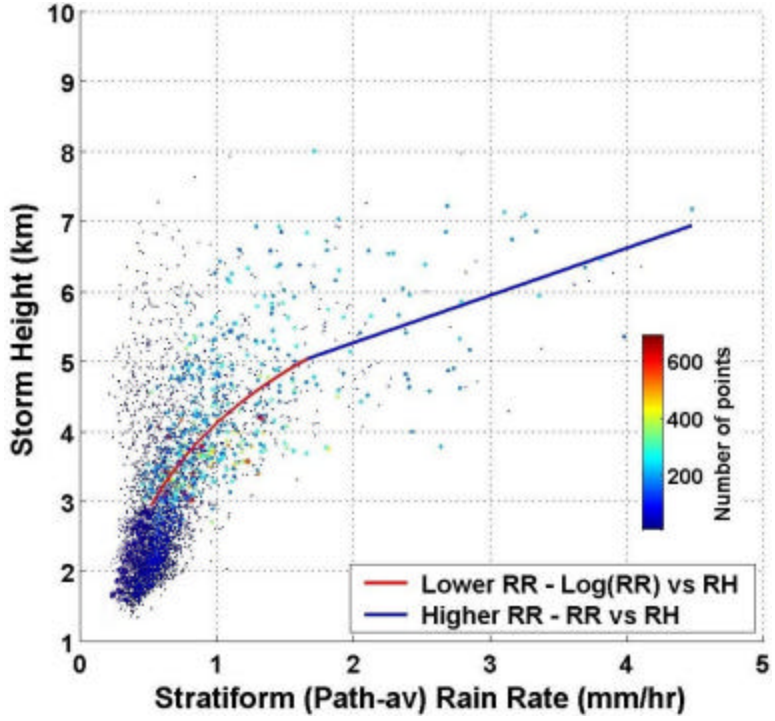
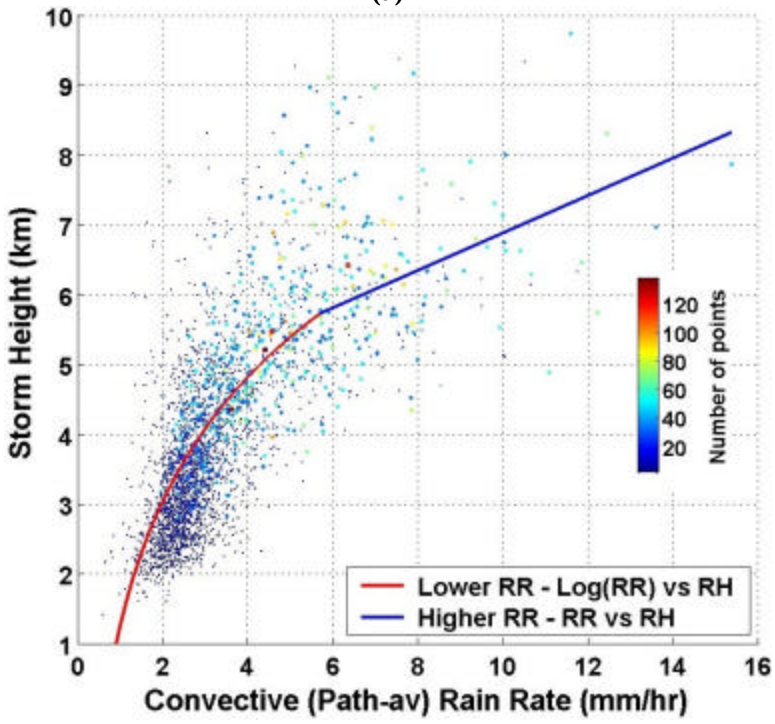


Figure 3-17 Conceptual figure of 'Log-Linear Combined' – log (RR) vs. RH (for lower RR) and RR vs. RH (for higher RR) regression scheme

The 'Log-Linear Combined' regression was selected as the best scheme for developing the final model, which relates rain rate and rain height over all seasons, regions and rain type. This scheme satisfies the criteria for a regression scheme enumerated in the section 2.3. Figure 3-18 shows the 'Log-Linear Combined' regression fit over the Atlantic north region for stratiform and convective rains. The distribution of the estimated rain height using 'Log-Linear Combined' regression scheme is comparable with the actual rain-height distributions.



(a)



(b)

Figure 3-18 'Log-Linear Combined' regression fit for RR vs. RH scatter of Atlantic northern region in January 1998 (a) for stratiform rain and (b) for convective rain.

3.4 Statistical Significance

Different statistical measures are available to evaluate the degree of confidence that the true RR vs. RH relationship is close to the regression-based relationship.

3.4.1 Standard Error and Goodness of fit

The regression scheme seeks to minimize the sum of the squared errors of prediction. The square root of the average squared error of prediction is used as a measure of the accuracy of prediction. This measure is called the standard error of the estimate and is designated as s_{est} [56]. The formula for standard error of estimate is:

$$s_{est} = \sqrt{\frac{\sum (RH - RH_{est})^2}{N}} \qquad \textbf{(3.1)}$$

where RH is the measured rain-height value from TRMM and RH_{est} is the estimated rain height using the regression scheme. N is the total number of (RR, RH) points. Another common statistic associated with regression analysis is the goodness of fit parameter (R^2). Goodness of fit is equal to one minus the ratio of the sum of squared estimated errors to the sum of squared deviations about the mean of the dependent variable [57]. The formula for goodness of fit is

$$R^2 = 1 - \frac{\sum (RH - RH_{est})^2}{\sum (RH - m_{RH})^2} \qquad \textbf{(3.2)}$$

where m_{RH} is the mean of the TRMM rain-height data.

The sum of squared deviations about its mean is a measure of the total variation of the dependent variable. The sum of squared deviations about the regression line is a measure of the extent to which the regression fails to explain the dependent variable. Hence, the R^2 statistic is a measure of the extent to which the total variation of the dependent variable is explained by the regression scheme. High values of R^2 suggest that the model is good for predictive and forecasting purposes. Table 3-1 shows the standard error of estimate and goodness of fit statistics for important regression schemes investigated in our study. The 'Log-Linear Combined' regression scheme shows low standard error of estimate of 0.87 km for Atlantic northern region in January 1998. Similar results were obtained for other regions and seasons. The goodness of fit parameter is also tabulated for various regression schemes.

Regression schemes	**Standard Error of Estimate (Km)**	**Goodness of fit R^2**
Linear Regression	2.5	0.40
Bilinear Regression	1.3	0.52
Weighted Bilinear Regression	1.2	0.55
Log (RR) vs. RH regression	1.4	0.60
'Combined Linear' regression	0.93	0.89
'Log-Linear Combined' regression	**0.87**	**0.93**

Table 3-1 Standard Error of estimate and goodness of fit parameters of various regression schemes analyzed for RR vs. RH scatter for Atlantic northern region in January 1998

To visualize the accuracy of our regression fit over all rain-rate values, the following exercise was carried out for the Atlantic northern hemisphere rain data (January 1998). The data were segmented into five groups based on rain rate. Figure 3-19 (a) shows the segments in the RR vs. RH scatter. In each segment, standard error of

height estimate are calculated and plotted along the mean rain rate value in each segment. Figure 3-19 (b) shows the standard error in rain height vs. rain rate plot along with the histograms of the estimated rain height in each segment. This figure shows that the standard error in rain-height estimate is consistently low (< 1 km) for all rain-rate values. Hence the 'Log-Linear regression model can be used to estimate rain height for all values of rain rate.

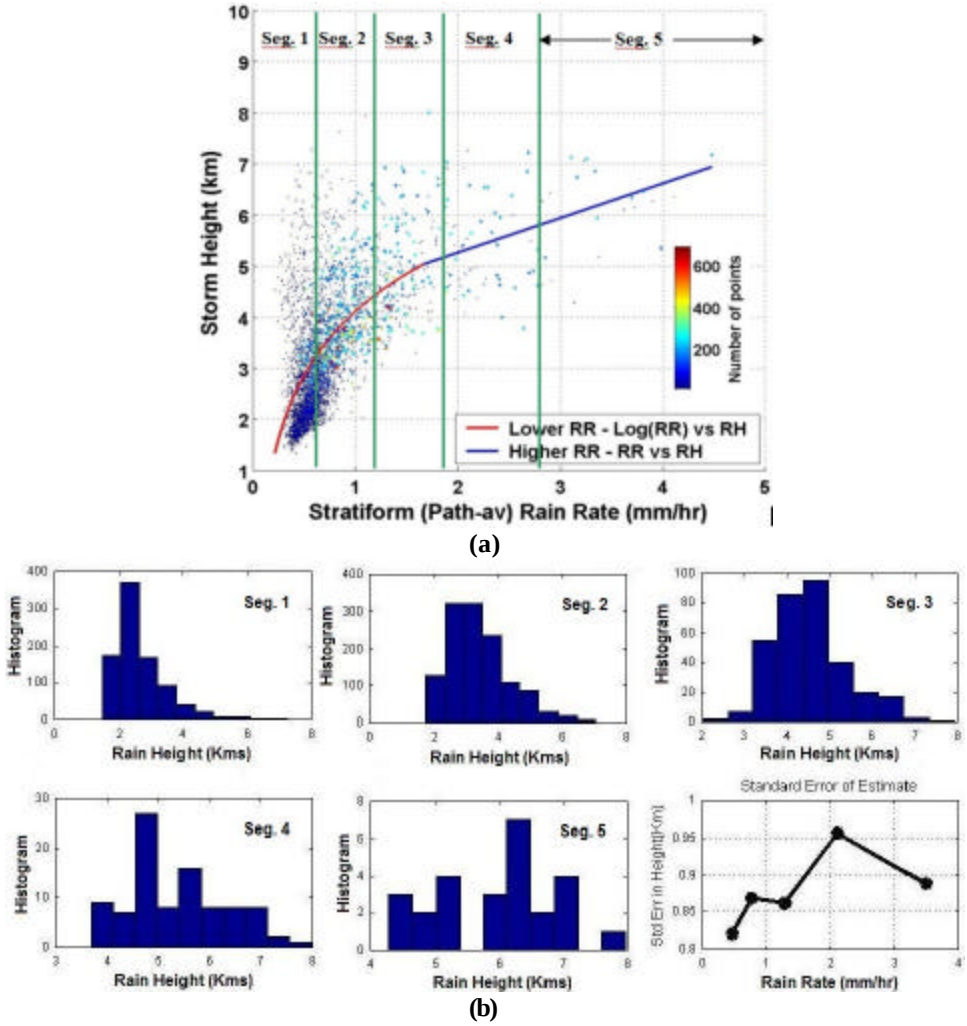


Figure 3-19 Segmented standard error in height (km). (a) Segments in the RR vs. RH scatter (b) Histogram of estimated rain-height data and the standard error in height in each segment vs. mean rain-rate plot.

3.4.2 T-test

Hypothesis testing is a procedure for determining if an assertion about a characteristic of a population is reasonable [56]. *Paired t-test* is usually used to compare paired sets of *X* and *Y* data and determine if they differ from each other in a significant way [57]. In this section, the paired *t-test* is used to find the closeness of the original TRMM rain height to that of the estimated rain height obtained from the regression model. The null hypothesis or the original assertion in this case is that the means of the two rain height (original and estimated) data are equal. The paired t-test determines the probability that a given hypothesis is true. The alternative hypothesis is that the means of the two data sets are not equal.

The *t-test* takes an additional input parameter known as the *significance level* (α), which is defined as the degree of certainty required to reject the null hypothesis in favor of the alternative. After analyzing the data based on t-statistics, the *t-test* produces a result to either "reject the null hypothesis" or "do not reject the null hypothesis". One should note that "do not reject the null hypothesis" result cannot be concluded that the null hypothesis is true. It means that one does not have sufficient evidence to reject the null hypothesis. *T-test* also produces an output known as *p-value*. The *p-value* is the probability of observing the given sample result under the assumption that the null hypothesis is true [57]. Small values of *p* in comparison to the significance level (α), cast doubt on the validity of the null hypothesis.

