GPS signal derived precipitable water vapor and its applications in rainfall prediction

Manandhar, Shilpa

2019

Manandhar, S. (2019). GPS signal derived precipitable water vapor and its applications in rainfall prediction. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/106016>

<https://doi.org/10.32657/10220/48092>

**GPS SIGNAL DERIVED PRECIPITABLE
WATER VAPOR AND ITS APPLICATIONS IN
RAINFALL PREDICTION**

SHILPA MANANDHAR
SCHOOL OF ELECTRICAL & ELECTRONIC ENGINEERING
2019

GPS SIGNAL DERIVED PRECIPITABLE WATER VAPOR AND ITS APPLICATIONS IN RAINFALL PREDICTION

SHILPA MANANDHAR

School of Electrical & Electronic Engineering

A thesis submitted to Nanyang Technological University in
partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2019

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarized materials, and has not been submitted for a higher degree to any other University or institution.

Date : 09/01/2019

Shilpa Manandhar

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanayang Technological University and that the research data are presented honestly and without prejudice.

Date : 09/01/2019

Yee Hui Lee

Authorship Attribution Statement

This thesis contains materials from 2 published journal papers and 1 under major revision journal paper, in the following peer-reviewed journals, where I was the first and the corresponding author.

Chapter 4 is published as: S. Manandhar, Y. H. Lee, Y. S. Meng and J. T. Ong, “A Simplified Model for the Retrieval of Precipitable Water Vapor from GPS Signal,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6245-6253, Nov, 2017.

The contributions of the co-authors are as follows:

- Assoc. Prof Lee Yee Hui provided the initial project direction and helped in editing the manuscript.
- I prepared the manuscript drafts. The manuscript was revised by Dr Yusong Meng. He helped in pointing out probable questions from the reviewers.
- Prof. Ong has always been supervising our team. He helped in analysis of few results.
- I used the software and generated all the necessary data. The post processing of the data was done by me. I produced the results and did initial analysis.

Chapter 6 is published as: S. Manandhar, Y. H. Lee, Y. S. Meng, F. Yuan and J. T. Ong, “GPS Derived PWV for Rainfall Nowcasting in Tropical Region,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4835-4844, Aug, 2018.

The contributions of the co-authors are as follows:

- Assoc. Prof Lee Yee Hui provided the initial project direction and helped in editing the manuscript.
- I prepared the manuscript drafts. The manuscript was revised by Dr Yusong Meng and Dr. Yuan Feng.
- I liaised in managing the weather station data from Brazil. I processed the GPS and the weather station data and generated the initial results. Dr. Yaun Feng gave his inputs in processing few data.
- Assoc. Prof Lee and Prof Ong helped in analyzing the graphs and improving the results.

Chapter 7 is under major revision as: S. Manandhar, S. Dev, Y. H. Lee, Y. S. Meng and S. Winkler, “A Data-Driven Approach to Improve the Rainfall Prediction for Tropical Region,” *IEEE Trans. Geosci. Remote Sens.*, 2018. [Major Revision, Decision received in Oct 2018]

The contributions of the co-authors are as follows:

- I and Dr. Dev started with the initial idea. I prepared the manuscript. The manuscript was thoroughly revised by Dr. Dev.
- Assoc. Prof Lee and Dr. Yusong Meng helped in revising the manuscript. They helped in improving the time series analysis section.
- Dr. Winkler helped in improving the machine learning section. He gave his inputs in improving the statistical results presented in the paper.
- I processed the data and did the initial analysis of the results.

Date : 09/01/2019

Shilpa Manandhar

Acknowledgment

I would like to take this opportunity to thank my supervisor, Prof. Lee Yee Hui, for supervising me during my PhD years. I got to learn a lot of technical and practical things under her guidance. She is a wonderful supervisor who has not only motivated me in doing my research but also has equally supported me in my other co-curricular interests. I feel lucky to have gotten the chance to do my PhD under her. I would like to thank my co-supervisor Dr. Yu Song Meng for his support during my PhD journey. He is a very hardworking and a dedicated researcher. Thank you Yusong for your effort in going through my works and giving me constructive feedback, always. I would like to thank Dr. Stefan Winkler and Prof Ong. for working closely with our project and guiding us in our work.

I feel extremely fortunate to have a group of lovely lab mates with whom I spent most of my PhD days. I would like to thank my senior cum mentor Sutha Mam for her guidance during my initial days of PhD. Thank you Sutha Mam for your constant support and well wishes. I would like to thank my seniors Jun Xiang, Soumya, Siya, Yuan Feng and Florian. I had had most wonderful memories and have shared a warm bonding with Soumya and Yuan Feng. I would also like to thank my junior Shravan who extends his helping hands whenever needed.

Apart from my lab-mates, I feel extremely lucky of having a bunch of friends who were there during my thick and thins. I would like to thank my friends Niroj, Milan, Dipu, Ujjal, Rakesh and Shiva for being around and supporting me. I would like to thank Anna and Vasilisa for all those forgetful and fun-filled memories. And of course, Pradeep, thank you very much for your constant encouragement, support and unconditional love. Without the support of my friends, my PhD journey would

not have been a joyous experience.

My acknowledgement is incomplete without thanking my parents. I would like to thank my father and mother, Shree Hari Manandhar & Sarojani Manandhar for their love and constant support. Especially my mother who has always encouraged me to live my dreams and follow my passion. I am also thankful towards my sister, Sarbada Manandhar, not only for loving and supporting me but also for looking after my family during my PhD years when I am away from home. Also a huge thanks to my grand-pa and grand-ma for their blessings and well-wishes. I would like to express my sincere gratitude towards each and every individual who has directly or indirectly helped me in reaching my goal. Thank you.

Abstract

The atmospheric water vapor is generally expressed in terms of Precipitable water vapor (PWV). It is an important indicator of water vapor climatology and variability in the lower troposphere and related climate processes. Global Positioning System (GPS) satellite signals are used to derive PWV values with high spatio-temporal resolution which is available under all-weather conditions. The PWV values are derived using the delay information from GPS signals. The signal delays in the troposphere are divided into two categories; Zenith Hydrostatic Delay (ZHD) and Zenith Wet Delay (ZWD). The ZWD delay is caused by the water vapor of the atmosphere and PWV values are derived using these ZWD values.

The accuracy of the GPS-PWV values depends on how accurately the ZWD values are derived. A comprehensive analysis is done to understand the effects of different mapping functions and different ZHD models on the retrieved PWV values. The results show that there is a very small effect of the mapping functions on the PWV values. Use of different ZHD models have greater effect on the PWV values.

When converting the ZWD values to PWV, it involves the use of a dimensionless constant called PI . This PI value is found to rely on a water-vapor weighted mean temperature (T_m) value which varies according to different locations and seasons. It is therefore, both site and time specific. Analysis of the PI value and its effect on the retrieved PWV from the data obtained from different regions show that although the PI value is time and site specific, the change in the median value of PI for different years is minimal and is dependent only on factors like the latitude coordinates of the particular site and the day-of-the year. Therefore, using the data obtained from 174 different sites, a latitude-coordinate and day-of-year based PI value model is pro-

posed for the retrieval of PWV. The proposed model has been successfully validated using data from different databases.

The GPS-PWV values are useful in applications like rainfall prediction, cloud detection, etc. With the rapid deployment of the GPS CORS (Continuously Operating Reference Stations), many researchers are studying the PWV values and its usefulness in the prediction of a rainfall event. This thesis firstly reviews the rainfall prediction algorithms from the temperate and sub-tropical regions. Secondly, We propose a prediction algorithm suitable for the tropical region. It was found that the absolute value of PWV plays an important role in long-term rainfall prediction, unlike for the temperate and the sub-tropical regions. In addition to the long-term rain prediction, a simple algorithm is proposed to perform the rainfall nowcasting in the tropical region. The algorithm applies GPS-derived PWV values and its second derivatives. The proposed algorithm incorporates the seasonal dependency of PWV values for the prediction of a rain event.

The occurrence of rainfall is complicated and dependent on a myriad of atmospheric parameters. Therefore, this research is extended into the systematic study of various parameters that affect the precipitation in the atmosphere. Different ground based weather parameters are identified as probable features for rainfall prediction and a detailed feature correlation study is presented. It has been shown that only a few parameters; PWV, solar radiation, seasonal and diurnal factors stand out for the rainfall prediction.

In summary, in this thesis, we have studied the GPS signal delays and related processes in detail. We have done a comprehensive study on different hydrostatic delay models and their effect on the PWV values. We have proposed a simplified *PI* model that is globally applicable. We have proposed rainfall prediction algorithms suitable for the tropical region. We have implemented a data-driven approach that uses other meteorological parameters to improve the rainfall prediction.

Contents

Statement of Originality	i
Supervisor Declaration Statement	ii
Authorship Attribution Statement	v
Acknowledgment	vii
Abstract	ix
Contents	xiv
List of Figures	xxii
List of Tables	xxiv
Abbreviations	xxvi
1 Introduction	1
1.1 Background and Motivation	1
1.1.1 Objectives and Contributions	2
1.1.2 Major contributions	2
1.2 Thesis Outline	3
2 Literature Review	6
2.1 Overview	6

2.2	Precipitable Water Vapor and its Significance	6
2.3	State-of-The-Art Techniques for Precipitable Water Vapor Retrieval	7
2.3.1	Radiosondes for PWV Retrieval	8
2.3.2	Microwave Radiometers for PWV Retrieval	12
2.3.3	Satellite-Based Instrument for PWV Retrieval	13
2.3.4	GPS Signals for PWV Retrieval	16
2.4	GPS Signal Delays for PWV Estimation	18
2.4.1	GPS	18
2.4.2	GPS Signal Delays	20
2.4.3	Mapping Functions	22
2.4.4	Empirical Model-Based Estimation of ZHD Values	26
2.4.5	PPP-Based Estimation of ZWD Values	27
2.4.6	PWV from ZWD Values	29
2.4.7	GIPSY Processing for Estimating PWV	31
2.4.8	Error Sources for GPS-PWV Values	35
2.5	Applications of GPS Derived PWV Values	36
2.5.1	GPS-PWV for Rainfall Monitoring	36
2.5.2	GPS-PWV for Rainfall Forecasting	37
2.5.3	Data-Driven Methods	38
2.6	Summary	39
3	Accuracy Assessment of GPS-Derived PWV	40
3.1	Introduction	40
3.1.1	Outline of the Contributions	41
3.2	Effects of Using Different Mapping Functions	42
3.2.1	Time Series Plots	43
3.2.2	Residual Plot of Differences	44
3.3	Effects of Using Different ZHD Models	45
3.3.1	Time Series Plots	45
3.4	GPS-PWV Vs Radiosonde-PWV	49
3.4.1	Root Mean Square Errors (RMSE)	49

3.4.2	Density Plots	51
3.5	GPS-PWV and Radiosonde-PWV Values at Varying Cut-Off Angles .	54
3.5.1	GPS-PWV Values at Different Cut-Off Angles	54
3.5.2	Methodology	59
3.5.3	Results and Discussion	61
3.6	Comparison of GPS-Derived PWV and MODIS-Derived PWV Values	63
3.7	Summary	66
4	Conversion of Zenith Wet Delays to PWV	68
4.1	Introduction	68
4.1.1	Outline of The Contribution	69
4.2	Mean Weighted Temperature (T_m) and Surface Temperature Relationships (T_s)	71
4.2.1	Review of Existing ($T_m - T_s$) Models	71
4.2.2	Derivation of New ($T_m - T_s$) Relationships	72
4.3	Modeling and Proposal of a New PI Function	74
4.3.1	Trend in PI Values with Day of Year	74
4.3.2	Study of Latitude Dependency	76
4.3.3	Proposal of A Simplified Model	79
4.4	Performance Evaluation and Validation	80
4.4.1	Performance Comparison	80
4.4.2	Further Validation of the Proposed Model	81
4.4.3	Independent Station Validation	84
4.5	Summary	86
5	GPS-Derived PWV Values for Long-Term Rainfall Prediction	87
5.1	Introduction	87
5.1.1	Outline of the Contribution	88
5.2	Time Series of GPS-Derived PWV and Rain	91
5.3	Review and Implementation of Existing Long-Term Rainfall Prediction Algorithms	95

5.3.1	Rain Prediction Algorithm from the Temperate Region	97
5.3.2	Rain Prediction Algorithm from Sub-Tropical Region	99
5.4	Long-Term Rainfall Prediction for Tropical Region	106
5.4.1	Optimum Threshold Criteria for The Tropical Region	107
5.4.2	Lead Time Values for Rainfall Prediction for Tropical Region .	114
5.5	Summary	117
6	GPS-Derived PWV Values for Rainfall Nowcasting	118
6.1	Introduction	118
6.1.1	Outline of the Contribution	118
6.2	PWV and Rainfall	119
6.2.1	Representation of PWV and Rainfall Data	120
6.2.2	Seasonal and Diurnal Variations in PWV	121
6.3	A Season Dependent PWV Threshold Value	123
6.4	Composite Analysis of PWV Derivatives with respect to Rainfall Ini- tiation	124
6.4.1	First Derivative of PWV Values:	126
6.4.2	Second Derivative of PWV Values:	127
6.5	Rainfall Nowcasting	127
6.5.1	Rainfall Nowcasting Algorithm	128
6.5.2	Evaluation Metrics	128
6.5.3	Results And Discussion	131
6.5.4	Performance Validation Using Data from Independent Years at the Same Station	132
6.6	Evaluation of the Proposed Algorithm	132
6.6.1	Performance Validation Using Data from Another Station in Singapore	133
6.6.2	Performance Validation Of The Proposed Methodology For A Station From Brazil	134
6.7	Summary	137

7	A Data-driven Approach to Improve the Rainfall Prediction Results	139
7.1	Introduction	139
7.1.1	Outline of the contribution	139
7.2	Approach & Tools	141
7.2.1	Supervised Machine Learning Technique	141
7.2.2	Evaluation Metrics	142
7.2.3	Downsampling Technique	142
7.3	Feature Identification & Correlation	143
7.3.1	Features	143
7.3.2	Feature Correlation Matrix	146
7.3.3	Time Series Observations	147
7.4	Rainfall Prediction	150
7.4.1	Assessment of Individual Features for Rainfall Prediction . . .	151
7.4.2	Selection of Optimum Features	152
7.4.3	Benchmarking	155
7.4.4	Different Lead Time Values	157
7.5	Summary	160
8	Conclusion & Future Work	161
8.1	Conclusion	161
8.2	Future Work	164
	Author's Publications	169
	References	170
A	Details on location of GPS, Radiosonde and Weather Stations	184

List of Figures

2.1	Radiosonde balloon with the radio transmitter	8
2.2	Location of the radiosonde stations in different regions. The black dots mark the location of the radiosonde stations.	11
2.3	IGS GPS Stations distributed around the globe. The green and red dots mark the IGS stations. Image Source: http://www.igs.org/network	16
2.4	GPS satellite constellation: 24 satellites in six orbit planes	19
2.5	Total atmospheric delay	20
2.6	Atmospheric delays mapped to the zenith direction.	21
2.7	Schematic representation of PWV retrieval using GIPSY OASIS software.	32
3.1	Locations of the co-located Radiosonde and IGS GPS stations. Markers in red (1-8) represent the tropical stations, in green (9-14) represent the sub-tropical stations and in yellow (15-26) represent the temperate stations, which includes 4 polar stations (station No. 20-23). Details of these stations are mentioned in table A.1.	42

3.2	Time series plot of PWV values (Year 2012) calculated by using different mapping functions for three randomly chosen IGS GPS stations; Station No. 1, 10 and 16 in (a), (b) and (c) respectively. The small figures corresponding to each plot, show the zoomed-in view of the portion highlighted by the dotted box. The details about the stations are mentioned in the figure title. The x-axis indicates the day number and y-axis indicates the PWV values. Curves in red, blue, green and black represent PWV values when mapping functions NMF, GMF, GPT2 and VMF are used respectively [<i>Best viewed in color</i>].	43
3.3	Histogram of the difference between PWV values (mm) when using NMF mapping function and other functions as indicated by the legend for (a) tropical region (b) sub-tropical region and (c) temperate region. The black line marks the position of 0 difference.	44
3.4	Time series plot of PWV values (Year 2012) calculated by using different <i>ZHD</i> models for three randomly chosen IGS GPS stations; Station No. 1, 10 and 16 in (a), (b) and (c) respectively. The small figures corresponding to each plot, show the zoomed-in view of the portion highlighted by the dotted box. The details about the stations are mentioned in the figure title. The x-axis indicates the day number and y-axis indicates the PWV values. Curves in red, blue, green and black represent PWV values when <i>ZHD</i> models Static, GPT, GPT2 and VMF1 are used respectively [<i>Best viewed in color</i>].	46
3.5	Histogram of the difference between PWV values (mm) when using Static <i>ZHD</i> model and other models as indicated by the legend for (a) tropical region (b) sub-tropical region and (c) temperate region. The black line marks the position of 0 difference.	47

3.6	Average Root Mean Square Error (mm) of difference between GPS-PWV and Rad-PWV for all the stations (ref. Fig. 3.1) for years 2012-2015. Values in star, circle, triangle and square markers represent the RMS Errors when Static, GPT, GPT2 and VMF1 <i>ZHD</i> models are used respectively. The absolute latitude values are plotted. The broken lines separate the tropical, sub-tropical, temperate and the polar regions. [<i>Best Viewed in Color</i>]	50
3.7	Density plot of GPS-PWV Vs Rad-PWV for (a) tropical region (b) sub-tropical region and (c) temperate regions for data from years 2012-2015. Color bars represent the total number of data points in each bin (1 mm). [<i>Best Viewed in color</i>]	52
3.8	(a)Schematic diagram of elevation cut-off angles for GPS satellite observations. (b)Number of GPS satellites observed at different Elevation Cut-Off Angles.	55
3.9	Residual values plotted against (a) Elevation and (b) Azimuth angles for the data from 31-Dec-2012 for IGS GPS station, NTUS. The blue dots for both the figures represent the uncorrected residuals and the red dots represent the corrected residuals. [Best viewed in color] . . .	56
3.10	Zenith wet delay and projected slant wet delay on 12-Dec-2018 for the NTUS IGS GPS station. Red dots represent the average zenith wet delay values and the blue dots represent the slant delays projected on the zenith direction. [Best viewed in color]	57
3.11	Slant water vapor (<i>mm</i>) for IGS GPS station NTUS. Different colors represent the different range of slant water vapor values as shown by the figure legend. [Best viewed in color]	58
3.12	Map of Singapore showing 4 GPS Stations (red marker) and 1 Radiosonde Station (green marker). Marker 1: Radiosonde Station, GPS Station Names; Marker 2: SRPT, Marker 3: SNPT, Marker 4: SLYG and Marker 5: SNYU. The map is generated from an online map customizer site (https://www.mapcustomizer.com/). [Best viewed in color]	60

3.13	(a) Correlation Coefficient in % and (b) RMSE in mm between GPS-PWV and radiosonde-PWV.	61
3.14	Scatter plot of MODIS-PWV and GPS-PWV values from the tropical, the sub-tropical and the temperate regions. The description of the markers are given in the legend and along the text. Solid black line indicates the $x=y$ axis. [Best viewed in color]	64
4.1	Location of 174 radiosonde stations. Markers numbered from 1–59 (in red color) represent the temperate stations, numbered from 60–121 (in green color) represent the sub-tropical stations and those from 122 – 174 (in pink) represent the tropical stations. The map is generated from an online map customizer site (https://www.mapcustomizer.com/). [Best viewed in color]	70
4.2	T_m vs T_s for stations from temperate region (in red dots), from sub-tropical region (in blue circles) and from tropical region (in green triangles) with linearly regressed equations represented in black, yellow and magenta for the respective regions. [Best viewed in color]	73
4.3	Plot of daily mean PI values against Day of Year (DoY), for 4 years of data from 2012 to 2015. The red and green circles are for the northern and the southern tropical stations, blue and cyan dots are for the northern and the southern sub-tropical stations and black and magenta crosses are for the northern and the southern temperate stations respectively. The solid red line represents the best fit line for respective regions. [Best viewed in color]	75
4.4	Boxplot of PI values for 174 radiosonde stations for 4 years (2012-2015). Red color represents the stations from negative latitude and black from positive latitude. [Best viewed in color]	77
4.5	(a) Study of median PI values with respect to absolute latitude values. (b) Altitude correction for the estimated PI values in (a). (c) Variation in parameter c with respect to the absolute latitude values. [Best viewed in color]	78

4.6	(a) Histogram of the difference between $Prop - PI$ and $GGOS - PI$ (b) Histogram of difference between $Prop - PI$ and $Rad - PI$. With bin size of 0.006 PI values. (4 years of data from 174 stations (Fig. 4.1))	82
4.7	(a). Histogram of difference between $PWV - GGOS - Prop$ and $PWV - GGOS - T_m$. (2 years of data from 384 IGS stations) (b) Histogram of difference between $PWV - IGS - Prop$ and $PWV -$ $IGS - T_m$. (2 years of data from IGS GPS stations (ref. Table A.1)). (c) Histogram of difference between $PWV - VLBI - Prop$ and $PWV -$ $VLBI - T_m$. (2 years of data from 15 VLBI stations). With bin size of 1 mm.	83
4.8	(a). Plot of PWV Vs DoY for different years. Green circle repre- sent PWV from GPS station using IGS database, black dots represent PWV from GPS station calculated from GGOS database, red stars represent PWV from VLBI data and blue triangles represent PWV from radiosonde data. 2 years of data from independent station of Japan, TSKB ($36.06^\circ N$, $140.05^\circ E$). [Best viewed in color]	85
5.1	Weather station located at NTU rooftop ($1.3^\circ N$, $103.68^\circ E$)	90
5.2	Time series plot of PWV (mm) for non-rainy days and rainy-days with rainfall rate (mm/hr) for IGS GPS station (NTUS). Time series for rainy-days are shown for 4 different days 5^{th} May 2010, 3^{rd} Oct 2011, 21^{st} May 2012 and 29^{th} Oct 2013. Time series for a non-rainy day is shown for 24^{th} Oct 2013. Here X-axis represents the hours of the day. The red dashed line is the constant PWV value at 55 mm.	92
5.3	(a) Histogram of PWV values during rain (b) Histogram of PWV values for no rain conditions. The red line with diamond markers represent year 2010, the blue line with star markers represent year 2011, the green line with circle markers represent year 2012 and the black line with square markers represent year 2013.[Best viewed in color]	94

5.4	Schematic diagram for a long term rainfall prediction. 6 hour PWV values from box A is used to predict rain within next 6 hours as indicated by box A1.	96
5.5	Flowchart of general rainfall prediction algorithm.	96
5.6	flowchart to illustrate the three-factor algorithm proposed for the sub-tropical region.	100
5.7	<i>TD</i> and <i>FA</i> rates for for different IGS-GPS stations indicated by individual figure title, when the monthly average PWV values are applied for the rainfall prediction. The details about the figure are given in the corresponding text. [Best viewed in color]	104
5.8	<i>TD</i> and <i>FA</i> rates when using different criteria for NTUS GPS station.	109
5.9	<i>TD</i> and <i>FA</i> rates when the threshold of maximum PWV values and the threshold of maximum increment in PWV values are used. . . .	112
5.10	<i>TD</i> rates for NTUS for (a) different length of data used (b) for different lead time values. [Best Viewed in Color]	115
5.11	Zoomed in version of Fig. 5.10, with two hours of time frame for both the figures (a) different length of data and (b) different lead time values. [Best Viewed in Color]	116
5.12	<i>TD</i> rates for SALU for (a) different length of data used (b) for different lead time values. [Best Viewed in Color]	116
6.1	Diurnal and seasonal variation in PWV and rainfall data for NTUS GPS station (2010 - 2013). The boxplots of PWV values with average of rainy (cyan line with circle markers) and non- rainy days (solid light green line) (a) NE, (b) SW and (c) inter-monsoons. Dashed black lines represent a threshold value of (a) 55 mm for NE, (b) 54 mm for SW and (c) 56.3 mm for inter-monsoons. The total number of rain events occurring at specific hours of the day (d) NE, (e) SW and (f) inter-monsoon seasons. The time for both GPS and weather station for NTUS station is recorded in local time format (UTC+8).	121