The two-sided t-test was performed on the actual rain-height and predicted rain-height values for stratiform rain data over Indian southern region. A significance level of $\alpha = 0.05$ was chosen for the test. Figure 3-20 shows the stratiform rain-rate and height maps over Indian southern hemisphere. The rain-rate data from this region is used to predict rain-height values with the aid of the 'Log-Linear Combined' regression scheme. The predicted stratiform rain-height map is also shown in Figure 3-20. The two-sided t-test, which compares the actual and predicted rain-height, does not have sufficient evidence to reject the null hypothesis (Note the null hypothesis is the assumption that the means of actual and predicted rain-height data are equal). The p-value for this t-test over Indian southern hemisphere is given as 0.9284. Higher p-value implies higher probability for the null hypothesis to be true.

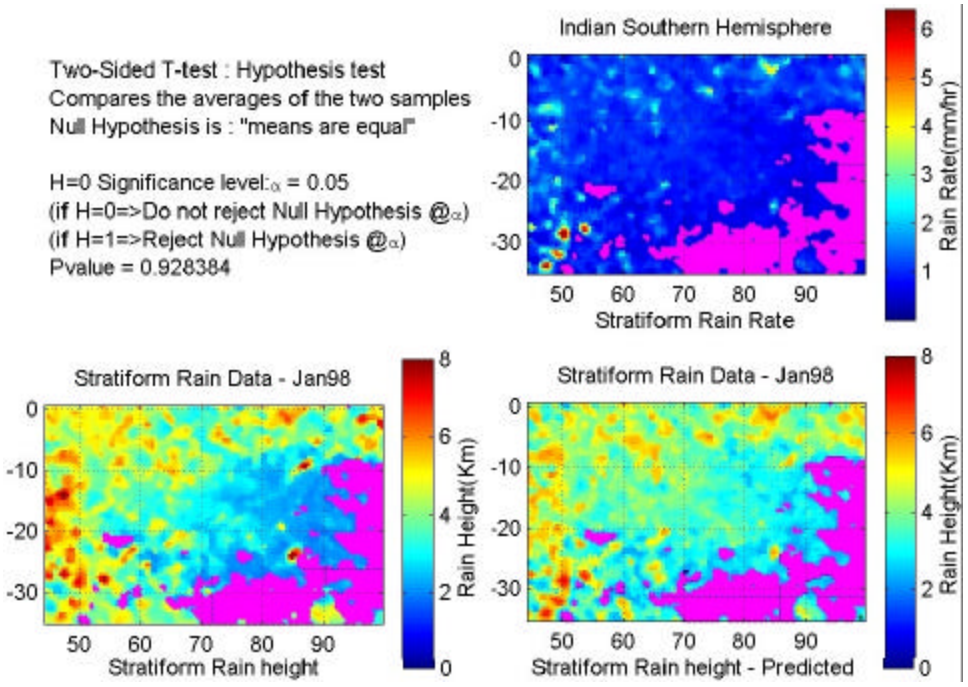


Figure 3-20 Two-sided T-test result based on stratiform Indian southern hemisphere region data. The null points in the data are shown in pink color

The actual rain-rate, rain-height and predicted rain-height maps for convective Indian southern hemisphere data are shown in Figure 3-21. The two-sided *t-test* over actual and predicted rain-height for convective rain data show result against the rejection of null hypothesis. The *p-value* for this convective rain case is 0.3772. If the *t-test* is repeated for the same convective rain case with a significance level of 0.1, then the *p-value* shows as 0.7425. The low p-value for the 0.05 significance level is mainly due by the few wide-spread convective points observed for higher rain-rates. In general comparison to the stratiform rain case, a lower p-value suggests lesser confidence in rain-height prediction from convective rain-rates using the regression model.

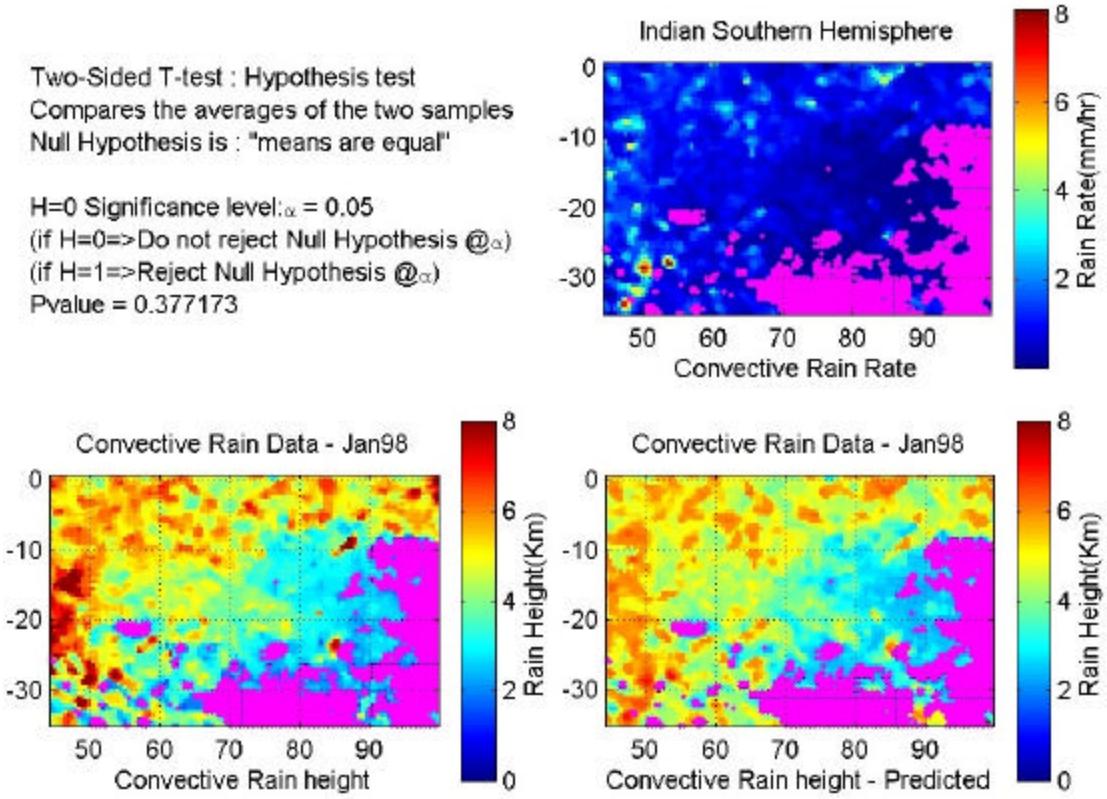


Figure 3-21 Two-sided T-test result based on convective Indian southern hemisphere region data

3.4.3 Comparison between Region and Hemispherical scatter based rain height prediction

In this section, the rain height is predicted from rain rate using the *Log-Linear Combined* regression lines over different regions. This predicted rain-height is compared with the measured TRMM rain-height from the same region. Similar to the previous section, we used the t-test between these rain-height values. Table 3-2 shows the p-value results of the hypothesis t-test with a significance level of 0.05 for stratiform rain data. The stratiform rain data produce very high p-values for all regions. The table also shows higher confidence in the northern and southern hemisphere scatter based rain-height prediction compared to regional data based rain-height prediction. This result is very significant, as one can omit the regional data based analysis for the rest of the statistical model development.

Region	p-value
Indian north	0.80
Atlantic north	0.91
Pacific northeast	0.78
Pacific northwest	0.75
Northern Hemisphere	**0.90**
Indian south	0.93
Atlantic south	0.92
Pacific southeast	0.89
Pacific southwest	0.85
Southern Hemisphere	**0.94**

Table 3-2 T-test results comparing the measured and predicted rain-height values over different regions for stratiform rain case with a significance level of 0.05.

Table 3-3 shows the p-value results of the t-test with a significance level of 0.1 for convective rain data. For the convective case with use of the entire-hemisphere data produced higher p-values than use of regional prediction. Rain data from January 1998 were used in all of the hypothesis tests.

Region	p-value
Indian north	0.73
Atlantic north	0.75
Pacific northeast	0.77
Pacific northwest	0.69
Northern Hemisphere	**0.75**
Indian south	0.74
Atlantic south	0.72
Pacific southeast	0.69
Pacific southwest	0.70
Southern Hemisphere	**0.79**

Table 3-3 T-test results comparing the measured and predicted rain-height values over different regions for convective rain case with a significance level of 0.1.

4 Regression-based Statistical Model

Chapter 3 describes various regression schemes to investigate the relationship between rain-rate and rain-height. The *Log-Linear Combined* regression, which satisfies the criteria for estimating rain-height from rain-rate, was selected as the universal regression method; valid over various regions, seasons and rain types. We started with *Log-Linear Combined* regression fits for specific months and regions. Generalization of trends exhibited by the regression parameters is required to extend this regression-based relationship to the future.

This chapter analyzes the seasonal trends in the slope and intercept statistics of the Log-Linear Combined regression scheme for the monthly rain data extending over four years (1998 – 2001). We used Fourier analysis and synthesis to capture the seasonal trends in slope and intercept statistics and to predict them for any future season. We developed a generalized statistical model to estimate rain height given rain rate in any region and season. Rain-height estimation using a statistical model based on data for the entire northern and southern hemispheres is as effective as models based on regional data. The final output of this thesis is a *regression-based*

statistical model, which is used to estimate rain-height for a given rain-rate, month, region and rain type.

4.1 Slope and Intercept Statistics

The slopes and intercepts of the *Log-Linear Combined* regression lines were tabulated for all months over 4 years (1998-2001). Table 4-1 shows the structure of the slope-and-intercept statistics table. We generated similar tables for all eight oceanic regions and hemispherical sections. Using these tables, rain height can be estimated given a rain rate, region and a month in these four years.

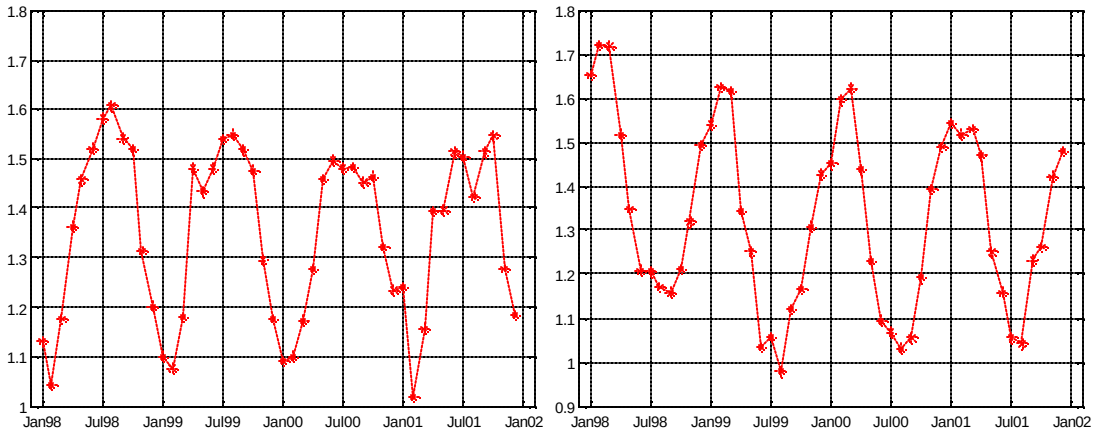
Statistics	Jan98	Feb98	Mar98	Apr98	…	Dec01	Jan02
Slope (m1) Log(RR) vs. RH					…		
Intercept (c1) Log(RR) vs. RH					…		
Slope (m2) RR vs. RH					…		
Intercept (c2) RR vs. RH					…		

Table 4-1 Format of the slope and intercept statistics table

4.1.1 Seasonal trend

Figure 4-1 is a plot of the slope statistics of Log (RR) vs. RH regression line developed for lower rain-rate sections (plot of row 1 of the slope and intercept statistics table). This figure is for convective rain data with all regions in each hemisphere combined. A sinusoidal seasonal trend is apparent. Similar seasonal

trends were found in slope and intercept statistics over all regional and hemispherical sections. Figure 4-1 shows that the northern and southern hemisphere slope curves are approximately 180° out of phase. Note a minimum occurs in February 1998 for the northern hemisphere, whereas a maximum occurs in the same month for the southern hemisphere. We need to model the seasonal trends to predict the slopes and intercepts of the *Log-Linear Combined* regression method for any future month.