6.2	(a) Composite analysis of first derivative of PWV (mm/5min) and (b) second derivative of PWV (mm/(5min) ²). Red diamonds represent mean values of PWV derivatives for 208 rain events of year 2010, blue stars for 234 rain events of 2011, green circles for 219 rain events of 2012 and black squares for 252 rain events of 2013.	125
6.3	Illustration of data points considered for rainy and non-rainy cases. A window size of 6 data points (30 min) is used for both the cases. . .	129
6.4	Flowchart to explain the rainfall nowcasting algorithm.	130
6.5	Probability distribution of PWV values for SiReNT GPS stations SRPT, SNYU, SNUS and SSTS. [Best viewed in color]	134
6.6	Diurnal and seasonal variation in PWV and rainfall data for SALU GPS station (2016 - 2017). The boxplots of PWV values with average rainy (cyan line with circle markers) and non-rainy days (solid light green line) (a) Wet, (b) Dry Seasons. Dashed black lines represent a threshold value of (a) 54 mm for dry and (b) 45 mm for wet season. (c) total number of rain events occurring at specific hours of the day for wet and dry seasons (d) First derivative values of 2 years. (e) Second derivative values of 2 years. The time for both GPS and weather station for SALU station is recorded in UTC format.	135
7.1	(a) Sky Imager (WAHRISIS) (b) Sky image taken by WAHRISIS . . .	141
7.2	(a) X and Y components of <i>HoD</i> and (b) X and Y components of <i>DoY</i> [Best viewed in color]	145
7.3	Degree of correlation between various features [Best viewed in color].	147

7.4	Time series of different weather variables. (a) PWV (blue solid line) in mm and Rainfall Rate (stem plot in green) in mm/hr (b) Surface Temperature in °C (solid magenta line), Dew point Temperature in °C (dashed cyan line) and Relative humidity in % (Red dots) (c) Solar Radiation in watt/m ² . The X-axis for all the subplots represent the time of the day for four consecutive days : 28 th , 29 th , 30 th Nov and 1 st Dec, 2015. The sky images captured by WAHRIS are also shown for specific time of the day. [Best viewed in color]	149
7.5	Performance of an individual feature for (a) Rainfall classification (b) Rainfall Prediction. The <i>TD</i> values are represented by black and <i>FA</i> is represented by blue. D and N indicate results for Day and Night respectively. Each reading shows 95% confidence intervals. The left y axis shows values for <i>TD</i> and the right y axis shows values for <i>FA</i> . (<i>Best Viewed in Color</i>)	151
7.6	Test results (<i>TD</i> , <i>FA</i> and <i>A</i>) when different features are added one at a time. The <i>TD</i> values are represented by solid black line, <i>A</i> is represented by dashed black line and <i>FA</i> is represented by blue dots. Each reading shows 95% confidence intervals. The left y axis shows values for <i>TD</i> and <i>A</i> and the right y axis shows values for <i>FA</i> . (<i>Best Viewed in Color</i>)	153
7.7	Test results (<i>TD</i> , <i>FA</i> and <i>A</i>) when different features are removed one at a time. The <i>TD</i> values are represented by solid black line, <i>A</i> is represented by dashed black line and <i>FA</i> is represented by blue dots. Each reading shows 95% confidence intervals. The left y axis shows values for <i>TD</i> and <i>A</i> and the right y axis shows values for <i>FA</i> . (<i>Best Viewed in Color</i>)	154
7.8	<i>TD</i> and <i>FA</i> rates for different features combined with seasonal and diurnal features at higher lead time values (a) <i>PWV</i> , (b) <i>SR</i> , (c) <i>T</i> , (d) <i>RH</i> and (e) <i>DPT</i>	158
7.9	Accuracy at different lead time values	159

List of Tables

2.1	Accuracy requirements (expressed as standard error) for upper-air measurements for synoptic meteorology	9
2.2	Inter comparison of independent techniques for the measurements of PWV	18
2.3	GPS Signals	19
2.4	Dry and Wet coefficients for Niell Mapping Function (NMF)	24
3.1	Correlation coefficients (%) at different times of day; AM and PM . .	62
3.2	Average correlation coefficients at 0 & 99 % of cloud free conditions .	65
4.1	$(T_m - T_s)$ relationship for specific stations from literature	71
4.2	Derived $(T_m - T_s)$ relationships for different regions	73
4.3	c and d values for different regions.	76
5.1	TD and FA rates when using threshold of maximum rate of change of PWV values (mm/hr)	98
5.2	TD rates when using threshold of maximum variation in PWV values (mm)	102
5.3	TD and FA rates when the three factors are implemented for the Tropical and Sub-Tropical stations	106
5.4	TD and FA rates when threshold of maximum PWV (55 mm) values are used on data from the tropical region.	108

5.5	<i>TD</i> and <i>FA</i> rates when threshold of maximum PWV values and threshold of maximum rate of PWV increment are used on data from the tropical region.	111
5.6	Comparison of <i>TD</i> and <i>FA</i> rates when the three-factor method and the proposed methods are implemented for the Tropical stations. . .	113
6.1	True detection and false alarm rate when using the proposed algorithm for rain prediction for NTUS station	131
6.2	Comparison of <i>TD</i> and <i>FA</i> rates from literature and from proposed method.	132
6.3	Validation Of proposed algorithm for a station from Singapore and Brazil	136
7.1	Confusion Matrix	143
7.2	Benchmarking	156
A.1	IGS GPS stations and radiosonde stations	185
A.2	SiReNT station location and data availability	186
A.3	Weather stations	186
A.4	Summary of all instruments and related data availability	187

Abbreviations

CORS	Continuously Operating Reference Stations
DD	Double Differencing
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA-40	ECMWF Re-Analysis-40
FA	False Alarm
GAMIT	GPS Analysis at Massachusetts Institute of Technology
GGOS	Global Geodetic Observing System
GIPSY	GPS Inferred Positioning System
GMF	Global Mapping Function
GNSS	Global Navigation Satellite Systems
GPS	Global Positioning System
GPT	Global Pressure Temperature
GWMT	Global Weighted Mean Temperature
IR	Infrared
JPL	Jet Propulsion Laboratory
LEO	Low Earth Orbit
MF	Mapping Function
MODIS	Moderate Resolution Imaging Spectroradiometer
MPS	Multipath Stacking
nIR	near-Infrared
NMF	Niell Mapping Function
NRCan	Natural Resources Canada
NWM	Numerical Weather Model

PPP	Precise Point Positioning
PWV	Precipitable Water Vapor
RINEX	Receiver Independent Exchange
RMSE	Root Mean Square Error
RO	Radio Occultation
SiReNT	Singapore Satellite Positioning Reference Network
SWD	Slant Wet Delay
SWV	Slant Water Vapor
TD	True Detection
VLBI	Very Long Baseline Interferometry
VMF	Vienna Mapping Function
WMO	World Meteorological Organization
ZHD	Zenith Hydrostatic Delay
ZTD	Zenith Total Delay
ZWD	Zenith Wet Delay

Chapter 1

Introduction

1.1 Background and Motivation

The study of the water vapor in the earth's atmosphere is becoming popular in the remote sensing community for a variety of applications and domains especially in the fields of weather prediction [1,2]. As the amount of water vapor in the atmosphere is linked to the formation of rain clouds and rain, it is being studied as a useful parameter for rainfall prediction [3–5] and cloud micro-physics.

The water vapor in the atmosphere is generally expressed in terms of the total water vapor present in a vertical column of unit cross section, commonly known as precipitable water vapor (PWV). PWV values are commonly derived using the measurements recorded by the radiosondes, measurements from the microwave radiometers and from the satellite instruments. However, these techniques rarely provide sufficient temporal and/or spatial resolutions for all weather conditions. This is where the Global Positioning System (GPS) offers a compelling alternative. The GPS signal delays have been extensively used in the derivation of the PWV values.

With the rapid deployment of the GPS continuously operating reference stations (CORS) in the local and the regional scale, and with the rapid development of GLONASS (GLObal NAVigation Satellite System) and the new emerging satellite constellations like BeiDou, the use of the satellite signal delays can produce PWV

values with very high spatio-temporal resolution. The existence of ubiquitous GPS stations around the globe can generate a huge amount of PWV data with a good resolution which can be effectively used to solve some of the fundamental problems in the field of remote sensing like rainfall prediction, cloud detection, etc. Mostly the PWV data are being explored in the field of rainfall prediction. Extracting the valuable information from the PWV data for rainfall prediction applications is challenging. In this thesis, we extensively study the GPS derived PWV values and its usefulness in the rainfall prediction applications.

1.1.1 Objectives and Contributions

Our primary objectives are to use the GPS signal delays to derive the PWV values accurately and to develop a rainfall prediction algorithm for tropical region using the PWV values. To meet these objectives, firstly the data from GPS stations distributed over different regions; tropical, sub-tropical and temperate, are studied to properly understand the dynamics of the PWV values. Secondly, different parameters and its effects on the accuracy of the GPS derived PWV values are thoroughly studied. The GPS derived PWV values are compared against radiosonde PWV values to verify its accuracy. Thirdly, different existing rainfall prediction models are reviewed and an algorithm is proposed for the tropical region.

1.1.2 Major contributions

- A comprehensive analysis is done on the effect of using different mapping functions, different ZHD models and different satellite elevation cut-off angles on the GPS-derived PWV values. The main contribution of this work is that the different season/location specific ZHD models are proposed for tropical, sub-tropical and temperate regions. GPS-derived PWV values are validated by comparing to the radiosonde PWV values. The GPS-derived PWV values are also compared to the Moderate Resolution Imaging Spectroradiometer (MODIS)-PWV values [J2], [C5], [C6].

- A simplified model is introduced to accurately convert the zenith wet delays *ZWD* into the PWV values. Different than in the literature, the unique contribution of the proposed simplified model is its global applicability and the fact that it is based on the parameters like day-of-the year, station height and station latitude, which are easily available [J6], [C11].
- A comprehensive review is provided on the importance of GPS-PWV values for rainfall prediction and on different long-term rainfall prediction models that use GPS-PWV values. Not many literature have implemented the rainfall prediction algorithm using GPS-PWV values in a tropical climate. Therefore, the proposed model is best-suited for the tropical climate and is the first for the tropical region [J1], [C2], [C18].
- A rainfall nowcasting algorithm which uses the GPS PWV values to predict a rainfall event in the next 5 minutes given the data of the past 30 minutes is proposed for the tropical region. This method is novel as it addresses the rainfall nowcasting problem which has not been addressed in the literature. The proposed algorithm uses the second derivative method which has not been discussed before [J5].
- A data-driven approach is implemented to improve the accuracy of the rainfall prediction algorithms using GPS PWV values and other meteorological parameters like temperature, relative humidity, solar radiation and dew point temperature. The major contribution of this work is the analysis of the effect of different ground based weather parameters on rainfall. The proposed method is unique as different from the literature, it implements the machine learning algorithm to improve the rainfall prediction accuracy [J3], [C3], [C4].

1.2 Thesis Outline

Chapter 1: Introduction The objectives and motivation of the thesis are discussed in this chapter. We also describe the outline of the thesis, and briefly

highlight the main contributions of this thesis.

Chapter 2: Literature Review In this chapter, a thorough literature review of the various aspects of this thesis is given. A general account of the current state-of-the-art practice is reviewed and relevant research gaps are identified.

Chapter 3: Accuracy Assessment of the GPS-Derived PWV Values

In this chapter, PWV values are derived for stations from different regions and the effects of using different mapping functions and different ZHD models are studied. The GPS derived PWV values are compared with the ground truth; radiosonde PWV values. Moreover, a comparative study between the GPS derived PWV values and the MODIS PWV values is presented in this chapter.

Chapter 4: Conversion of Zenith Wet Delays into PWV In this chapter, different existing T_m - T_s models are reviewed briefly. A simplified model is proposed to convert the zenith wet delays into the PWV values. The proposed model performs with similar accuracy as the T_m - T_s models. It has an added advantage of being easily applicable anywhere in the globe as it depends only on the factors like station height, station latitude and day-of-the year. Moreover, it does not require the additional parameter like temperature.

Chapter 5: GPS-Derived PWV Values for Long-Term Rainfall Prediction In this chapter, the GPS derived PWV values are studied with respect to the rainfall events and the usefulness of PWV values for rainfall prediction is pointed out. Different rainfall prediction algorithms that are developed using data from the temperate and the sub-tropical regions are reviewed and are implemented using the data from the tropical region. An accurate long-term rainfall prediction algorithm is then proposed for the tropical region.

Chapter 6: GPS-Derived PWV Values for Rainfall Nowcasting In this chapter, the GPS derived PWV values are studied for rainfall nowcasting for the tropical region. Season dependent PWV threshold values are proposed and an algorithm is developed to nowcast a rain event in next 5 minutes given the data of the past 30 minutes.

Chapter 7: A Data-driven Approach to Improve the Rainfall Prediction Results In this chapter a machine learning based frame work is implemented

to accurately predict the rainfall events. Along with the GPS PWV values, different meteorological parameters like temperature, relative humidity, dew point temperature and solar radiation are used for this purpose. A comprehensive study on the importance of the individual feature on the rainfall prediction is given in this chapter.