(a) **(b)**

Figure 4-1 Seasonal trend in values of slopes based on *Log-Linear Combined* regression fits. The slopes of the lower rain-rate bin regression lines (Log (RR) vs. RH regression) are shown for (a) Convective entire southern Hemisphere (b) Convective entire northern hemisphere

4.1.1 Mean value substitution

A simple method to generalize the seasonal trend of the slope and intercept statistics is to substitute with their mean values over all seasons. To validate this approach, we attempted to estimate the 95% confidence interval of each slope and intercept estimates. Wide confidence intervals will allow the use of mean slope or intercept

values for all regions. The slope confidence interval calculations require an estimate of the standard error of slope, which is calculated as shown below,

$$s_{slope} = \sqrt{\frac{s_{est}}{\sum RR - m_{RR}}} \qquad \textbf{(4.1)}$$

where *RR* designates the measured rain-rate values from TRMM, m_{RR} is the mean of the TRMM rain-rate data and $s_{est.}$ is the standard error of the estimates shown in equation (3.1).

The confidence interval (*C.I.*) for the slopes can be estimated as follows,

$$C.I. = m \pm s_{slope}(t) \qquad \textbf{(4.2)}$$

where m is the estimated slope of the regression line, s_{slope} is the standard error in slope estimates and *t* is the t-value for a two-tailed test given a desired α-level (95% or α = 0.05). The confidence interval provides a range in which the 95% of true population slope is likely to fall. Figure 4-2 shows the two slope statistics of the *Log-Linear Combined* regression scheme for the stratiform southern hemisphere data. Each slope estimate was plotted with its 95% confidence interval. The confidence interval is narrow for all slope estimates in this example. Similar results were found for slope and intercept estimates for all regions and rain types. Hence it is reasonable to analyze the sinusoidal trend exhibited by the slope and intercept statistics.

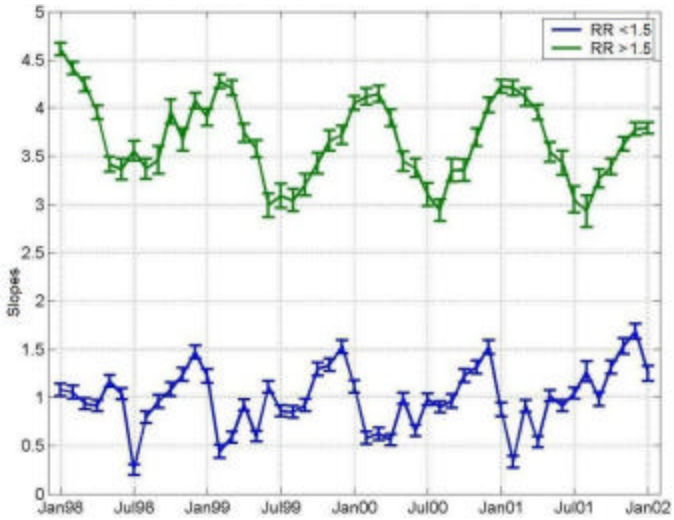


Figure 4-2 Value of slopes based on *Log-Linear Combined* regression scheme for stratiform southern hemisphere data shown with the 95% confidence interval. The blue line shows the intercept of the Log(RR) vs. RH regression fits (for lower RR bin). The green line shows the intercepts of the RR vs. RH regression fits (for higher RR bin).

4.2 Fourier analysis

Fourier analysis is a method of describing periodic waveforms in terms of trigonometric functions [58]. The slope and intercept statistics show sinusoidal trends in time with approximately one year as the period of each cycle. The wave function (usually amplitude, frequency, or phase versus time) can be expressed as a Fourier series, uniquely defined by the Fourier coefficients. If these coefficients are represented by *a0, a1, a2, a3... aN, ...* and *b1, b2, b3... bN ...* , then the Fourier series *x(t)* is given as:

$$x(t) = \frac{a_0}{2} + \sum_{k=1}^{N} \left[a_k \cos(k w_0 t) + b_k \sin(k w_0 t) \right] \qquad \textbf{(4.3)}$$

where *t* denotes time and w_0 is the fundamental angular frequency.

The spectrum of a discrete sequence can be obtained using Fast-Fourier Transformation (FFT). Figure 4-3 shows the FFT spectrum magnitude and phase plot of the slope curves for the northern-hemisphere convective-rain data. The DC, fundamental and harmonic components are shown in red. Equation (4.4) shows the magnitude and phase of the selected Fourier coefficients.

$$\{a_0, (a_{fun}, f_{fun}), (a_{har2}, f_{har2}), (a_{har3}, f_{har3}), ..., (a_{harN}, f_{harN})\} \qquad \textbf{(4.4)}$$

where a_k represents the magnitude and f_k represents the phase of the Fourier coefficients.

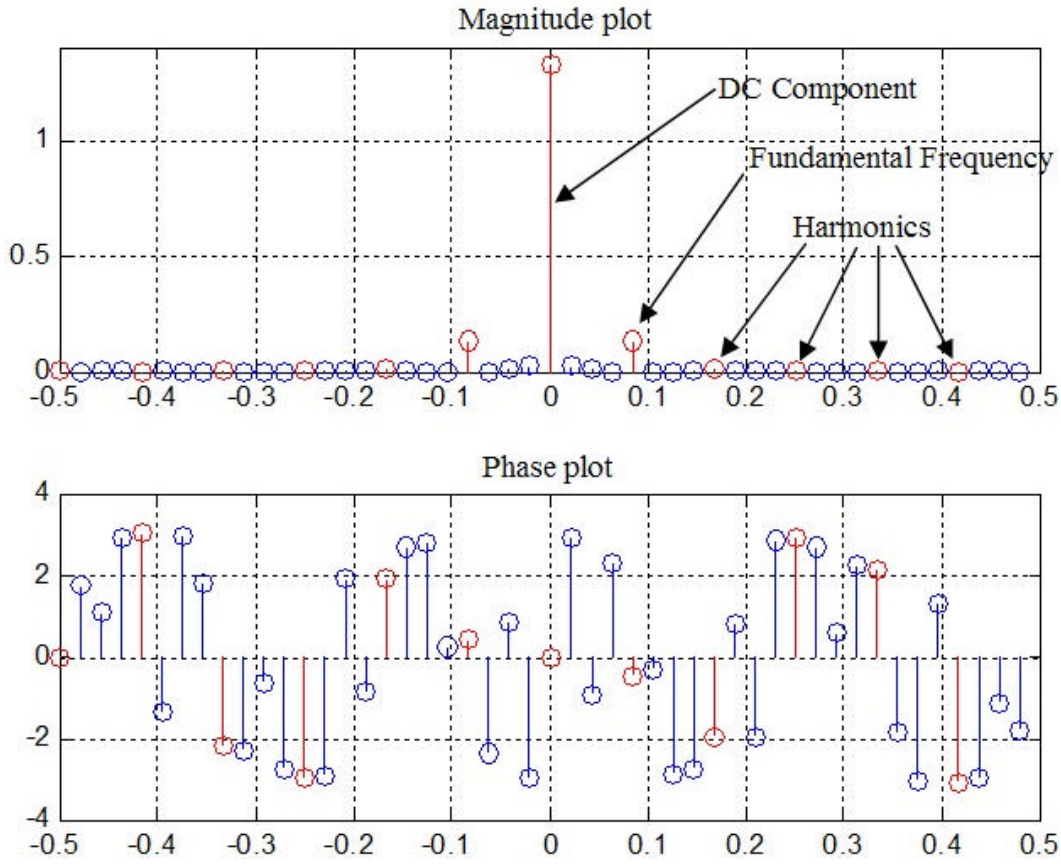


Figure 4-3 FFT spectrum magnitude and phase plot of the slopes statistics curve for convective northern hemisphere data. The DC, fundamental and harmonic components are marked.

4.3 Fourier Synthesis

Fourier synthesis is a method of reconstructing a signal with a specific, desired periodic waveform [58]. After analyzing various slope and intercept statistics, we concluded that the use of DC, fundamental and two harmonics (N=3) is sufficient to regenerate the basic sinusoidal trend, as shown in (4.5):

$$\hat{x}(t) = a_0 + a_{fun}\cos(2\pi f_{fun} t + f_{fun}) + a_{har2}\cos(2\pi f_{har2} t + f_{har2}) + a_{har3}\cos(2\pi f_{har3} t + f_{har3}) \quad \textbf{(4.5)}$$

where f_{fun} is the fundamental frequency and f_{har2}, f_{har3} are harmonic frequencies.

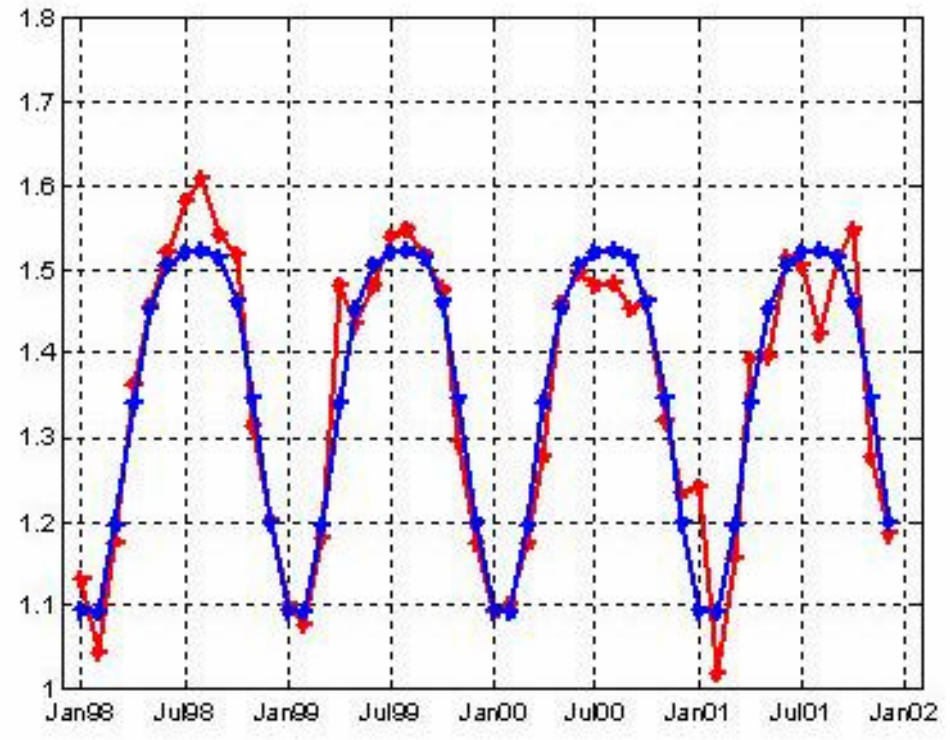


Figure 4-4 Values of slope based on Log(RR) vs. RH regression fits to the measured northern-hemisphere convective rain data. The red line shows the measured data, and the blue line is the Fourier-synthesized (regenerated) curve from the fundamental and two harmonic components.