Chapter 8: Conclusion & Future Work We conclude the thesis in this chapter. Future works on the applications of the GPS PWV values in the remote sensing fields are discussed. Following this chapter, we provide a detailed list of the various notations that are used throughout this thesis, for easy reference.

Chapter 2

Literature Review

2.1 Overview

This chapter provides a thorough literature review of different aspects that have been covered by this thesis. As mentioned in the previous chapter, this thesis is about the use of the GPS derived PWV values for remote sensing applications. This chapter presents the introduction to PWV and its significance to remote sensing applications. Different existing methodologies which are implemented to derive these PWV values are presented. A detailed review is done on the application of the PWV values in the rainfall prediction and few research gaps have been identified. In this thesis, we propose methodologies and algorithms to bridge most of the research gaps.

2.2 Precipitable Water Vapor and its Significance

The atmospheric water vapor is generally expressed in terms of Precipitable water vapor (PWV). PWV is the total atmospheric water vapor contained in a vertical column of unit area and is expressed in the unit of length; *mm*. PWV is mainly concentrated in the lowest 3 km of the atmosphere and its content generally increases with temperature. It is an important indicator of water vapor climatology and variability in the lower troposphere and related climate processes [6, 7]. It is strongly

linked to the hydrological cycle and dynamical processes in the tropical region where the overall PWV is high [8].

The earth's weather and climate are heavily influenced by the amount of water vapor and other greenhouse gases presented in the lower part of the neutral atmosphere; troposphere. The troposphere can contain a large volume of water vapor that traps radiant energy and thus causes the temperature to increase, resulting in global warming [9]. This is known as the greenhouse effect. A warmer climate increases the amount of water vapor, which further reduces the amount of long-wave radiation escaping from the earth to space, and thereby leading to an even warmer climate. Hence, the atmospheric water vapor is very important for the earth's climate system and a key to understand the hydrological cycle [10].

The large amounts of water vapor in the atmosphere in low-latitude regions contribute to many meteorological phenomena such as tropical storms, El Niño and La Niña [11]. It plays a crucial role for the development of hazardous cumulus convection [12] and is generally the source of convective clouds [13]. Its variation and distribution largely affect the formation and development of the convective precipitation systems. Thus, study of water vapor is very important for short-term and long-term weather forecasting and climate monitoring.

2.3 State-of-The-Art Techniques for Precipitable Water Vapor Retrieval

PWV values can be determined through the measurement data using various different instruments like radiosondes, microwave-radiometers, satellite-based instruments and GPS signal delays. The radiosonde and microwave radiometers are the traditional techniques [14–16], which are useful in validating the PWV values from GPS and satellite-based instruments. This section provides a detailed literature review on how these techniques are used as well as the state-of-the art of these techniques. An inter-comparison of the advantages and disadvantages of the individual instrument is also reported.

2.3.1 Radiosondes for PWV Retrieval

A radiosonde is an expendable, balloon-borne device, which is equipped with a radio transmitter that transmits the measured vertical profile of meteorological parameters to a land-based or sea-based receiving and processing stations. There are many radiosonde stations around the globe from where the radiosondes are generally launched twice a day (at 00 : 00 UTC and at 12 : 00 UTC). A radiosonde measures the meteorological parameters like pressure, temperature, relative humidity at different altitudes. These meteorological parameters from radiosonde are used in processing the precipitable water vapor.

Radiosonde Instrument

Figure 2.1: Radiosonde balloon with the radio transmitter

A radiosonde instrument as shown in Fig. 2.1 is equipped with a radio transmitter that transmits the measured vertical profile of meteorological variables to a land-based or sea-based receiving and processing station. During the 1 to 2 hours ascent of radiosonde, the sensors measure the variations of pressure, temperature, relative humidity with the changing altitude. Wind speed and wind direction are also obtained by use of a GPS or a radio direction finding antenna that tracks the

position of the balloon during the flight. It has been assigned two radio frequency bands 400-406 MHz and 1675-1700 MHz for transmitting radiosonde data. The meteorological measurements are made at intervals that vary from 1 to 6 s, depending on the type and manufacturer of the radiosonde. All of the world's radiosondes are required to meet certain performance standards that have been established by the WMO (ref. Table 2.1).

Variable	Range	Accuracy requirement
Pressure	Surface to 5 hPa	$\pm$ hPa
Temperature	Surface to 100 hPa	$\pm$ 0.5 K
	100 to 5 hPa	$\pm$ 1 K
Relative humidity	Troposphere	$\pm$ 5%(RH)
Wind direction	Surface to 100 hPa	$\pm$ 5° for wind speed < 15 m/s
		$\pm$ 2.5° for wind speed > 15 m/s
Wind speed	100 to 5 hPa	$\pm$ 5°
	Surface to 100 hPa	$\pm$ 1 m/s
	100 to 5 hPa	$\pm$ 2 m/s
Geopotential height of significant levels	Surface to 100 hPa	$\pm$ 1% near the surface decreasing to $\pm$ 0.5% at 100 hPa

Table 2.1: Accuracy requirements (expressed as standard error) for upper-air measurements for synoptic meteorology

The temporal and spatial resolution of the radiosonde is quite low. The radiosonde balloons are generally launched only twice a day at 00 : 00 UTC and 12 : 00 UTC and the radiosonde stations are not as densely populated as ground-based GPS stations.

PWV Retrieval from Radiosonde Profiles

This section describes about the mathematical procedures to derive the PWV values from the radiosonde data [17]. Water vapor pressure, e_{wv} in hPa, can be calculated using the radiosonde temperature and relative humidity data as shown in eq. (2.1).

$$e_{wv} = RH \times \exp(-37.2465 + 0.213166T - 2.56908 \times 10^{-4}T^2) \quad (2.1)$$

where, RH is relative humidity in percentage, T is the absolute temperature in Kelvin. By applying the gas state equation, the water vapor density ρ_v can be computed using eq. below.

$$\rho_v = \frac{e_{wv}}{R_v T} \quad (2.2)$$

PWV is then calculated by using eq. (2.3).

$$PWV = \frac{1}{\rho} \sum \frac{(h_{j+1} - h_j)(\rho_v^{j+1} - \rho_v^j)}{2} \quad (2.3)$$

where ρ is the density of water, h_j is the altitude at j^{th} level. j represents the different level of altitude during radiosonde flight. The maximum height that it can go is around 17 km in average. There are two main purposes of the upper-air soundings, one is to describe and analyze current weather conditions, and the other is to provide inputs to short and medium range computer-based weather forecast models. Other uses of radiosonde data include climate studies, air pollution investigations, and defense applications. In this thesis, radiosonde data are used in the estimation of PWV values.

Radiosonde Database

We use the radiosonde data from the radiosonde database maintained by the University of Wyoming [18]. There are more than 1000 operational radiosonde stations distributed all over the world [19]. Fig. 2.2 shows the distribution of most of the radiosonde stations maintained by Wyoming University. Each station has its own unique station number (radiosonde station Id). Each station records data twice a day (00:00 AM and 12:00 PM).

Radiosonde-Derived PWV Values

Use of radiosondes to derive PWV values is a traditional technique and it has been studied for a long time but it is important as radiosonde derived PWV values are

Figure 2.2: Location of the radiosonde stations in different regions. The black dots mark the location of the radiosonde stations.

generally used as the ground truth. Most of the literature show that the PWV derived from recent techniques like satellite or GPS is first validated against radiosonde derived PWV values before they are used in any applications. Therefore, the radiosonde derived PWV values are generally used in the validation of PWV measured by other methods like satellite and GPS [20–23]. Wang et. al [20], have shown the characterization of water vapor in the sub-tropical coastal regions of China by using the PWV values derived from GPS stations. This paper first validates the water vapor from GPS by comparing it to the PWV from radiosonde. This paper reports a root mean square error (RMSE) of around 3 mm between PWV derived from GPS and radiosonde. Gui et. al [21], present an evaluation of precipitable water vapor derived from radiosonde, satellite and GPS instruments using data from the temperate regions of China. This paper also compares the PWV values derived from GPS against the radiosonde PWV values. It reports a RMSE of 2.60 mm between GPS derived PWV and radiosonde derived PWV at 00 : 00 UTC and of around 3 mm at 12 : 00 UTC. In [22], the RMSE of up to 3 mm is reported between ra-

diasonde and GPS PWV values from a station located at the temperate region of Turkey. Similarly, [24,25] present the similar comparison between the GPS derived PWV values and the radiosonde derived PWV values with RMSE of 3.7 mm and 7.2 mm for stations from Japan and Korea respectively. The literature therefore, highlights that the PWV values from radiosonde correlate well with that of PWV values from GPS stations with RMSE values that varies for different locations. Here it is important to note that most of the works reported in the literature use data from the temperate and the sub-tropical regions. Limited work has been done for tropical region. Thus, in this thesis, along with the temperate and the sub-tropical regions, results corresponding to many tropical stations are presented and discussed.

Here it is clear from the literature that the radiosondes are widely used in the retrieval of PWV values to validate the new emerging technologies for PWV retrieval from GPS and satellite-based instruments. The major shortcoming of this instrument is its lower spatial and temporal resolutions inspite of its high operational costs. [14,23,26]. Due to its lower temporal resolution, very limited data are available for the understanding of different complex tropospheric phenomenon [26]. Moreover, the radiosondes are generally not released during major weather phenomenon like thunderstorms, hurricanes and heavy rain. Therefore, radiosonde data might be limited for studying different severe weather phenomenon [20].

2.3.2 Microwave Radiometers for PWV Retrieval

Microwave radiometer is a device that measures energy emitted at millimeter-to-centimeter wavelengths (frequencies of 1 - 1000 GHz) known as microwaves. This band includes both microwave frequency range and infrared frequency range. Majority of the active sensors operate in the microwave band (1-300 GHz). Active sensors provide their own source of energy to illuminate the objects they observe. Most passive systems used in remote sensing applications operate in the infrared, visible band (>300 GHz). Passive sensors detect natural energy (radiation) that is emitted or reflected by the object being observed. Reflected sunlight is the most common source of radiation measured by passive sensors. Microwave radiometers are usually

equipped with multiple receiving channels that helps in characterizing the emission spectrum of the atmosphere. Different important meteorological quantities like vertical temperature and humidity profile, precipitable water vapor and column liquid water path can be derived using microwave radiometers with high temporal resolution. The microwave radiometers are operated at a frequency of 23.8 and 31.5 GHz for measuring water vapor and cloud liquid water content respectively as water vapor and liquid water have the dominance in emission and absorption in this microwave region [27]. Similar to radiosondes, ground-based microwave radiometers are also a conventional technique used in derivation of PWV values [28, 29]. In recent literature, the microwave radiometers are generally used in verification of the water vapor derived from GPS and satellites-based instruments [20, 30, 31].

Although, the microwave radiometers have found to derive the PWV values with good accuracy and with a high temporal resolution, the ground-based microwave radiometers are restricted due to its poor spatial coverage. Microwave radiometers have sparse station distribution [16]. It has high instrument cost and hence it cannot be deployed easily [14].

PWV observations from microwave radiometers are still of limited value in climate studies particularly in predicting and tracking heavy rainfall cases [20]. Microwave radiometers can provide reliable PWV readings only under no rain-fall conditions [31]. They are not able to perform in all-weather conditions [16]. It might even require site or season specific calibrations [14, 15].

2.3.3 Satellite-Based Instrument for PWV Retrieval

MODIS Instrument

Moderate Resolution Imaging Spectroradiometer (MODIS) is a key instrument aboard two satellites; Terra and Aqua. The two satellites; Terra and Aqua orbit around the earth such that Terra passes from north to south across the equator in the morning, while Aqua passes south to north over the equator in the afternoon [32]. MODIS uses the near-infrared (nIR) algorithm for retrieving precipitable water vapor data for day time. Since the two satellites pass through any point on Earth twice a day,

the temporal resolution of the PWV values derived from MODIS is also twice a day similar to that of the radiosonde data. But the spatial resolution of MODIS is better compared to the radiosonde data as each pixel of the nIR MODIS PWV data cover an area of 1km by 1km and the MODIS satellites orbit all around the globe.