Figure 4-4 compares the lower rain-rate bin regression line for convective rain in the northern hemisphere with the s3elected Fourier-series fit. The red line represents the measured slope statistics curve $x(t)$. Magnitude and phase of the Fourier coefficients were obtained by analysis of $x(t)$. The Fourier synthesized curve is the blue line in the figure. Using these Fourier coefficients, the slopes and intercepts of *Log-Linear Combined* regression scheme can be predicted for any given month in the future.

An example of the intercept statistics for stratiform data from the southern hemisphere is in Figure 4-5. The shaded region corresponds to the 95% confidence bound. The intercept statistics were Fourier analyzed and appear in the figure as blue (lower RR bin) and green (higher RR bin) lines. Most of the synthesized curves fall within the 95% confidence bound of the measured intercept values. Hence this method allows good prediction of slopes and intercepts.

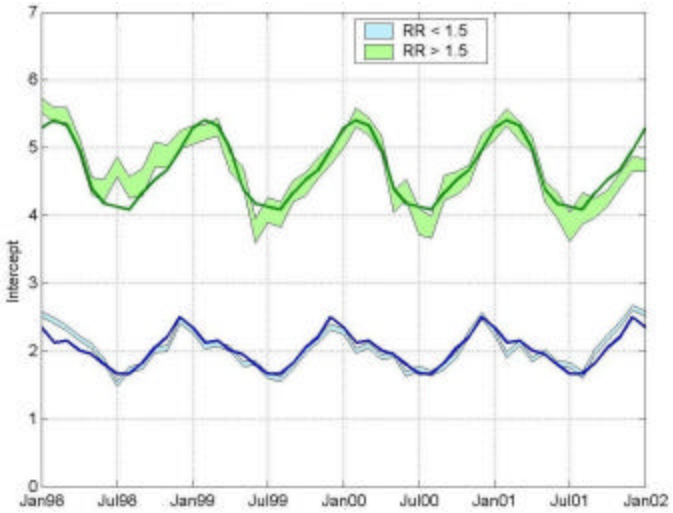


Figure 4-5 Intercept curves of the *Log-Linear Combined* regression scheme for stratiform rain in the southern hemisphere. The blue and green lines show the synthesized intercept values from the Log (RR) vs. RH regression and RR vs. RH regression respectively. The shaded region represents the 95% confidence interval about the measured values.

4.4 Table of Fourier coefficients

Final Table	DC	Fundamental Frequency		Second Harmonic		Third Harmonic	
		Mag	Phase	Mag	Phase	Mag	Phase
Slope (m1) Log(RR) vs. RH							
Intercept (c1) Log(RR) vs. RH							
Slope (m2) RR vs. RH							
Intercept (c2) RR vs. RH							

Table 4-2 Format of the Fourier Coefficients Table

Fourier coefficients generated from the Fourier analysis of the slope and intercept statistics are tabulated in the format shown in Table 4-2. The Fourier-coefficient tables can be developed for the regional rain data or the data for the entire hemisphere. The disadvantage of using the region-data based Fourier coefficient tables is that one has to maintain a separate table for each oceanic region. Section 3.4.3 compares the estimated rain height based on regression fits to the regional RR vs. RH measurements and those for entire hemispheres. The difference between these rain-height estimates was found to be statistically insignificant, which makes the region-based Fourier-coefficient tables unnecessary.

Since the hemispherical tables are used for all of our subsequent analysis, only four Fourier-coefficient tables are required in the *regression based statistical model*. They are:

1. Stratiform Northern Hemisphere
2. Stratiform Southern Hemisphere
3. Convective Northern Hemisphere
4. Convective Southern Hemisphere

4.5 Rain-height estimation using Regression-based Statistical Model

The *regression-based statistical model* is a complete algorithm for estimating rain height given a rain rate over any hemisphere, season and rain type. Figure 4-6 shows the flowchart of the final algorithm. In the flowchart, the solid arrows indicate the logical flow and dashed arrows show the inputs to each block.

For a given rain type and hemisphere, a Fourier coefficient table is selected. Fourier synthesis provides slopes and intercepts (m_1, m_2, c_1, c_2) for a given month for the RH vs. RR curves. The slope and intercept values of the two regression lines are used to find the breakpoint rain-rate (BP). Breakpoint rain rate splits the RR vs. RH scatter into two rain-rate sections where the regression lines meet. The input rain-rate from

AMSR is classified into one of the two rain-rate sections. The formula used to estimate rain-height is based on this classification.

For the lower rain-rate section, the rain-height is estimated as follows:

$$RH_{est} = m_1 \times \log(RR) + c_1 \qquad for\ RR < BP \qquad \textbf{(4.6)}$$

Similarly for the higher rain-rate section, the rain height estimation formula is:

$$RH_{est} = m_2 \times RR + c_2 \qquad for\ RR > BP \qquad \textbf{(4.7)}$$

For each measured rain-rate from AMSR, rain height is estimated using the *regression-based statistical model* as shown in Figure 4-6 and used in the *rain-effect correction algorithm*. The next chapter validates this *regression-based statistical model* in estimating rain heights.

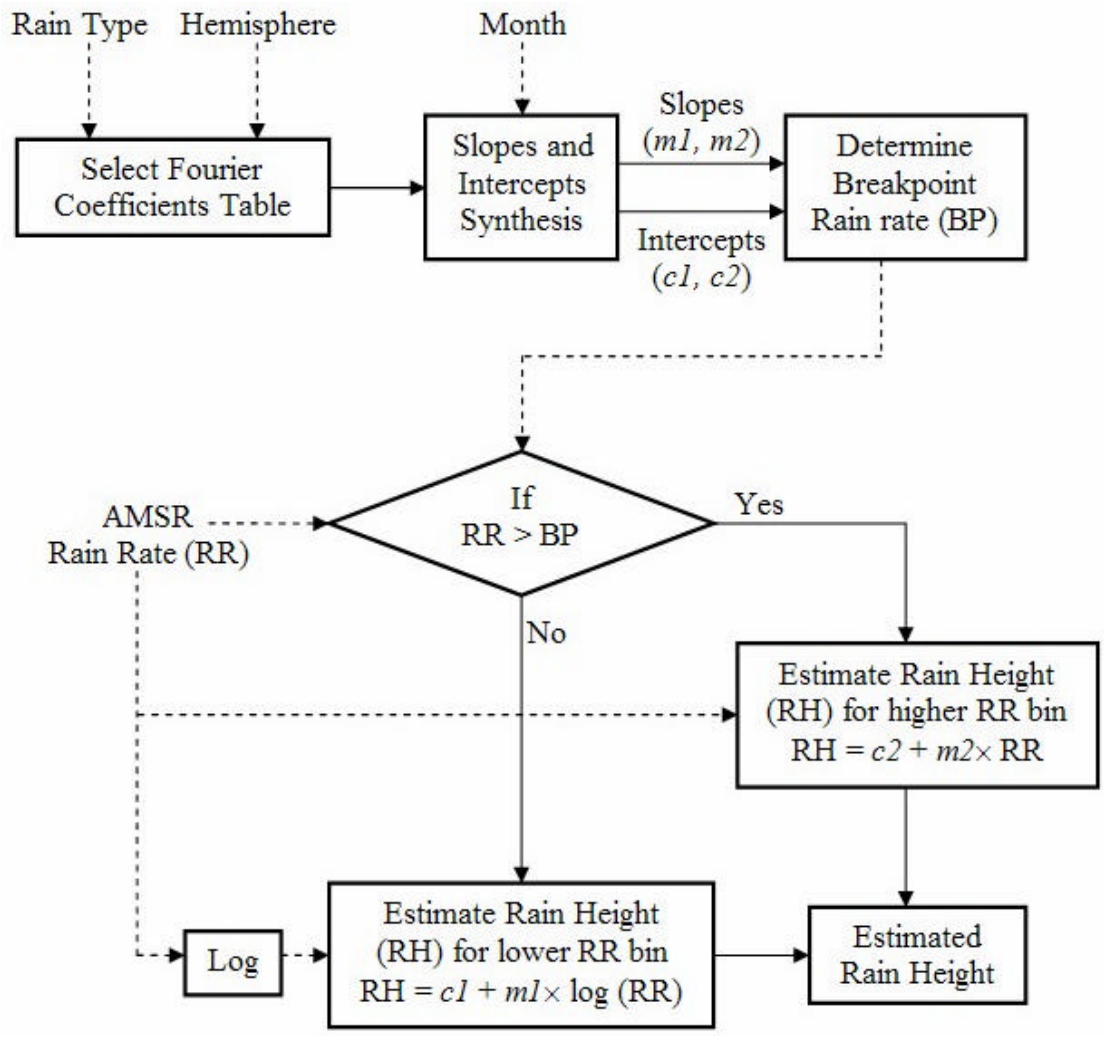


Figure 4-6 Flowchart to estimate rain-height using the *regression-based statistical model*

5 Validation of Statistical Model

Chapter 4 describes the *regression-based statistical model*, which is used to estimate rain height from rain rate for a given hemisphere, month and rain type. Rain-height is an important parameter in the *rain effect correction algorithm* discussed in section 2.3. This algorithm attempts to correct for rain attenuation and backscatter effects on SeaWinds signal. It is important to validate the *regression-based statistical model* if we are to use it in the *rain effect correction algorithm*.

5.1 Comparison of RH Estimates from Regional Regression lines and the Statistical model

Section 3.4 compares the rain-height estimates obtained from regional *Log-Linear Combined* regression fits and the measured TRMM rain-height values. Hypothesis tests revealed high confidence in these estimates. The *regression-based statistical model* uses the Fourier coefficient table and predicts the slopes and intercepts of the *Log-Linear Combined* regression fits. These predicted slopes and intercepts are used in the statistical model to estimate rain-height from rain rate.

In this section, we compare the rain-height estimates obtained from the *Log-Linear Combined* regression fits with the rain height estimates derived from the statistical model, using the stratiform North-Atlantic rain data for January 2002. Note that only the rain data from January 1998 to December 2001 were used in the development of the statistical model. Hence, the slopes and intercepts for January 2002 were predicted by the Fourier coefficient tables and the statistical model estimated rain height. A *Log-Linear Combined* regression fit of the stratiform data for the North-Atlantic region RR vs. RH scatter was also performed. The slopes and intercepts from these regression lines were used to obtain other rain-height estimates. This exercise compares both the rain-height estimates to allow visualizing the accuracy of the *regression-based statistical model* in rain-height estimation.

Figure 5-1 shows the distribution of the rain-height estimates obtained using two different methods. The black histogram is the distribution of the measured TRMM rain-height values in the North-Atlantic region for January 2002. The regression lines on the scatter plot shown in the Figure 5-1 also indicate the close correspondence of the statistical model and the region-based regression fits. The blue histogram represents the distribution of the rain-height estimates obtained using the regression lines for TRMM measurements over the entire northern hemisphere. The red histogram represents their distribution obtained using the statistical model. This comparison highlights the effectiveness of the statistical model.