PWV and Cloud-Mask Retrieval from MODIS observations

Terra and Aqua satellites detect the electromagnetic radiation in 36 visible, near Infra-Red (nIR), and Infra-Red (IR) spectral bands between 0.645 and 14.235 μm and the swath width of the MODIS data for Terra and Aqua are 2300 and 2330 km, respectively [33,34]. The nIR MODIS PWV is derived from two non absorption channels centered near 0.905 and 0.94 μm and three absorption channels centered near 0.865, 0.936, and 1.24 μm and is available during the daytime only at a 1 km by 1 km spatial resolution [34,35]. The IR MODIS PWV is derived from bands between 4.47 and 14.24 μm , and is available for both day and night with a 5 km by 5 km field of view. The MODIS IR PWV products have lower accuracy compared to the MODIS nIR PWV products [36] and thus in this thesis, we use the nIR MODIS products for further discussion and analysis.

The nIR MODIS-derived PWV values are stored as the MODIS Level-2 PWV products of collection-5 from Terra (MOD05_L2) and from Aqua (MYD05_L2). The MODIS level-2 collection-5 also includes the cloud mask products. MODIS retrieved cloud mask product has 4 levels : 192, 128, 64 and 0. 192 indicates that the MODIS pixel has a probability of 99% of being cloud free, 128 indicates a probability of 95 % of being cloud free, 64 indicates a probability of 66 % of being cloud free and 0 indicates 0 % of being cloud free (i.e. fully cloudy condition).

MODIS Database

The MOD05_L2, MYD05_L2 and the geo-location products can be downloaded from the given link at [32]. A python script is written to process these products and derive the PWV values and the cloud mask values. Given the desired coordinates, the script file processes the PWV values and the cloud mask values taking the given

coordinates as the center point, with spatial resolution of 1 km by 1 km around the coordinates and temporal resolution of 2 data points per day (TERRA and AQUA). We process the MODIS data for a number of stations covering different regions. As we compare the MODIS derived PWV values with the PWV values derived from GPS stations, we take the GPS station coordinates to process the MODIS PWV and cloud mask values.

MODIS-Derived PWV Values

It has not been long since researchers have started to study about the applications of the MODIS derived PWV values. Therefore, recent literatures show many ongoing works that compare the MODIS derived PWV values to the PWV values derived from radiosonde and GPS [21–23, 37]. Gui et. al [21], have analyzed the temporal and spatial distribution of MODIS derived PWV values using data from years 2011–2013 in China. The inter-comparison between the PWV derived from radiosonde, GPS and MODIS showed that the RMSE between the MODIS-derived PWV values and GPS-derived PWV values is around 5.76 mm and that between MODIS-derived PWV values and radiosonde-derived PWV values is around 5.31 mm for clear sky conditions. The RMSE between the GPS-derived PWV and the radiosonde-derived PWV is around 2.60 mm. In general, the result showed that the MODIS-derived PWV values were higher than the PWV values derived from radiosonde and GPS. In another study [22], PWV variations obtained from 6 year GPS observations in Turkey are compared with the PWV values obtained from MODIS observations. The RMSE difference of about 3 mm to 7 mm is reported for cloud-free conditions. In [34], based on 17 months of data from Southern California a RMSE of 4 mm has been reported. Similarly, in [38], the evaluation of MODIS PWV against the GPS PWV values showed a range of RMSE from 2 mm to 8 mm for stations from different regions of China.

From the literature it can be inferred that higher RMSE differences exist between MODIS and GPS PWV values compared to the GPS and radiosonde PWV values. The MODIS-derived PWV values generally over estimates the PWV values from

GPS. The overestimation is more prominent at higher range of PWV values [39]. Here, it is also important to note that the MODIS-derived PWV values are accurate only when the sky is cloud free (clear) [21]. The PWV values derived from MODIS during the cloudy condition are ambiguous and not useful. In this thesis, we present the PWV values derived from MODIS for different stations from the temperate, the sub-tropical and the tropical regions and compare with the corresponding PWV values from GPS.

2.3.4 GPS Signals for PWV Retrieval

Figure 2.3: IGS GPS Stations distributed around the globe. The green and red dots mark the IGS stations. Image Source: <http://www.igs.org/network>.

Literature have shown both advantages and shortcomings of using radiosondes, microwave radiometers and MODIS for PWV retrieval. Radiosondes and microwave radiometers are limited by their spatio-temporal resolutions and operational/maintenance cost. Similarly, MODIS can give higher spatial resolution but it

has poor temporal resolution. Most importantly, these instruments have limitations in working under all weather conditions. Radiosondes and microwave radiometers are not useful for severe weather conditions like hurricanes and heavy rain and the PWV data from MODIS are not accurate for cloudy conditions. Therefore, PWV values from these instruments are biased towards a particular weather condition only. To overcome the drawbacks of these systems, GPS signal is extensively being used to retrieve the PWV values for many applications [1, 6, 31, 40–42].

IGS GPS Database

There are in total more than 500 IGS (International Global Navigation Satellite System Service) stations that are distributed all around the globe as shown in Fig. 2.3 [43]. Each IGS GPS station has its unique station ID. Each station records its RINEX (Receiver Independent Exchange) observation files. The RINEX observation files can be downloaded from the ftp site [44].

GPS-Derived PWV Values

The PWV values can be derived from the GPS signal delays at an interval of 5 minutes. Along with a good temporal resolution, the spatial resolution of the GPS stations are also better. More details on GPS-PWV retrieval is given in the following sections.

There is a dramatic change in the field of GNSS with the rapid development of GLONASS (the Russian GLObal NAVigation Satellite System), and newly emerging constellations, like the Chinese BeiDou Navigation Satellite System and the European Galileo system. With the rapid development of the current GNSS constellation, there is a promising prospect for the real-time retrieval of tropospheric parameters like wet delays and PWV [45, 46]. Therefore, GPS meteorology offers an improved spatial and temporal resolutions for water vapor retrieval compared to the traditional techniques [16, 20]. GPS is widely being used as an all-weather, low-cost remote sensing instrument for weather forecasting and climatology. Particularly in the field of remote sensing and weather forecasting, the GPS-derived PWV has been used

for the analysis of severe weather conditions such as storms, floods [42, 47], heavy rainfall events monitoring, rainfall now-casting, cloud micro physics and dynamics studies [1, 3, 4, 31, 48].

Table 2.2, summarizes and inter-compares the advantages and disadvantages of different technologies. The comparison between the existing technologies concludes that GPS is an emerging technology for the retrieval of PWV values with higher spatio-temporal resolutions and its advantages out weigh its disadvantages. The next section gives a detailed review on derivation of PWV values from GPS signal delays.

Parameters	Radiosonde	Radiometer	MODIS	GPS
Temporal Resolution	Low	High	High	High
Spatial Resolution	Low	Low	High	High
Vertical Resolution	High	Low	Low	Low
All Weather Applicability	Low	Low	Low	High
Cost	High	High	Low	Low

Table 2.2: Inter comparison of independent techniques for the measurements of PWV

2.4 GPS Signal Delays for PWV Estimation

2.4.1 GPS

The NAVSTAR Global Positioning System has been developed by the Department of Defense of the United States. The main purpose is to provide the military forces of the USA and its allies with a means to accurately determine their position, velocity and time, anywhere on the Earth, at any time. It operates worldwide in all weather conditions 24 hours a day. The GPS system consists of three major segments: SPACE, CONTROL and USER. The SPACE segment currently consists of 28 operational satellites in six orbital planes with four satellites in each plane spacing evenly and several spare satellites. The satellites operate in circular orbits of approximately 26,559.7 km radius at an inclination angle of 55 degrees, with about a period of 12-hours [49]. The GPS constellation as shown in Fig. 2.4 provides

Figure 2.4: GPS satellite constellation: 24 satellites in six orbit planes a global coverage with four to eight simultaneously observable satellites above an elevation angle of 15° .

The satellite positions in the sky are the same at the same sidereal time, each day, i.e. the satellites appear 4 minutes earlier each day in civilian time. The CONTROL segment consists of five Monitor Stations (in Hawaii, Kwajalein, Ascension Island, Diego Garcia, Colorado Springs), three Ground Antennas (in Ascension Island, Diego Garcia, Kwajalein) and a Master Control Station (MCS) located at Schriever Air Force Base in Colorado. The monitor stations track all satellites in view, accumulating the information. This information is then processed at the MCS to determine satellite orbits, clock errors and to update each satellite's navigation message. Updated information is then transmitted to each satellite via the ground antennas. The USER segment consists of antennas and receiver-processors that provide positioning, velocity, and precise timing to the users. The GPS satellites transmit at two frequencies: 1.575 GHz (designated as L_1) and 1.228 GHz (designated as L_2). Codes are modulated onto these carriers. The precision (encrypted) or P(Y) code is modulated on both the frequencies, whereas the coarse-acquisition (C/A) code is modulated only on L_1) (ref. Table 2.3).

Carrier	C/A code	P code
L1: 1575.42 MHz (19.0 cm)	1.023 MHz (293 m)	10.23 MHz (29.3 m)
L2: 1227.60 MHz (24.4 cm)	Not modulated	10.23 MHz (29.3 m)

Table 2.3: GPS Signals

2.4.2 GPS Signal Delays

When GPS signals travel from a GPS satellite to a receiver on the ground, the signal is delayed by a number of parameters in atmosphere. This total atmospheric delay is mostly caused by the ionosphere and troposphere of the earth as explained by the flowchart in the Fig. 2.5. The delay in the ionosphere is generally caused by the ions and electrons present in the ionosphere region and the ionospheric delay is dispersive in nature i.e. it is frequency dependent [50]. Therefore, ionospheric delays are determined by using the linear combination of ionospheric free equations which requires observations from both the frequencies transmitted by GPS satellites (L_1 at 1.575 GHz and L_2 at 1.228 GHz) [50–53].

Figure 2.5: Total atmospheric delay

Now, the remaining delay is caused by the troposphere and is called as the neutral delay. The total neutral delay in the zenith direction is called as zenith total delay (ZTD). The ZTD is non-dispersive in nature and it cannot be estimated like that of the ionospheric delays [50]. The ZTD can be divided into two parts: zenith hydrostatic delays (ZHD) and zenith wet delays (ZWD) as shown in Fig. 2.5. The ZHD contributes about 90 % of the total zenith delay [1, 54] and it generally depends on the surface pressure (P_s), temperature and refractive index of the troposphere [55] and empirical models are available to estimate it [56]. Unlike ZHD , it is difficult

Figure 2.6: Atmospheric delays mapped to the zenith direction.

to estimate the ZWD s because the wet delays are a function of the atmospheric water vapor profile and humidity, which is random and changes with temperature. eq. (2.4) shows the simple expression of the zenith delays [50].

$$ZTD = ZHD + ZWD \quad (2.4)$$

The ZTD at any point of time is obtained by projecting the slant total delay at an elevation angle (e), to the zenith direction using mapping functions that depends on the elevation angle of the satellites. Fig. 2.6 shows a simple illustration of slant and zenith delays. In Fig. 2.6, the elevation angles of different slant paths are represented by e_1 , e_2 and e_3 . As shown by Fig. 2.6, at any time, the delay in the zenith direction represents an average of a number of slant path delays projected in the zenith direction.

$$STD(e) = MF_h(e) \cdot ZHD + MF_w(e) \cdot (ZWD + \cot(e)(G_N \cdot \cos\phi + G_E \cdot \sin\phi)) + \epsilon \quad (2.5)$$

eq. (2.5) [57–59] formulates the delay in slant direction (STD). Here, e and

ϕ represent the elevation and the azimuth angles of the GPS satellites. MF_h and MF_w are the dry and the wet mapping functions respectively to map the delays at an elevation angle, e to the zenith direction or vice-versa. G_N and G_E are the delay gradient parameters in the north and the east directions respectively. ϵ is the postfit phase residuals. These postfit phase residuals include all unmodeled phase errors which are not considered in GPS analysis like multipath and satellite orbit errors along with the atmospheric inhomogeneity which is caused by the presence of water vapor. It is important to remove such errors like multipath for the *STD* analysis, except for those due to atmospheric inhomogeneity.