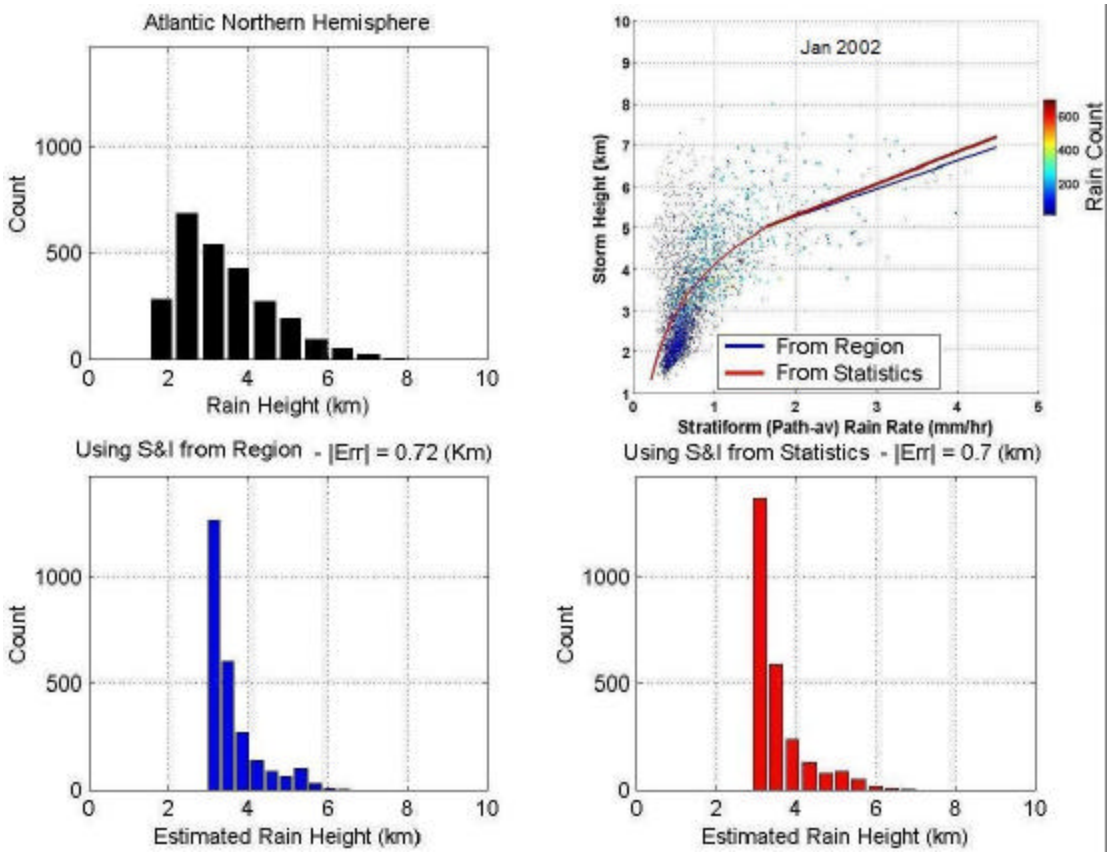


Figure 5-1 Comparison of distribution of rain-height estimates obtained from the *Log-Linear Combined* regression fits and from the statistical model.

From Figure 5-1, the closeness of estimated rain-height distributions obtained from two different methods is apparent. The Kolmogorov-Smirnov (ks-test) goodness-of-fit hypothesis test was used to compare these estimated rain-height distributions. In this test, the null hypothesis assumes that the cumulative distribution functions of both the rain-height estimates are the same. In the stratiform North-Atlantic case (January 2002), the ks-test suggests 'not to reject the null-hypothesis' and it provides a very high confidence in the result with a p-value of 0.92. Similar to the T-test comparison made in the section 3.4.2, the measured rain-height was compared with

the estimated rain-height values obtained with both the methods. The null hypothesis for the T-test assumes the rain-height populations to have the same mean. The T-test result on measured and estimated rain-height suggests 'not to reject the null hypotheses' with very high p-values as shown in Table 5-1. Similar results were obtained in the hypothesis testing using rain data from other regions and rain types.

Hypothesis T-test Comparison	p-value
Measured rain-height and the estimated rain-height obtained from regression fits in the regional RR vs. RH scatter.	0.952
Measured rain-height and the rain-height estimates obtained from the statistical model.	0.945

Table 5-1 T-test results comparing the measured rain height and the estimated rain height

Another statistical parameter that can be used to compare the rain-height estimates obtained using the two methods is the mean absolute error between the measured and the estimated rain-height values in a region.

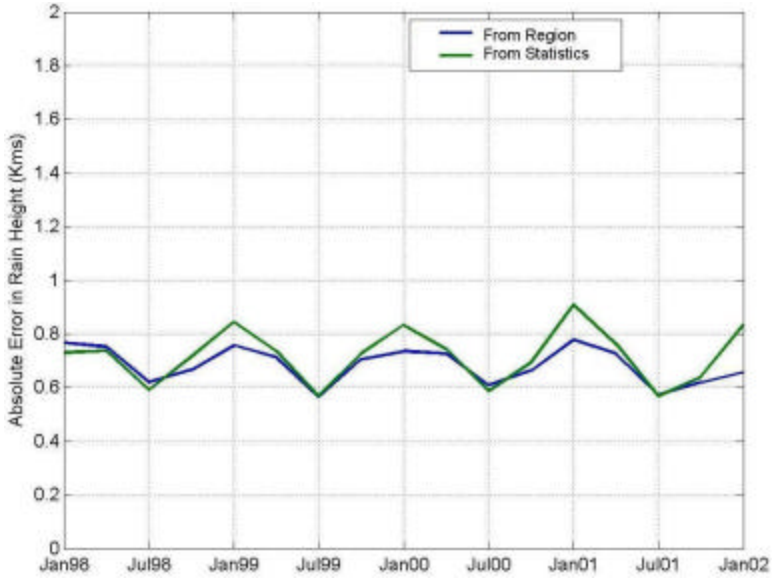


Figure 5-2 Comparison of mean absolute error in rain-height estimates obtained from the *Log-Linear Combined* regression fits and from the statistical model.

Figure 5-2 shows the mean absolute error in rain-height estimates for the stratiform North-Atlantic rain data in all quarters (months of January, April, July, October) over four years (1998 – 2001). The mean absolute errors in rain-height estimates are below 1 km for all seasons. The hypothesis testing and mean-absolute-error comparison show the accuracy of the *regression-based statistical model* in estimating rain-height from measured rain rates.

This estimated rain-height values are used in the *rain-effect correction algorithm* to correct for the rain attenuation and backscatter effects on the normalized radar backscattering coefficient (s_s^o) measured by the SeaWinds scatterometer. A simulation approach developed to find the impact of the estimated rain-height in the surface scattering coefficient correction process is discussed in the next section.

5.2 Simulation procedure

Rain-height estimates from the *regression-based statistical model* will be in error compared to the actual rain-height values. These errors are particularly great for higher rain rates due to the large spread of the data points around the regression lines. If these rain-height estimates are used in the *rain-effect correction algorithm*, they will affect the corrected surface scattering coefficient (s_s^o). Simulations were performed to determine the nature of these errors in the normalized radar cross-section coefficient.

The simulation approach can be divided into two parts. The first part is the 'forward simulation', which estimates the received scattering coefficient (s_r^o) as measured by the SeaWinds instrument. The estimation of received scattering coefficient (or the measured equivalent scattering coefficient as shown in equation (2.4)) is necessary, because the data from ADEOS II SeaWinds instrument were not available during the time of this study. The forward simulation models the rain effects on the SeaWinds received signal with the use of *measured TRMM rain-height* values. The rain heights obtained from TRMM are referred as *measured* heights ($h_{measured}$) in this simulation.

The second part, known as the 'reverse simulation', tries to correct rain effects to obtain an estimated surface scattering coefficient ($s_{s-corrected}^o$). The estimated rain-height from the *regression-based statistical model* is used in the reverse simulation to model the effects of rain. These rain-height estimates are denoted as *estimated* heights ($h_{estimated}$) in this simulation. Modeling of rain effects is similar in the forward and the reverse simulations except for the use of *measured* and *estimated* rain heights. Comparing the corrected surface scattering coefficient from the simulations and the actual surface scattering coefficient shows the impact of rain-height estimates in the *rain-effect correction algorithm*. A detailed description of this simulation approach for stratiform fully filled rain cells is in [27]. For fully filled rain cells, one assumes that uniform stratiform rain is present throughout the SeaWinds footprint.

5.2.1 Forward Simulation

The forward simulation estimates the received normalized radar cross-section using

$$\tilde{s}_r^o = \tilde{s}_s^o \times \exp(-2 \times K_{er} \times SL_{measured}) + h \times h_{measured} \times x_{measured} \qquad \textbf{(5.1)}$$

Equation (5.1) is similar to equation (2.4). To estimate the received scattering coefficient, we must estimate the other parameters of the equation. For the simulation, the surface values $\tilde{s}_s^o$ are modeled by the standard QuikSCAT geophysical model function (GH table). The model function provides surface scattering coefficient $\tilde{s}_s^o$ for a given wind speed, wind direction and polarization.

The attenuation constant K_{er} was estimated for the simulations using empirical K-R relations. The final relation relating attenuation and rain rate was produced using the ITU tables and the linear and log interpolation formulas shown in [38]. The K-R relations used for horizontal and vertical polarization are:

$$\begin{aligned} &\text{For Horizontal Polarization:} \quad K_{er} = 0.0262\, R^{1.1858}\ dB/km \\ &\text{For Vertical Polarization:} \quad K_{er} = 0.0262\, R^{1.1644}\ dB/km \end{aligned} \qquad \textbf{(5.2)}$$

It must be noted that the values for the attenuation K_{er} in (5.2) are expressed in *dB/km* and need to be converted to *Np/m* for performing the simulations. The measured TRMM rain-height is used for $h_{measured}$ and is modified to get slant height $SL_{measured}$ values.

The volume backscatter coefficient is calculated as shown in (2.6). The reflectivity factor *Z* in (2.6) is estimated using empirical Z-R relations:

$$\text{For Stratiform Rain: } Z = 300\,R^{1.49}$$
$$\text{For Convective Rain: } Z = 150\,R^{1.55} \qquad \textbf{(5.3)}$$

The measured TRMM rain-rates are used to estimate Z values. The factor that accounts for the attenuation of the rain echo $x_{measured}$ is calculated as shown in equation (2.7). Again, measured rain-heights are used in this calculation. The estimated received scattering coefficient $\tilde{s}_r^o$ is passed to the reverse simulation, where the rain effects are corrected with the use of rain-height estimates from statistical model.

5.2.2 Reverse Simulation

The reverse simulation is intended to correct for rain-effects in the estimated received scattering coefficient $\tilde{s}_r^o$. The corrected surface scattering coefficient $\tilde{s}_{s-corrected}^o$ is evaluated using

$$s^{\,o}_{s-corrected} = \frac{\tilde{s}_r^{\,o} - (h \times h_{estimated} \times x)}{\exp(-2 \times K_{er} \times SL_{estimated})} \qquad \textbf{(5.4)}$$

Equation (5.4) is very similar to equation (2.8). The difference between the forward and reverse simulation is in the rain-height values used. In reverse simulation, the *estimated* rain-heights from the *regression-based statistical model* are used. If the estimated height from the statistical model were perfect then the corrected surface scattering coefficients would be equal to those obtained using the QuikSCAT geophysical model functions. Clearly, this is not generally the case and thus, by evaluating the difference between $\tilde{s}_s^o$ and $\tilde{s}_{s-corrected}^o$, we can quantify the error in the correction process due to the incorrect estimates of height obtained from the *regression-based statistical model*. Figure 5-3 shows the flowchart of the complete simulation process.

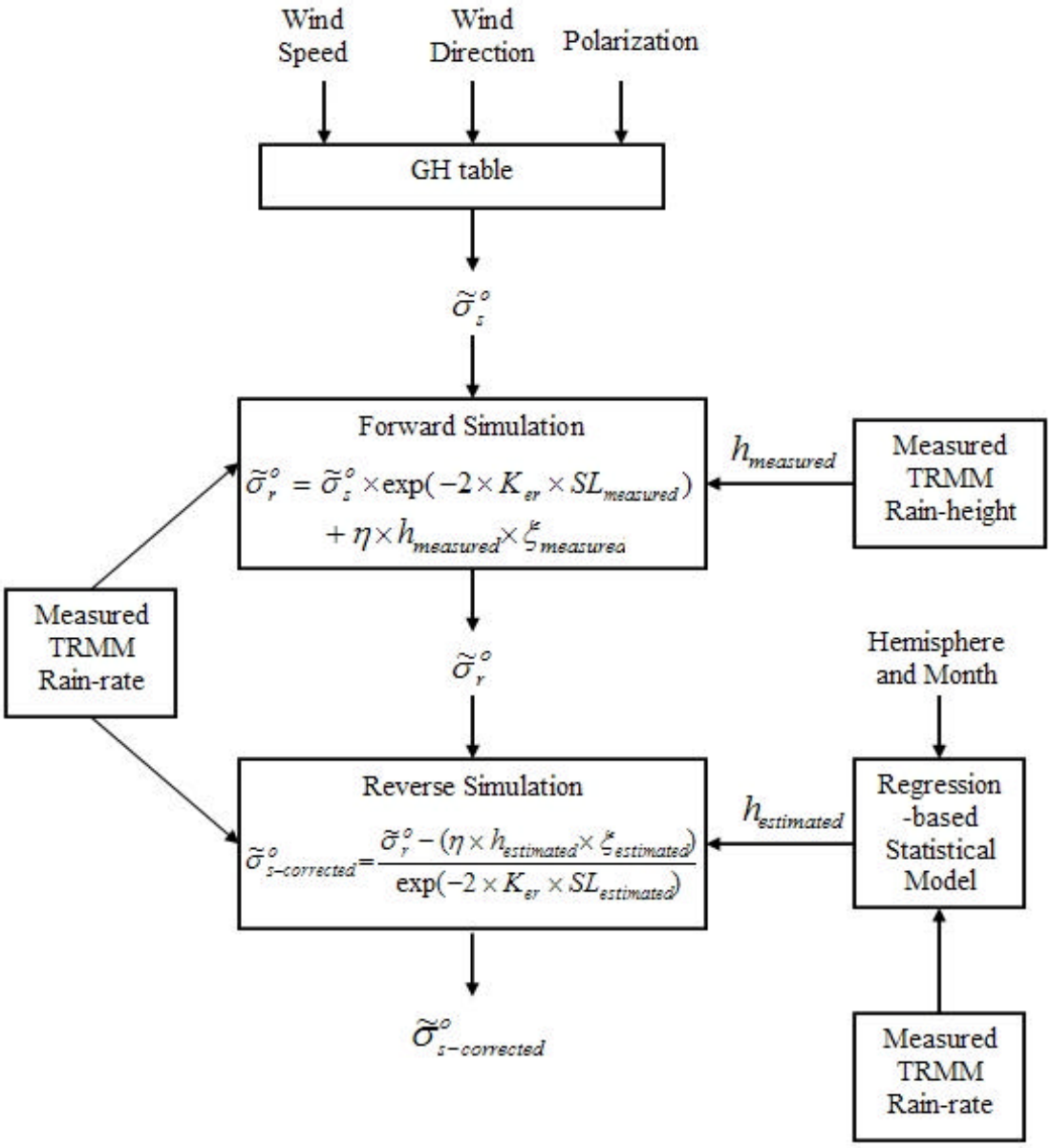


Figure 5-3 Flowchart showing the simulation approach to evaluate the impact of estimated rain-height (using statistical model) in the *rain-effect correction algorithm*.

5.3 Simulation Results

The simulation approach developed in the previous section will help to evaluate the impact of erroneous rain-height estimates on the correction process. In this section, simulation steps shown in Figure 5-3 were performed for every pair of TRMM measured rain-rate and rain-height values in a given region. The resulting scattering coefficients ($\tilde{s}_s^o$, $\tilde{s}_r^o$ and $\tilde{s}_{s-corrected}^o$) are plotted in a figure *vs*. rain-rate. To obtain a greater understanding of these errors in terms of magnitude and variability, we performed these simulations for different combinations of wind speed, wind direction and polarization.

Figure 5-4 shows the simulation results for stratiform rain from the North-Atlantic region for wind speed of 10 m/s, wind direction of 0 degrees (upwind), and horizontal polarization. In this figure, the actual surface scattering coefficient $\tilde{s}_s^o$ obtained from the GH table is shown as the black line. The blue dashed line represents the received normalized radar cross-section $\tilde{s}_r^o$ from the forward simulation. Correction of the received scattering coefficient for rain attenuation and backscatter results in the surface scattering coefficient $\tilde{s}_{s-corrected}^o$ (shown as red and green lines in the figure). The difference between the two is that the green line comes from correction using the statistical model developed for this specific region and the red line comes from correction using the statistical model developed for the entire northern hemisphere. Figure 5-5 shows similar results on rain data from the North-Atlantic region for the

same wind speed (10 m/s) and polarization (horizontal), but for a wind direction of 90 degrees (crosswind).

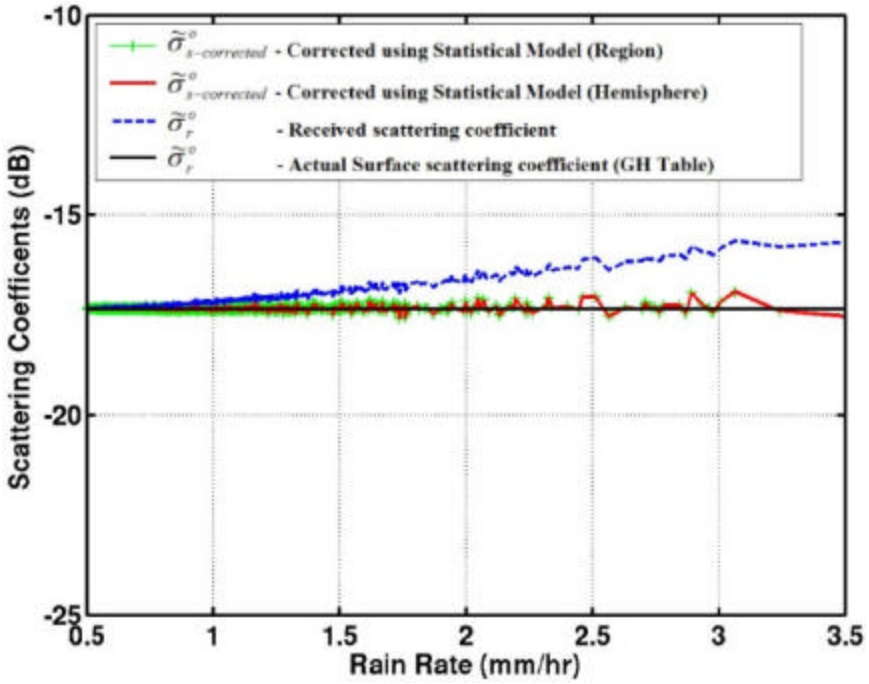


Figure 5-4 Simulation results for Atlantic north region data at WS = 10 m/s and WD = 0 degree

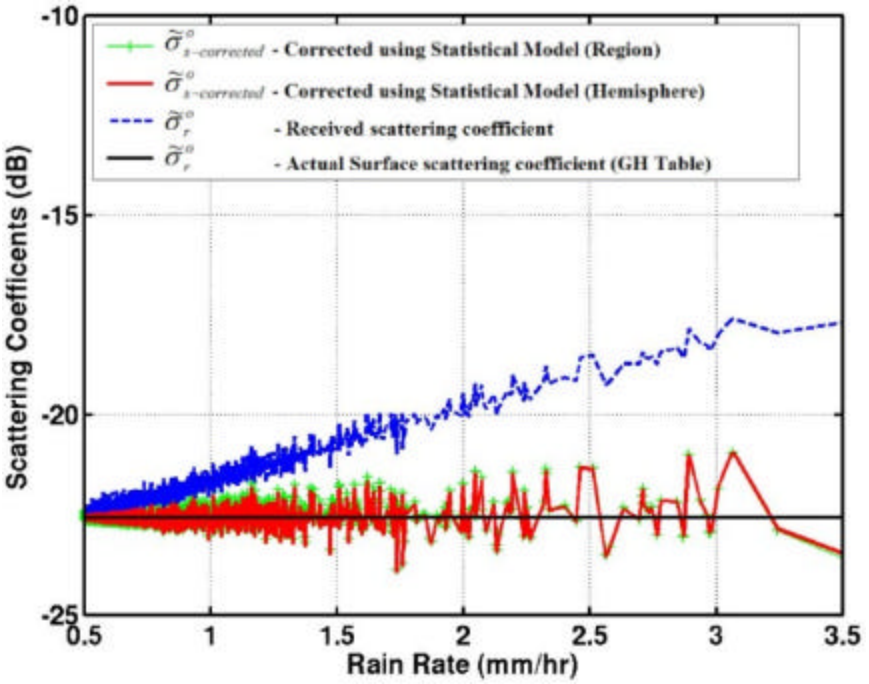


Figure 5-5 Simulation results for Atlantic north region data at WS = 10 m/s and WD = 90 degree

The errors in the corrected surface scattering coefficient are greater at higher rain rates. These larger errors result from errors in the regression model due to larger spreads of the data points from the estimated regression lines.

The simulation results show that the errors between the corrected and true $\tilde{s}_s^o$ values were greater at lower wind speeds than at higher wind speeds. The lower errors at higher wind speeds result from stronger surface signals (high $\tilde{s}_s^o$ values) found for high wind speeds. Our simulations showed similar error estimates for horizontal and vertical polarizations so the *regression-based statistical model* is applicable for both polarizations. We also observed that the error increases as the look direction approaches crosswind because of the weaker surface signals with that look direction. Similar results were obtained for many different regions, polarizations, wind speeds, and wind directions for both stratiform and convective rains.

In all of these figures, the corrected scattering signals obtained from the statistical model developed from regional (green line) and hemispherical (red line) data are very close to each other. This emphasizes the point that use of data for an entire hemisphere is sufficient in estimating rain-heights for all oceanic regions. Thus, our final *rain-effect correction algorithm* need not be applied on a regional basis (longitude regions), resulting in simplification of the computational load on the correction procedure.

The *regression-based statistical model* was found to be effective for rain-height estimation for most cases. Hence the rain-height estimates from this model can be used in the correction algorithm. Sometimes the errors in corrected surface scattering coefficient become intolerable. Typically, these errors occur for cases where the backscatter from the rain is two or three times greater than the surface signal. In these cases, meaningful correction of the surface signal is not possible. Therefore, rain flags to discard the data points must be set when the rain rate exceeds a threshold. Estimation of the rain-rate threshold for stratiform rain data is discussed in [59]. The details of this rain-rate threshold estimation procedure are beyond the scope of this thesis.

6 Conclusion

Measurements of ocean scattering by the SeaWinds on ADEOS II satellite were corrupted by the presence of rain, and the same will be true for all other Ku-band scatterometers. The Radar Systems and Remote Sensing Laboratory at the University of Kansas has developed an algorithm that corrects the received signal using rain-rate and rain-height estimates. Rain- effect corrections are possible up to a rain-rate threshold that depends on the surface wind speed and direction. Above the threshold, one must discard the data by setting a rain flag. The AMSR radiometer on ADEOS II could provide the rain-rate estimates. However, no measurement of rain-cell height was available on ADEOS II. Although ADEOS-II failed, the methods for correction will be useful in future satellites carrying both wind-vector scatterometers and microwave radiometers.

This study evaluated numerous methods to estimate rain height using the data that were available from ADEOS II. The use of climatological height was not promising, since climatology provides only very crude height estimates, and also the climatology for rain at sea is not well known outside the tropics. An attempt to relate rain height to sea-surface temperature also failed, as they correlate well in some oceanic regions but very poorly in others. A method to estimate rain-height from rain-rate was finally

adopted, since TRMM rain rate and rain height showed good correlation. Due to the scatter of measurements, standard linear-regression methods cannot be applied to relate rain-rate and rain-height. After analyzing numerous regression schemes, the *Log-Linear Combined* regression technique using a linear-log regression for low rain rates and a linear regression for high rates was selected. Hypothesis testing and other statistical-based analyses showed that rain-height estimates obtained using the *Log-Linear Combined* regression scheme provide reasonable accuracy.