2.4.3 Mapping Functions

In simple terms, the mapping functions can be defined as the ratio of the path delay at a particular direction at an elevation angle (e) to the path delay in the zenith direction. [54]. In literature, there are two methods adopted to model the mapping functions; a) based on actual meteorological data b) based on empirical models. The mapping functions based on the actual meteorological data use information from ray-tracing through numerical weather models (NWMs) at certain times and locations. On the other hand, the mapping functions based on empirical models rely on statistical climatological values [60, 61]. The accuracy of the empirical models are generally lower compared to those based on the real observation but yet the empirical models are frequently used mostly because the empirical models can be easily implemented, faster to process and do not require internet connections for downloading the latest observation data [60]. Some of the commonly used mapping functions found in the literature are Niell mapping function (NMF) [42, 62], Global mapping function (GMF) [56], Global pressure temperature 2 mapping function (GPT2) [60] and Vienna mapping function 1 (VMF1) [63]. Out of these mapping functions, NMF, GMF and GPT2 are the empirical models and VMF1 is based on the real observation data. All of these mapping functions have the same basic equation as shown in eq. (2.6) [54, 60, 61, 64], which was developed by Marini [65].

$$MF_i(e) = \frac{1 + \frac{a_i}{1 + \frac{b_i}{1 + c_i}}}{\sin(e) + \frac{a_i}{\sin(e) + \frac{b_i}{\sin(e) + c_i}}} (i = h, w) \quad (2.6)$$

Here, a_i , b_i and c_i are the coefficients for hydrostatic and wet functions; $i = h$ for hydrostatic and $i = w$ for wet mapping functions. The coefficients a_i , b_i and c_i define the dependency on refractivity through the atmosphere. In other words, they give a measure of the thickness of the neutral atmosphere [66]. As the coefficients decrease the mapping function approaches $1/\sin(e)$. Different mapping function models differ on how the coefficients a_i , b_i and c_i are calculated.

Niell Mapping Function (NMF)

Prior to NMF, the coefficients a_i , b_i and c_i were estimated using surface measurements like surface temperature and pressure. Since surface measurements were not readily available everywhere, empirical method like NMF was developed to estimate the coefficients. The parametrization of NMF is based on ray traces of radiosonde profiles spanning the latitudes 43°S to 75°N. By assuming longitudinal homogeneity and symmetry between the southern and the northern hemisphere, this mapping function was extended globally. The equatorial region from 15°S to 15°N is defined by the estimates of the 15°N latitude. The polar regions with latitudes higher than 75° on the northern and southern hemisphere are both approximated by the 75°N latitude values. The hydrostatic coefficients are calculated by using eq. (2.7).

$$a_h(L_a, t) = a_{avg}(L_a) - a_{amp}(L_a) \cos(2\pi \frac{t - 28}{365.25}) \quad (2.7)$$

Here, L_a is the latitude, t is the day-of-year, a_{avg} represent the average value and a_{amp} represent the annual amplitude. A similar procedure is followed for parameters b_h and c_h . However, for wet coefficients only an interpolation in latitude parameter is needed. The typical values for the NMF coefficients are shown in Table 2.4 [61, 67].

Coefficient	Latitude (deg)				
	15	30	45	60	75
Dry Coefficients					
a_{avg}	1.2769934e-3	1.2683230e-3	1.2465397e-3	1.2196049e-3	1.2045996e-3
b_{avg}	2.9153695e-3	2.9152299e-3	2.9288445e-3	2.9022565e-3	2.9024912e-3
c_{avg}	62.610505e-3	62.837393e-3	63.721774e-3	63.824265e-3	64.258455e-3
a_{amp}	0.0	1.2709626e-5	2.6523662e-5	3.4000452e-5	4.1202191e-5
b_{amp}	0.0	2.1414979e-5	3.0160779e-5	7.2562722e-5	11.723375e-5
c_{amp}	0.0	9.0128400e-5	4.3497037e-5	84.795348e-5	170.37206e-5
Wet Coefficients					
a_w	5.8021897e-4	5.6794847e-4	5.8118019e-4	5.9727542e-4	6.1641693e-4
b_w	1.4275268e-3	1.5138625e-3	1.4572752e-3	1.5007428e-3	1.7599082e-3
c_w	4.3472961e-2	4.6729510e-2	4.3908931e-2	4.4626982e-2	5.4736038e-2

Table 2.4: Dry and Wet coefficients for Niell Mapping Function (NMF)

Here, it can be noted that the coefficients of the continued fraction representation of the hydrostatic mapping function depend on the latitude and the day-of-year; the dependence of the wet mapping function is only on the site latitude. The coefficients were derived with an assumption of symmetry between the southern and the northern hemisphere. Therefore, one of the major drawbacks of NMF is its biasness towards the northern hemisphere.

Vienna Mapping Function (VMF)

VMF follows the same basic equation as shown in eq. (2.6). The significant improvement in VMF over NMF is that the coefficients a_h and a_w are fitted to raytracing with the Numerical Weather Model (NWM) of the European Centre for Medium Range Weather Forecast (ECMWF). NWMs provide a highly accurate description of the atmospheric state every 6 hours with a spatial resolution of 2.5° in longitude and 2° to 2.5° in latitude. The coefficients a_h and a_w are available in the grid format for latitude and longitude values (2.5° by 2°) at ftp site [68]. The coefficients are provided by the Global Geodetic Observing System (GGOS) service and maintained by the Vienna Technical University. VMF retains the functional representations for the remaining hydrostatic and wet coefficients (b_i, c_i) as that of the NMF. An up-

dated version of the VMF is VMF1, which has improved empirical representation of the coefficients (b_i, c_i) . The hydrostatic and wet coefficients of VMF1 are given as: $b_h = 0.002905$; $c_h = 0.0634 + 0.0014\cos(2 L_a)$; $b_w = 0.00146$; $c_w = 0.0439$ [69].

Global Mapping Function (GMF)

It should be noted that the original VMF1 parameters are site-dependent and only available for particular sites. This site-dependency represents a significant limitation of VMF1. To overcome this, an empirical VMF1 function which can be globally applicable is introduced. This new mapping function is the Global Mapping Function (GMF), which is a backup VMF1 function. The coefficients of the GMF are obtained from an expansion of the VMF1 parameters into spherical harmonics on a global grid. Similar to NMF, the values of the coefficients require only the station coordinates and the day of year as input parameters. The parameterization of the coefficients has been refined to include a dependence on longitude. Compared to the 6-hourly values of the VMF1 a slight degradation in short-term precision occurs when using the empirical GMF [70].

Global Pressure Temperature Mapping Function (GPT2)

GPT2 uses 10 years (2001–2010) of 37 monthly mean pressure level data from the ECMWF Re-Analysis (ERA-Interim) to determine mean values (A_0) as well as annual (A_1, B_1) and semiannual amplitudes (A_2, B_2) for any selected parameter, r on a regular 5° grid [60, 71]. The equation is as shown in eq. (2.8). Here DoY is the day-of-year. GPT2 provides values for the hydrostatic and wet mapping function coefficients, pressure, temperature, lapse rate and water vapour pressure. The GPT2 derived parameters can be found at the ftp site [68].

$$r(t) = A_0 + A_1\cos\left(\frac{\text{DoY}}{365.25}2\pi\right) + B_1\sin\left(\frac{\text{DoY}}{365.25}2\pi\right) + A_2\cos\left(\frac{\text{DoY}}{365.25}4\pi\right) + B_2\sin\left(\frac{\text{DoY}}{365.25}4\pi\right) \quad (2.8)$$

Chapter 3 of this thesis presents a comparative study of effect using different mapping functions on the derived GPS-PWV values.

2.4.4 Empirical Model-Based Estimation of ZHD Values

As mentioned earlier ZHD delay contributes almost 90 % of the total zenith delay. At maximum the ZHD values can reach up to 2.3 m [72]. A typical variation of 1.85 m to 2.05 m in ZHD values have been reported in [73] for temperate stations of Iran. This section reviews about three models which are generally used in deriving the zenith hydrostatic delays: Saastamonien equation, VMF1 model and Static model.

Saastamonien equation is one of the empirical models that uses the pressure and temperature values to calculate the hydrostatic delays. eq. (2.9-2.10) show the Saastamoinen equation [1, 6, 74, 75].

$$ZHD = ((2.2799 \pm 0.0024) \cdot P_s) / f(\phi \cdot H) \quad (2.9)$$

$$f(\phi \cdot H) = 1 - 0.00266 \cdot \cos(2\phi) - 0.00028H \quad (2.10)$$

Here, P_s is the surface pressure (hPa), H is the ellipsoidal height of the station (km) and ϕ is the latitude of the station. Basically, to calculate the ZHD values, the Saastamonien equation requires only the surface pressure values. Accurate surface pressure values can be derived from weather stations or radiosondes. But for many GPS stations, co-located weather station or radiosonde station might not be available. Therefore, alternatively pressure values derived from the empirical models like GPT or GPT2 are used in eq. (2.9) to derive the ZHD values [1, 74–76]. VMF1 models are also used in deriving the hydrostatic delays [76, 77]. Along with the VMF1 mapping coefficients, the VMF1 hydrostatic delays are provided in the GGOS website [68]. The ZHD values from GGOS are based on the meteorological data from NWMs [76].

The ZHD values can also be calculated using a static model. The static model

is as shown in eq. (2.11) [57, 59, 78].

$$ZHD = 1.013 \cdot 2.27 \cdot e^{0.000116h} \quad (2.11)$$

As shown by the eq. (2.11), the static model is based on station altitude (h). The Saastamonien equation shows the use of both station height and pressure values, but the static model is more simplified and uses only height factor to deduce the ZHD values. In Chapter 3 of this thesis, we calculate the ZHD values using all three models Saastamonien (GPT+Saastamonien and GPT2+Saastamonien), VMF1 and Static and the results are subsequently inter-compared and discussed.

2.4.5 PPP-Based Estimation of ZWD Values

The zenith wet delay contributes only 10 % of the total zenith delay [1, 50, 54]. However, it is an important parameter for the retrieval of PWV values. Unlike the zenith hydrostatic delays, the wet delays cannot be estimated easily as the wet delays change rapidly over time and space. In literature, there are no accurate models for determining the wet delays. Therefore, wet delays are estimated as unknowns along with other parameters like station coordinates and receiver clock-offsets during the GNSS processing [6, 56, 79]. In literature, two different modes are commonly used for processing these observations: 1) double-difference (DD) solution mode and 2) Precise Point Positioning (PPP) mode [6, 48, 80]. In DD mode, GPS observations are simultaneously measured by two or more receivers. Out of these receivers, one is known as the reference station. The reference station calculates its geometric location by using the satellite positions, which are then used to estimate the unknown parameters. The PPP technique is based on the absolute position solution where the GPS observations of a single receiver are used to directly estimate the tropospheric delays. PPP requires the a-priori information like precise satellite orbit and clock products, which are made available from different sources like the International GNSS service (IGS), Jet Propulsion Lab (JPL) and Natural Resources Canada (NRCAN) [48, 57]. PPP can achieve cm-level positioning accuracy by just using single receiver. Since, PPP does not need the tracking data from a reference station or a reference network,

its equipment cost is cheaper and it is computationally more efficient compared to the DD technique [57,81]. Therefore, recent literature show the use of PPP technique over the DD technique for calculating the ZWD values [20,48,56,79,81–83].

There are different software packages that implement the PPP technique to estimate the ZWD values. Some of them are GNSS-Inferred Positioning SYstem, GIPSY by JPL, Bernese software, GPS Analysis at Massachusetts Institute of Technology (GAMIT), Canadian Spatial Reference System (CSRS-PPP service) by Natural Resources Canada, MagicGNSS operated by GMV Aerospace and Defense, gLAB by Catalonia Technical University and the European Space Agency, POINT by University of Nottingham, and RTKLIB by Takatsu [57]. All these softwares use the same strategy to estimate the ZWD . The difference is that different softwares might use different default models to calculate the ZHD , use different default mapping functions to calculate the delay in the zenith direction and different default models to correct the ionospheric effects. Based on such differences, the performance of the software might differ [57]. But for these software packages the settings can be modified to make it work using specific variables. For example, the default ZHD values for GIPSY version 6.4 is calculated using the Static ZHD model (ref. eq. (2.11)), which is a function of the station ellipsoidal height only. But GIPSY has a function called *tropnominal* which can be used to input different other models to calculate ZHD . Similarly, the default mapping function for GIPSY version 6.4 is NMF, but it has a *tropmap* function that allows users to choose between different mapping functions like NMF (default), GMF, GPT2 and VMF mapping functions. Therefore, the choice of the software might be determined by the ease of its availability and the constellation(s) (GPS, GLONASS, BeiDou, Galileo) that it is applicable for. In this thesis, we use GIPSY software version 6.4 by JPL, which is being used by many researchers in this field [77,84–86] as it is easily available and is free for educational purposes, it is well documented and it is suitable for GPS constellation which has more than 500 ground receivers all around the globe.