To apply this regression-based relationship for all seasons, the slope and intercept statistics were Fourier analyzed to capture the seasonal trends. Finally, we proposed the *regression-based statistical model*, which predicts slopes and intercepts of regression lines and estimates the rain height from rain rate for a given hemisphere, season, and rain type. Statistical hypothesis testing showed that rain-height estimates derived from the *regression-based statistical model* provide reasonable accuracy. A simulation procedure was developed and implemented in this thesis to analyze the impact of rain-height estimates in correcting the normalized radar cross-section coefficients.

Simulation results showed good rain-effect correction for surface scattering coefficient for lower rain-rates. These results validate the use of *regression-based statistical model* for rain-height estimation in the *rain effect correction algorithm* for many situations. However, for some high rain-rates and other wind conditions,

correction is not possible. For these cases, rain-rate thresholds have to be introduced in the correction algorithm to set rain flags. Threshold estimation procedure will be the subject of continuing study to complete the *rain effect correction algorithm*.

This study concentrated only on stratiform rain, which is assumed to fill the footprints of SeaWinds and AMSR. A similar study is necessary for convective rain which usually fills the footprints only partially.

In this simulation, the only errors analyzed are those due to height estimates from rain rates. Both correction and simulation are subject to other errors, such as those in the rain-rate estimates from AMSR, attenuation estimates from AMSR, errors in the K_{er}-R and Z-R empirical formulas used in the simulation, and threshold estimation. The sizes of these errors can be estimated by separate studies.

References

[1] R. Atlas, et al., “Global surface wind and flux fields from model assimilation of Seasat data”, J. Geophys. Res., vol. 92, pp. 6477-6487, 1987.

[2] Katsaros, K., et al., “QuikSCAT’s SeaWinds facilitates early identification of tropical depressions in 1999 hurricane season”, Geophys. Res. Lett., vol. 28, pp. 1043-1046, 2001.

[3] W.J. Pierson, “Examples of, reasons for, and consequences of poor quality of wind data from ships for marine boundary layer: Implications for remote sensing”, J. Geophys. Res., vol. 95, pp. 12,313 – 13,340, 1990.

[4] F.M. Naderi, M.H. Freilich and D.G. Long, “Spaceborne radar measurements of wind velocity over the ocean – An overview of the NSCAT scatterometer system”, Proc. the IEEE, vol. 79, No. 6, pp. 850-866, 1991.

[5] R.K. Moore and W.J. Pierson, “Measuring Sea State and Estimating Surface Winds from A Polar Orbiting Satellite”, Proc. Int. Symp. on EM Sensing of Earth from Satellites, Miami Beach, FL, 1966.

[6] Moore, R. K., “Satellite Radar and Oceanography – An Introduction”, in Oceanography from Space, G. C. Ewing, Ed. Woods Hole, MA: Woods Hole Oceanographic Institute, 1965, pp. 355-366.

[7] Moore, R. K., and W. J. Pierson, “Measuring Sea State and Estimating Surface Winds from a Polar Orbiting Satellite”, presented at Inter. Symp. on Electromagnetic Sensing of the Earth from Satellites, Miami Beach, FL, 1965.

[8] Guinard, N.W, Ransone, J.T, Jr, Daley, J.C, "Variation of the NRCS of the sea with increasing roughness", J. Geophys. Res., vol. 76, pp. 1525-1538, 1971.

[9] Grant, C.R, Yaplee, B.S, "Backscattering from water and land at centimeter and millimeter wavelengths", Proc. IRE, vol. 45, pp. 972-982, 1957.

[10] Moore, R. K., et al., “Simultaneous Active and Passive Microwave Response of the Earth – The Skylab Radscat Experiment”, in Proc. 9th Intl. Symp. on Rem. Sens. Of Environ. Ann Arbor, MI: University of Michigan, 1974, pp. 189-217.

[11] Wentz, F. J., S. Peteherych and L. A. Thomas, “A model-function for ocean radar cross-section at 14.6 GHz”, J. Geophys. Res., 89, 3689-3704, 1984.

[12] Moore, R.K. and Fung, A. K., "Radar determination of winds at sea", Proc. IEEE, vol. 67, pp. 1504-1521, 1979.

[13] E. P. Attema, “The Active Microwave Instrument on-board the ERS-1 satellite”, Proc. IEEE, vol. 79 no. 6, pp. 791-799, June 1991.

[14] Michael H. Freilich, David G. Long, Michael and W. Spencer, “SeaWinds: A Scanning Scatterometer for ADEOS-II – Science Overview”, Proc. IEEE, pp. 960-963, 1994.

[15] M. W. Spencer, Chialin Wu, and D.G. Long, “Improved Resolution Backscatter Measurements with the SeaWinds Pencil-Beam Scatterometer”, IEEE Transactions on Geoscience and remote sensing, vol. 38, no. 1, January 2000.

[16] Kirimoto, T. and R. K. Moore, "Scanning Wind-Vector Scatterometers with Two Pencil Beams", in Proc. Conf. Frontiers of Rem. Sens. of Oceans & Troposphere, vol. NASA Conf Pub. 2303. Shoresh, Israel: NASA, 1985, pp. 89-104.

[17] Houze, R.A., "Cloud Dynamics", Academic Press, pp. 573, 1993.

[18] Houze, R.A., "Stratiform precipitation in regions of convection", *Bull. Amer. Meteor. Soc.* 78, 2179-95, 1997.

[19] F.T. Ulaby, R.K. Moore, and A.K. Fung, "Microwave Remote Sensing", vol. I, Addision-Wesley Publishing Company, 1986.

[20] B.W. Stiles, and S.H. Yueh, "Impact of rain on spaceborne Ku-band wind scatterometer data", IEEE trans. on geoscience and remote sensing, vol. 40, no. 9, pp. 1973-1983, Sep 2002.

[21] R. K. Moore, Y. S. Yu, A. K. Fung, D. Kaneko, G.J. Dome, and R. E. Werp, "Preliminary study of rain effects on radar scattering from water surfaces", IEEE J. Geophys. Res., 95, 18,353-18,355, 1990.

[22] Contreras, R. F., W.J. Plant, W.C. Keller, K. Hayes, and J. Nystuen, "Effects of Rain on Ku band Backscatter from the Ocean", J. Geophys. Res., vol. 108, no. c5, 3165, 2003.

[23] D.K. Smith, F.J. Wentz, and C.A. Mears, "Analysis of QuikSCAT rain-flagging winds within the tropical cyclone environment", 25th conference on Hurricanes and Tropical Meteorology.

[24] J.N. Huddleston and B.W. Stiles, "A multi-dimensional histogram rain flagging technique for SeaWinds on QuikSCAT", Proc. IGARSS, vol.3 Honolulu, HI, 2000 pp. 1232-1234.

[25] C.A. Mears, D. Smith, and F.J. Wentz, "Detecting rain with QuikSCAT", Proc. IGARSS, vol. 3, Honolulu, HI, 2000, pp. 1235 – 1237.

[26] M. Portabella, and A. Stoffelen, "Rain detection and quality control of SeaWinds", J. Atmos. Oceanic Technol., vol. 18, no.7, pp. 1171-1183, 2001.

[27] R.K. Moore, D. Braaten, B. Natarajakumar and V.J. Kurisunkal, "Correcting scatterometer ocean measurements for rain effects using radiometer data: Application to SeaWinds on ADEOS-II", Proc. IGARSS, Toulouse, France, July 2003.

[28] Deb. Chatterjee, "Overview of Lookup Table for determining the radiometer-based attenuation a_R for the SeaWinds scatterometer", Radar System and Remote Sensing Laboratory Technical Memorandum 11960-1, University of Kansas, 66045, 1998.

[29] R.K. Moore, N. Kambhammettu, and S. Song, "Correcting SeaWinds Scatterometer Measurements for Atmospheric Attenuation", Proc. IGARSS, vol. 1, Lincoln, Nebraska, pp. 1392-1394, 1996.

[30] N. Kambhammettu and R.K. Moore, "Study of Radiometric Correction of SeaWinds/ADEOS-II Scatterometric Wind Measurements", Proc. IGARSS, Florence, Italy, pp. 1621-1623, 1995.

[31] R.K. Moore, B. Natarajakumar, V.J. Kurisunkal and D. Braaten, "Correlation of rain rate and rain height: A study relating to correction of SeaWinds scatterometer data for rain", General Assembly of International Union for Radio Science (URSI), Maastricht, Netherlands, August 2002.

[32] Gregory J. Lehenbauer, "Using the WSR-88D radar to determine cloud top heights and the spatial characteristics of cloud fractional coverage", Master's thesis, University of Kansas, Physics and Astronomy, 1997.

[33] Prata, A. J., Turner, P.J., "Cloud-top height determination using ATSR data", Rem. Sens. Env., No. 59, 1, pp. 1-13, 1997

[34] Minzer, R. A., Shenk, W. E., Teagle, R. D. and Steranka, J., Stereographic cloud heights from imagery of SMS/GOES satellites, Geophys. Res. Lett., 5(1), pp. 21-24, 1978.

[35] D. Poli, G. Seiz, M. Baltsavias, "Cloud-top height estimation from satellite stereopairs for weather forecasting and climate change analysis", IAPS, Vol. XXXIII, Amsterdam, 2000.

[36] Prata, A. J., and P. J. Turner, Cloud-top height determination using ATSR data, Rem. Sens. Env., 59(1), 1-13, 1997.

[37] Kummerow, C., W. Barnes, T. Kozu, J. Shiue, and J. Simpson, "The Tropical Rainfall Measuring Mission (TRMM) sensor package", J. Atmos. Oceanic Technol., vol. 15, pp. 808-816, 1998.

[38] Tropical Rainfall Measurement Mission (TRMM) Precipitation Radar Algorithm, Instruction Manual version 2.0, January 2000.

[39] B. Murtha, “Exploring the possibilities of detecting the rain rate of storms over remote locations in the oceans, using a side-looking SAR,” Radar Systems and Remote Sensing Laboratory Technical Memorandum 11960-2, University of Kansas, 66045, 1999.

[40] D. A. Short and K. Nakamura, “TRMM Radar Observations of Shallow Precipitation over the Tropical Oceans”, Journal of Climate, vol. 13, pp. 4107-4124, December 2000.

[41] C. S Mason., B. Petrie, and B. J. Topliss., “Satellite measurement of sea-surface temperatures and the study of ocean climate”, Can. Tech. Rep. Hydrogr. Ocean Sci., vol. 193, p. 101, 1998.

[42] Zong-Liang Yang, D. Gochis, W. J. Shuttleworth and Gue-Yue Niu, “The impact of sea surface temperature on the North American monsoon: A GCM study”, Geophysical Res. Lett., vol. 30, no. 2, pp. 1033, 2003.

[43] P. Terray, P. Delecluse, S. Labattu, L. Terray, “Sea Surface Temperature associations with the Indian summer monsoon”, European Geophysical Society, vol. 5, 03455, 2003.

[44] McClain, E. P., Pichel, W. G., and Walton, C. C., “Comparative performance of AVHRR-based multichannel sea surface temperatures”, Journal of Geophysical Research, vol. 90, pp. 11587-11601, 1985.

[45] Pichel, W. G., “Operational production of multichannel sea surface temperatures from NOAA polar satellite AVHRR data” Palaeogeography, Palaeoclimatology, Palaeoecology, 90, pp.173–177, 1991.

[46] Goldhirsh J., Musiani Bert, “Dimension statistics of rain cell cores and associated rain rate isophlets derived from radar measurements in the mid-Atlantic coast of the United States”, *IEEE Trans. Geosci. and Remote Sens.*, Vol. 30, No. 1, pp. 28-37, January 1992.

[47] http://www.eorc.nasda.go.jp/TRMM/reentory/reentory_e.htm

[48] http://www.gsfc.nasa.gov/news-release/releases/2001/01-84.htm

[49] http://www.eorc.nasda.go.jp/TRMM/reentory/pdf/PR_Impact_e.pdf

[50] R.K. Moore, B. Natarajakumar, V.J. Kurisunkal and D. Braaten, “Correlation of rain rate and rain height: A study relating to correction of SeaWinds scatterometer data for rain”, General Assembly of International Union for Radio Science (URSI), Maastricht, Netherlands, August 2002.

[51] Samprit Chatterjee, Ali S. Hadi, Bertram Price, “Regression analysis by example, 3rd Edition”, John Wiley & Sons, 2000

[52] Edwards, A. L., "An Introduction to Linear Regression and Correlation", W. H. Freeman, 1976

[53] Douglas C. Montgomery, Elizabeth A. Peck, G. Goeffrey Vining, "Introduction to Linear Regression Analysis, 3rd Edition", John Wiley & sons, 2001

[54] Ryan T. P., "Modern regression methods", John Wiley & sons, 1997

[55] Draper N. R., Smith H., "Applied regression analysis, 3rd Edition", John Wiley & sons, 1998

[56] Robert V. Hogg, Allen T. Craig, “Introduction to Mathematical Statistics – 5th Edition”, Prentice Hall, December 1994

[57] Hine. J., Wetherill., G. B., "A Programmed Text in Statistics: The T-test and X Goodness-of-fit", Kluwer Academic Publishers Group, September 1975

[58] Alan V. Oppenheim, Ronald W. Schafer, John R. Buck, "Discrete-Time Signal Processing – 2nd Edition", Prentice Hall, February 1999.

[59] R.K. Moore, D. Braaten, B. Natarajakumar and V.J. Kurisunkal, "Correcting scatterometer ocean measurements for rain effects using radiometer data: Application to SeaWinds on ADEOS-II (On completion of Threshold Estimation for Stratiform Fully Filled Cells)", Radar Systems and Remote Sensing Laboratory Technical Memorandum 11960-3, University of Kansas, 66045, 2003.

[60] Tropical Rainfall Measuring Mission Science Data and Information System Interface Control Specification between the Tropical Rainfall Measuring Mission Science Data and Information (TSDIS) and the TSDIS Science User (TSU) TSDIS-P907 Volume 4: File Specifications for TRMM products – Level 2 and Level 3 – Release 5.21

Appendix A

Data Products

The first spaceborne precipitation radar (PR), an active microwave instrument aboard the Tropical Rainfall Measuring Mission (TRMM) satellite, was designed to measure rain and its vertical structures over the vast tropical oceans and continents. The PR's sensitivity has exceeded expectations, allowing routine detection of rain-fall intensities as low as 0.5 mm h^{-1} [37]. The PR operates at a wavelength of 2.2 cm, allowing penetration of even the heaviest obscuring cloud layers and detection of underlying precipitation. This, combined with a vertical resolution of 250 m, makes the PR an effective platform for investigating shallow precipitation [40].

TRMM

The Tropical Rainfall Measuring Mission (TRMM) precipitation radar (PR) is the first spaceborne rain radar and the only instrument on TRMM that can directly observe vertical distributions of rain. The frequency of TRMM PR is 13.8 GHz. The PR can achieve quantitative rainfall estimation over land as well as ocean. The precipitation radar provides statistics on rain rate and rain height (storm height) in

different oceanic regions on a monthly basis from January 1998. TRMM coverage extends only for ±35° latitude.

TRMM Precipitation-Radar Algorithms

The TRMM PR, standard algorithms were developed by the TRMM science team. They are classified into Level 1 (1B21, 1C21,), Level 2 (2A21, 2A23, 2A25) and Level 3 (3A25, 3A26) [60]. Level 1 and Level 2 products are data observed in each cell. Level 3 data give the monthly statistical values of rain parameters in 0.5° x 0.5° grid boxes required by the TRMM mission. The data products used in our analysis are 2A25 and 3A25. The algorithm 2A25 retrieves profiles of radar reflectivity factor, Z, with rain-attenuation correction and rain rate for each radar beam. The algorithm 3A25 gives the space-time averages of accumulations of 1C21, 2A21, 2A23, and 2A25 products.

3A25 Data Products – for Rain Rate, Height and Count

The algorithm 3A25 provides space-time statistics over a month from the level 1 & level 2 PR output products. Four types of statistics are calculated.

1. Probability of occurrence
2. Mean and Standard Deviation
3. Histogram
4. Correlation Coefficient

The standard space scale is a 5° x 5° latitude x longitude cell. A subset of the product, however, is also produced over 0.5° x 0.5° cells. The relevant products are:

1. Rain Rates (millimeter/hour; 4-byte real)

StratRainMean2 (148,720,4) – mean Stratiform rain rates at 4 levels (2Km, 4Km, 6Km, & Path-Averaged)

ConvRainMean2 (148,720,4) – mean Convective rain rates at 4 levels (2Km, 4Km, 6Km, & Path-Averaged)

2. Rain Counts (unitless; 4-byte integer)

StratRainPix2 (148,720,4) – Stratiform rain counts at 4 levels (2Km, 4Km, 6Km, & Path-Averaged)

ConvRainPix2 (148,720,4) – Convective rain counts at 4 levels (2Km, 4Km, 6Km, & Path-Averaged)

3. Storm Height (meters; 4-byte real)

stormHeightMean (148,720,3) – mean of storm height for stratiform, convective and all types.

The path-averaged rain rates are used in our studies. They are calculated by summing the rain rate values from the storm top (first gate where rain is detected) to the last gate (gate nearest to the surface uncontaminated by the surface cluster) and divided by the number of gates in the interval.

Sea Surface Temperature (SST) Data Product

The multi-channel Sea Surface Temperature (MCSST) data set is derived from the NOAA advanced very high-resolution radiometer (AVHRR). These data sets are available throughout the globe at 0.35° x 0.35° grid resolution. The data sets must be decimated and trimmed to match the rain-height data set in our analysis.

Appendix B

Bilinear Regression

Let us consider X and Y to be the two datasets to be regressed. X can be assumed as the independent variable from which the dependent Y is to be determined. When the scatter exhibits two different trends, bilinear regression can be used. In bilinear regression, a breakpoint (say x_0) is considered in the independent variable X. For values of X less than x_0, a linear regression line is fitted and for values greater than x_0, another linear regression line is fitted. These two regression lines are made to meet at a point to maintain consistency.

The objective function for bilinear regression is given as,

$$S(m_1, m_2, c_1, c_2) = \sum_{i=X<x_0} (y_i - c_1 - m_1 x_i)^2 + \sum_{j=X>x_0} (y_j - c_2 - m_2 x_j)^2$$

where, m_1, c_1 are the slope and intercept of the line for lower values of X and m_2, c_2 are the slope and intercept of the line for higher values of X. To maintain consistency, the regression lines have to be made to meet at a point. Hence, an additional criterion must be introduced in the objective function to force the regression lines to meet at a point.

$$S(m_1, m_2, c_1, c_2) = \sum_{i=X<x_0} (y_i - c_1 - m_1 x_i)^2 + \sum_{j=X>x_0} (y_j - c_2 - m_2 x_j)^2 + l\,(c_1 - c_2 + (m_1 - m_2)x_0)$$

Where l is the Lagrange multiplier and x_0 is the break point between the two regression lines. The slopes and intercepts of the regression lines are the unknown parameters to be determined. We obtain the partial derivative of the objective function with respect to each of the unknown parameters (m_1, m_2, c_1, c_2) and equate it to zero.

$$\frac{\partial S}{\partial m_1} = -2 \sum_{i=X<x_0} (y_i - c_1 - m_1 x_i)\, x_i + l x_0 = 0$$

$$\frac{\partial S}{\partial m_2} = -2 \sum_{j=X>x_0} (y_j - c_2 - m_2 x_j)\, x_j - l x_0 = 0$$

$$\frac{\partial S}{\partial c_1} = -2 \sum_{i=X<x_0} (y_i - c_1 - m_1 x_i) + l = 0$$

$$\frac{\partial S}{\partial c_2} = -2 \sum_{j=X>x_0} (y_j - c_2 - m_2 x_j) - l = 0$$

Solving these equations gives the unknown parameters. The final solution in the form of a matrix is given as

$$\begin{bmatrix} c_1 \\ c_2 \\ m_1 \\ m_2 \\ l \end{bmatrix} = \begin{bmatrix} \sum_{X<x_0} 1 & 0 & \sum_{i=X<x_0} x_i & 0 & \tfrac{1}{2} \\ \sum_{X<x_0} x_i & 0 & \sum_{X<x_0} x_i^2 & 0 & {}^{x_0}\!/_2 \\ 0 & \sum_{X>x_0} 1 & 0 & \sum_{X>x_0} x_i & -\tfrac{1}{2} \\ 0 & \sum_{X>x_0} x_i & 0 & \sum_{X>x_0} x_i^2 & -{}^{x_0}\!/_2 \\ \tfrac{1}{2} & -\tfrac{1}{2} & {}^{x_o}\!/_2 & -{}^{x_o}\!/_2 & 0 \end{bmatrix}^{-1} \times \begin{bmatrix} \sum_{X<x_o} y_i \\ \sum_{X<x_o} x_i y_i \\ \sum_{X>x_o} y_i \\ \sum_{X>x_o} x_i y_i \\ 0 \end{bmatrix}$$

Weighted Bilinear Regression

For weighted bilinear regression, each point in the scatter is provided with a weighting factor. Let n_i be the weighting factor to be considered in the regression. The objective function for weighted bilinear regression is modified as follow:

$$S(m_1, m_2, c_1, c_2) = \sum_{i=X<x_0} n_i (y_i - c_1 - m_1 x_i)^2 + \sum_{j=X>x_0} n_j (y_j - c_2 - m_2 x_j)^2 + l(c_1 - c_2 + (m_1 - m_2)x_0)$$

The objective function S should be partially differentiated with respect to the unknown parameters and equated to zero.

$$\frac{\partial S}{\partial m_1} = -2 \sum_{i=X<x_0} n_i (y_i - c_1 - m_1 x_i) x_i + l x_0 = 0$$

$$\frac{\partial S}{\partial m_2} = -2 \sum_{j=X>x_0} n_j (y_j - c_2 - m_2 x_j) x_j - l x_0 = 0$$

$$\frac{\partial S}{\partial c_1} = -2 \sum_{i=X<x_0} n_i (y_i - c_1 - m_1 x_i) + l = 0$$

$$\frac{\partial S}{\partial c_2} = -2 \sum_{j=X>x_0} n_j (y_j - c_2 - m_2 x_j) - l = 0$$

The final solution in matrix form is given as,

$$\begin{bmatrix} c_1 \\ c_2 \\ m_1 \\ m_2 \\ l \end{bmatrix} = \begin{bmatrix} \sum_{X<x_0} n_i & 0 & \sum_{i=X<x_0} n_i x_i & 0 & \frac{1}{2} \\ \sum_{X<x_0} n_i x_i & 0 & \sum_{X<x_0} n_i x_i^2 & 0 & \frac{x_0}{2} \\ 0 & \sum_{X>x_0} n_i & 0 & \sum_{X>x_0} n_i x_i & -\frac{1}{2} \\ 0 & \sum_{X>x_0} n_i x_i & 0 & \sum_{X>x_0} n_i x_i^2 & -\frac{x_0}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{x_o}{2} & -\frac{x_o}{2} & 0 \end{bmatrix}^{-1} \times \begin{bmatrix} \sum_{X<x_o} n_i y_i \\ \sum_{X<x_o} n_i x_i y_i \\ \sum_{X>x_o} n_i y_i \\ \sum_{X>x_o} n_i x_i y_i \\ 0 \end{bmatrix}$$