2.4.6 PWV from ZWD Values

After the *ZWD* values are retrieved from the GPS signal delays, they can be converted to the PWV values by using eq. (2.12).

$$PWV = PI \cdot ZWD \quad (2.12)$$

where,

$$PI = \frac{10^6}{\left(\frac{k_3}{T_m} + K'_2\right) \cdot R_v} \quad (2.13)$$

Here, *PI* is a dimensionless constant derived by using eq. (2.13). A few examples in the literature use symbol, *Q* and also use symbol $\bar{K}$ [87,88] to denote this constant and many use symbol π to denote it. For this thesis we use symbol *PI* to represent this conversion factor [89,90]. Using *PI* instead of π has an advantage of not being mistaken with the mathematical symbol π with a value of 3.14. Here, in eq. (2.13), R_v is the specific gas constant for water vapor of 461.5181 J/kg·K, K_3 and K'_2 are the refractivity constants of $(3.739 \pm 0.012) \cdot 10^3 \text{ k}^2/\text{Pa}$ and $(22.1 \pm 2.2) \cdot 10^{-2} \text{ K}/\text{Pa}$, respectively [50]. T_m (in Kelvin) is the weighted mean temperature defined by eq. (2.14).

$$T_m = \frac{\int \frac{e_{wv}}{T} dz}{\int \frac{e_{wv}}{T^2} dz} \quad (2.14)$$

Here, e_{wv} is the water vapor pressure and T is the air temperature. These parameters are traditionally calculated using data from the radiosonde sounding profiles. But as the radiosondes have very low temporal resolution, the T_m values are generally calculated using the surface temperature information (T_s). When T_m values are predicted using the T_s values, the relationship becomes site-specific as the surface temperature varies as per the geographic region and seasons. Therefore, researchers

have also looked into empirical models to model the T_m values.

($T_m - T_s$) Relationships

The most commonly used (T_m-T_s) relationship was reported by Bevis et al. [50] as,

$$T_m = a + b \cdot T_s \quad (2.15)$$

where $a = 70.2$ and $b = 0.72$.

It is noted that eq. (2.15) was derived based on the analysis of 8718 radiosonde profiles spanning approximately a two-year period from sites within the United States with a latitude range of 27° to 65° and an altitude range of 0 to 1.6 km [50]. This equation has been used by many researchers working in this field [91–93], notwithstanding researchers from the tropics and Antarctica regions. But studies have shown that the accuracy of the Bevis relationship in eq. (2.15) is found to be latitude dependent. It tends to overestimate T_m by up to 5 K at mid and high latitudes and tends to underestimate T_m by up to 6 K at low latitudes [94]. Therefore, instead of using eq. (2.15), some researchers were aware that the coefficients a and b in eq. (2.15) are region and season specific, and thus proposed their own (T_m-T_s) relationships [4, 88, 95–97].

Empirical models for T_m

Alternative to the T_m-T_s relationships, empirical values of T_m are also available. The GGOS database provides the gridded data of the global weighted mean temperature with a temporal resolution of 6 hours (at UTC 00 : 00, 06 : 00, 12 : 00, and 18 : 00) and a spatial resolution of 2.5° in longitude and 2° in latitude, using data from ECMWF. The GGOS T_m data can be accessed from [98].

Similarly, there are empirical models for T_m values [99] which are derived based on the temperature values estimated by the GPT and/or GPT2 model (which gives the surface temperature and pressure values based on the station location and day-

of-year parameters). Yao et. al have derived a global weighted mean temperature (GWMT) model based on GPT model and Bevis equation (eq. (2.15)) [100]. The radiosonde database was used for the derivation of this model. But because of the lack of the radiosonde stations near the sea, the GWMT model (also known as GTm-I) was reported to show abnormal results in some areas [101]. Therefore, Yao et. al have subsequently updated the GTm-I model with simulated T_m values over sea and included annual, semi-annual cycles of T_m values and have proposed the GTm-II and GTm-III models respectively [101, 102]. But these models do not consider the effect of altitude on T_m and since these models need the surface temperature values (T_s), they cannot be used when ground-truth temperature values are absent [103].

Therefore, in this thesis, apart from studying the $T_m - T_s$ models for different regions, we also study the behavior of PI values (ref. eq. (2.13)) based on different location, season and altitude and propose a simplified PI model for the conversion of ZWD values into PWV.

2.4.7 GIPSY Processing for Estimating PWV

This section describes the overall processing done in GIPSY software to retrieve the wet delays.

ZWD Retrieval from GPS Signal Delays

Fig. 2.7 shows a flowchart of how the PWV values are derived from the GPS signals. As shown by the flowchart, for the processing, RINEX files are required. RINEX files consist of GPS signal carrier and pseudo range information. The RINEX files are stored in the ftp server [44]. The server has daily RINEX files for all the IGS GPS stations. After the RINEX files are downloaded, TEQC is used to check the quality of the RINEX files and to merge the RINEX files. TEQC is one of the powerful tools for pre-processing the satellite observables. In our processing, RINEX files for each day is processed by merging three RINEX files; RINEX file from the corresponding day, RINEX file from one day before and RINEX file from one day after. This is done to ensure that there is no jump in the data when the days are changed. TEQC

Figure 2.7: Schematic representation of PWV retrieval using GIPSY OASIS software.

is used in merging the RINEX files. These RINEX files are then used by the GIPSY v-6.4 software to process the tropospheric delays. We run GIPSY in PPP mode and we know from literature that for PPP mode GIPSY requires the accurate satellite orbit information and clock error information. GIPSY gets these information during its processing from the anonymous ftp server [44] which is maintained by JPL.

GIPSY also requires information on ionospheric models, ocean tide loading models, mapping functions, *a-priori* ZHD values and elevation cut-off angles. We use the recommended script *gd2p* and recommended settings for these [78]. We apply the IERS2010 for the solid Earth tides [104] and FES2004 model for ocean tide loading [105]. The default elevation cut-off angle (ref. Fig. 2.6) for GIPSY is 7° but we can use the *w_elmin* function of GIPSY to set different values of cut-off angles. Generally, elevation cut-off angle of 10° is used [48, 95]. For our processing, we use 10° cut-off angle unless mentioned otherwise. GIPSY uses NMF as the default mapping function but the *trop_map* function of GIPSY can be used to run the scripts using different mapping functions like NMF, GPT, GMF and VMF1. We run the

gd2p scripts with different mapping functions and inter-compare the results.

To process the wet delays, GIPSY also requires *a-priori* ZHD values. As we know from the literature, GIPSY uses the static ZHD model eq. (2.11), as a default, to calculate the *a-priori* ZHD values. *tropnominal* function of GIPSY can be used to run the *gd2p* script with *a-priori* ZHD values derived from different models like GPT, GPT2 or VMF1. GPT and GPT2 use the saastamonien equations; eq. (2.9) and eq. (2.10), for which the values of pressure are derived from the GPT and GPT2 models which use the observation data whereas VMF1 uses the data from NWMs to deduce the ZHD values. We use all these different ZHD model during our processing and analyze its effect on the retrieved PWV values. Once the *a-priori* ZHD values are given, GIPSY processes the ZWD values. Now these ZWD values are converted to PWV by multiplying with a dimensionless constant factor, PI as shown by eq. (2.12). PI values are generally calculated using the mean weighted temperature T_m . These mean weighted temperature can either be derived from radiosonde profiles or $T_m - T_s$ linear regression or from the T_m models [98]. In this thesis, we propose a simplified PI model that is globally applicable and have similar accuracy as the temperature dependent model but is easier to implement.

ZWD from Tropospheric Products

Alternatively, the PWV values can also be processed from the delays published by the NASA website. The ZTD values are stored in the ftp server [106]. The values are calculated using the Bernese software which processes the RINEX observation files every 5 minutes for all IGS GPS stations. The software uses the GMF mapping function and minimum cut-off angle of 7° . Now the ZWD can be calculated from the ZTD by simply subtracting the hydrostatic delays (eq. (2.4)). We can process the ZHD from any of the models described previously. Use of the ZTD values directly from the ftp server saves a lot of time as we do not need to process the RINEX files using GIPSY software. But using the GIPSY software to process the RINEX files by ourselves, develops our understanding on how the observables are processed. Moreover, by using the GIPSY software, we can alter many settings like mapping

functions, *a-priori* ZHD models and cut-off angles and observe the changes.

SWD Retrieval from GPS Signal Delays

We know that the ZWD values calculated at any given time stamp represent an average of the multiple slant delay values as explained by the schematic diagram of Fig. 2.6. In this section we discuss the procedure to calculate the slant wet delay (SWD) values. Once we have processed the ZWD values using the GIPSY software as described in the earlier section, the SWD are derived by using eq. (2.16). This equation represents the wet delay part of the total slant delay from eq. (2.5).

$$SWD(e) = MF_w(e) \cdot (ZWD + \cot(e)(G_N \cdot \cos\phi + G_E \cdot \sin\phi)) + \epsilon \quad (2.16)$$

Here, after the ZWD values are processed from the software, *residuals.pl* function of the GIPSY software can be called for the post-processing. Using the *residuals.pl* function, we can get the information on the number of slant paths that are used in processing each of the ZWD values. For each ZWD value, the information on the slant path includes, the number of GPS satellites, the elevation angle (e) and the azimuth angle (ϕ) of each GPS satellite, the East and the North delay terms (G_N & G_E) and the value of residuals (ϵ) for each GPS slant path. The elevation angle dependency is used as the delay is expected to increase for lower elevation angles. Lower the elevation angle, longer the traveling time. Similarly, the delay is effected by the azimuthal asymmetries. The azimuthal asymmetries in the atmospheric refractive parameters can lead to errors in the estimated values [107]. G_N and G_E are the delay gradient parameters in the north and the east directions respectively. These parameters take into account the effect of the horizontal gradient of refractivity. This information along with the processed ZWD values can then be used in the eq. (2.16) to calculate the SWD values. As discussed in Chapter 2.4.2, the residuals are affected by both the atmospheric inhomogeneity and the multipaths. The residuals must be corrected to remove the multipaths. Residuals can be corrected using Multipath Stacking Algorithms (MPS). The MPS algorithm which is used in

this thesis are defined in Chapter 3. After the multipath errors are removed from the residuals, a positive residual value indicates that the slant path passes through a higher water vapor content and a negative residual value indicates that the slant path passes through a lower water vapor content compared to the zenith direction.

Following the equation in eq. (2.12), the slant water vapor (SWV) can be derived by multiplying the SWD by the conversion factor PI .

2.4.8 Error Sources for GPS-PWV Values

It is important to study the possible error sources for the ZWD values as it directly impacts the PWV values. For accurate retrieval of the PWV values, the ZWD values should be processed accurately.

As we know that the ZWD values are the difference between the two bigger delays ZTD and ZHD , the errors in the ZWD values are mostly from the errors in the estimated ZTD and ZHD values. The reported quality of ZTD estimates from near real-time processing is 3-10 mm [79]. There is a linear dependency between the daily mean of the total electron content (TEC) unit and the estimated vertical position [108]. An increase of the TEC unit from 25 to 175 will result in a ZTD error ranging from 0.6 to 4 mm if higher-order ionospheric corrections are not applied. Therefore, for ZWD processing the ionospheric correction models are necessary to correct the possible errors from the TEC of the ionospheric level.

Errors in the a-priori ZHD values caused by the use of inaccurate surface pressure values can result in an error of -0.1 to -0.2 mm/hPa in vertical position estimates, and this could also lead to an error in the calculated ZWD values [58]. Therefore it is important to use the accurate ZHD models for calculating the ZWD values. In Chapter 3, we show the use of different ZHD models and its effect on the final PWV values.

Furthermore, the tropospheric mapping functions(MF), which are used to map the tropospheric delay from any slant angle to zenith can impose an added error to the estimated PWV values. However, the effect of the MF reduces with an increase in the elevation cut-off angle [81]. In Chapter 3, we show the use of different mapping

functions and its effect on the final PWV values.

2.5 Applications of GPS Derived PWV Values

With the rapid deployment of GPS stations in local and regional levels, GPS derived PWV values are widely used in applications like the analysis of severe weather conditions [109] such as storms, thunderstorms [89, 110], flash-floods [47, 111–113], heavy rainfall events monitoring [48], rainfall forecasting [3, 4, 15], cloud microphysics and dynamics studies [114, 115]. In this thesis, we particularly focus on the applications of the GPS derived PWV values for rainfall monitoring and prediction.

2.5.1 GPS-PWV for Rainfall Monitoring

Many researchers have studied the PWV values and its usefulness in monitoring a rainfall event. Takiguchi et al. [116] reported that the onset of monsoon seasons in Thailand is followed by a rapid growth in GPS derived PWV values. Barindelli et. al [117] reported a peak in the precipitable water vapor value in response to a heavy precipitation event, followed by a steep decrease (5-10 mm in about 1 h) in the observed PWV values as the rain clouds moved past the station for a temperate station from Italy. Similarly in our study, we analyzed the trend in PWV values before and after a rainfall event. Time series plots of the PWV and the rainfall data are studied. A general trend of increment of PWV values before a rain event is noted. Cumulative distribution function (CDF) plots of PWV for rainy and non-rainy days in Singapore are presented in Chapter 5 which shows higher PWV values on the rainy days compared to the non-rainy days.

Shi et al. [48] presented some severe rainfall cases and a series of moderate rainfall cases to indicate the feasibility of GPS derived PWV for rainfall monitoring in China. The general trend of PWV w.r.t rain is highlighted in [48] and a fixed PWV threshold value of 50 mm is taken as a threshold to detect a rain event. It has been mentioned that the fixed PWV threshold value may not be suitable for different scenarios and seasonal factors can be considered to make it robust. In a recent study from Taiwan

[31], Yeh et. al reported different values of PWV thresholds under different rainfall strengths. The average PWV thresholds are reported to be 41.8 mm, 52.9 mm, 62.5 mm and 64.4 mm for no rain, light rain, moderate rain and heavy rain conditions, respectively. But the seasonal and diurnal variations in PWV values have not been considered while deriving the thresholds.

2.5.2 GPS-PWV for Rainfall Forecasting

Benevides et al. [3] presented a study of rainfall and GPS derived PWV values for a temperate station of Lisbon. This paper proposed a simple algorithm to forecast rainfall within 6 hours after the steep increase in PWV values. The algorithm used the maximum rate of change of PWV values as the only criteria for rainfall prediction. The algorithm was found to produce reasonable forecasts of precipitation with a success rate of 75 % but with a substantial amount of false alarms of 60-70 %. Following up, Yao et. al [4] introduced two auxiliary conditions to improve the success rate of rainfall prediction reported by [3]. In [4], the threshold of maximum rate of change of PWV values was taken as the main factor and the threshold of maximum variation in PWV values and the threshold of monthly averaged PWV values were taken as the auxiliary conditions to improve the detection rate. The algorithm was developed based on the data from the sub-tropical regions of China. After the application of this method, the results indicated a 7 % improvement in the correct detection rate compared to [3], but with a comparable false alarm rate of 66 %.

From literature, it was found that most of the works were done considering the data from temperate and sub-tropical regions. Very less work relating to the tropical climate has been reported. Therefore, in this thesis we analyze the importance of GPS-derived PWV for rainfall forecasting in tropical regions.

Literature also highlights that anomaly in PWV might be useful in the detection of rainfall. Li et al. [118] suggested that PWV anomaly values along with other meteorological parameters like temperature may be useful for predicting rainfall. Sharifi et al. [119] proposed to use the relative humidity anomaly along with the

PWV anomaly values to improve the prediction accuracy of rainfall.

2.5.3 Data-Driven Methods

In literature, there are few papers that use the machine learning algorithms to combine different features for rainfall prediction [120–122]. Mostly these papers are focused on the implementation and the inter-comparison of different machine learning algorithms. For the remote sensing community, however, it is more interesting to identify the important variables for rainfall prediction, their interactions under different conditions (rain and no-rain) and their level of contribution in rainfall prediction. Therefore, in this thesis, we study the interaction between different meteorological parameters like temperature, relative humidity, solar radiation, dew point, and their seasonal and diurnal behaviors along with the PWV values and their contribution in improving the rainfall prediction accuracy.

A machine learning based algorithm is applied to detect and predict a rainfall event. Artificial Neural Network (ANN) and Support Vector Machine (SVM) are the most commonly used supervised machine learning techniques for dealing with such geoscience-based problems [123]. We use SVM as it is computationally efficient and is suitable for our application. SVM is a parametric method that generates a hyperplane or a set of hyperplanes in the vector space by maximizing the margin between classifiers to the nearest neighbor data [124, 125].

For the problem of prediction of a rainfall event, the number of cases with rain is less compared to the no-rain instances. If the machine learning algorithm is trained with an imbalanced dataset, which is heavily skewed towards no-rain cases, then the test results will be biased. Therefore, there is a general practice of balancing the training data before using them in any machine learning algorithm. The two most common balancing techniques are the upsampling and the downsampling techniques. For upsampling technique, the observations from the minority class (rain cases) are repeated for certain number of times such that the ratio of minority class to the majority class (no-rain cases) becomes acceptable. For the downsampling technique, observations from the majority class are randomly chosen such that the ratio between

the two is acceptable. There is a general practice of using 1:1 majority to minority ratio [126]. Upsampling might cause the problem of longer training time especially for huge database and over-fitting, therefore, we use the downsampling technique with a 1 : 1 ratio in this thesis.

2.6 Summary

In this Chapter, an overview of the importance of PWV values and different existing technologies to retrieve the PWV values were discussed. Table. 2.2 summarizes and inter-compares the advantages and disadvantages of different technologies. The comparison between the existing technologies concluded that GPS is an emerging technology for the retrieval of PWV values with higher spatio-temporal resolutions. Unlike radiosondes, radiometers and satellite-based instrument, GPS can be operated in any weather conditions. Moreover, there is a dramatic change in the field of GNSS with the rapid development of GLONASS (the Russian GLObal NAVigation Satellite System), and newly emerging constellations, like the Chinese BeiDou Navigation Satellite System and the European Galileo system. With the rapid development of the GNSS constellations, there is a promising prospect for the real-time retrieval of wet delays and PWV values, which would increase the spatial coverage of the PWV values and would be of great benefit for meteorological applications like nowcasting and severe weather event monitoring [45, 46].

In this Chapter, a detailed review on the retrieval of PWV values from GPS signals was presented. Various empirical models were introduced and the Chapter also presented the applications of GPS-PWV values in rainfall monitoring and forecasting. In the subsequent chapters, different models and techniques are presented to overcome most of the shortcomings that were pointed out in the literature.

Chapter 3

Accuracy Assessment of GPS-Derived PWV

3.1 Introduction

In this chapter, we discuss the effect of using different ZHD models and different mapping function models on the calculated PWV values. We compare the calculated PWV values from the GPS signal delays with the PWV values calculated from the radiosonde profiles. Radiosonde PWV values are used as the ground truth to assess the accuracy of the GPS-derived PWV values. In this chapter, we also compare the GPS-derived PWV values with the MODIS-derived PWV values and discuss the results.

With the rapid development of the GPS technology, much research is over the GPS-derived PWV values in different fields of remote sensing. Thus, it is important to assess the accuracy of the GPS-derived PWV values before they are used in any application. We know that the PWV values are derived from GPS signal delays using the ZWD values and the conversion factor PI (eq. (2.12)). Therefore, the accuracy of the GPS-derived PWV values mainly depends on the ZWD values and the PI values. In this chapter we discuss the different parameter settings that can be used in the GIPSY software to calculate the ZWD values and their effect on the derived

PWV values.

3.1.1 Outline of the Contributions

The main contributions of this chapter are listed as below:

- Extensive evaluation of the effect of different *ZHD* models and the mapping function models on the GPS-derived PWV values for stations from the tropical, the sub-tropical and the temperate regions;
- Comprehensive comparison of the GPS-derived PWV values with the radiosonde-derived PWV values for different region and for different satellite cut-off angles;
- Comprehensive comparison of the GPS-derived PWV values with the MODIS-derived PWV values for different regions at different cloud mask values;

The rest of this chapter is organized as follows. Firstly, the GPS-PWV values are derived using different mapping function models and are compared to the PWV values obtained by using the default mapping function model. Secondly, the GPS-PWV values are derived using different *ZHD* models and are compared to the PWV values obtained by using the default *ZHD* model. Thirdly, the GPS-PWV values are compared to the ground truth, which is the PWV values derived from the co-located radiosonde station. The root mean square error (RMSE) of the difference between the GPS and the radiosonde derived PWV values under different conditions are discussed. Finally, the GPS-PWV values are compared to the MODIS-PWV values and the results are discussed.

Database:

The PWV values are processed for 26 IGS GPS stations (GPS-PWV). Fig. 3.1 shows the distribution of all these stations. For each station, 4 years (2012-2015) of PWV values are processed. Similarly, the radiosonde PWV values (Rad-PWV) are processed for 26 radiosonde stations which are co-located to the IGS GPS stations (ref. Fig. 3.1). The chosen GPS and the radiosonde stations are separated by less

than 50 km. The radiosonde data are downloaded from the Wyoming University website for upper air observatory [18] for the same period of time as that of the GPS data (2012-2015). A detailed information about the IGS GPS stations and the radiosonde stations can be found in appendix A in table A.1.

Figure 3.1: Locations of the co-located Radiosonde and IGS GPS stations. Markers in red (1-8) represent the tropical stations, in green (9-14) represent the sub-tropical stations and in yellow (15-26) represent the temperate stations, which includes 4 polar stations (station No. 20-23). Details of these stations are mentioned in table A.1.

3.2 Effects of Using Different Mapping Functions

The GPS-PWV values calculated using different mapping functions are compared to the GPS-PWV values calculated using the default mapping function of GIPSY software, which is the Neill Mapping Function (NMF). For this set of experiment, GIPSY is set to use its default *ZHD* model (Static *ZHD* model).

3.2.1 Time Series Plots

Figure 3.2: Time series plot of PWV values (Year 2012) calculated by using different mapping functions for three randomly chosen IGS GPS stations; Station No. 1, 10 and 16 in (a), (b) and (c) respectively. The small figures corresponding to each plot, show the zoomed-in view of the portion highlighted by the dotted box. The details about the stations are mentioned in the figure title. The x-axis indicates the day number and y-axis indicates the PWV values. Curves in red, blue, green and black represent PWV values when mapping functions NMF, GMF, GPT2 and VMF are used respectively [*Best viewed in color*].

Fig. 3.2 shows the time series plots of PWV values calculated using NMF, GMF, GPT2 and VMF1 mapping functions for three different stations; a station from the tropical region in (a), a station from the sub-tropical region in (b) and a station from the temperate region in (c). The station information are mentioned in the individual figure title and the details about different curves are given in the figure caption. The time series is plotted using GPS-PWV values from year 2012.

A full-year time series plot of the GPS-PWV values show that the curves corre-

sponding to the different mapping functions are all overlapped. Thus, it shows that there is no distinct difference between the GPS-PWV values calculated using either of the mapping functions for all the regions. Fig. 3.2 also shows zoomed-in pictures on the right for a closer look. The zoomed-in figures allow us to clearly see that there is negligible difference between the GPS-PWV values calculated using different mapping functions. To quantitatively assess the effects of using different mapping functions, the differences between the GPS-PWV values calculated using different mapping functions (GMF, GPT2 and VMF1) and the PWV values calculated using the the default mapping function, NMF are calculated and the results are discussed in the next section.

3.2.2 Residual Plot of Differences

Figure 3.3: Histogram of the difference between PWV values (mm) when using NMF mapping function and other functions as indicated by the legend for (a) tropical region (b) sub-tropical region and (c) temperate region. The black line marks the position of 0 difference.

Fig. 3.3 shows the histogram of difference in PWV values when using the default mapping function and the other mapping functions for the tropical, the sub-tropical and the temperate regions in (a), (b) and (c) respectively. The results from the polar stations are included in the results of the temperate region. Different markers in the figure represent the differences when using NMF and the other mapping functions, as indicated by the figure legend. To plot this figure, one full year (2012) GPS-PWV values from all the stations (totally 26) as shown by Fig. 3.1 are used.

When the residual plots are inter-compared, an interesting region-based trend can be noted. For the tropical region, the differences are left skewed, for the sub-tropical region, the differences have a peak at 0 and for the temperate region, the differences are right skewed. This explains the dependency of the different mapping functions on the latitude of the station. Although there are differences when using the default and the other mapping functions, the differences are very small, which was also observed from the typical time series plot in Fig. 3.2. The maximum difference is about 0.4 mm and the Root Mean Square Error (RMSE) of difference between PWV values when using NMF and others (GMF, GPT2 and VMF1) is under 0.25 mm. These smaller differences explain that there is not much effect on PWV values when using NMF or other mapping functions to calculate the PWV values, irrespective of different regions.

3.3 Effects of Using Different ZHD Models

Apart from the different mapping functions, we also study the effect of the different *ZHD* models. As the *ZWD* values are basically the difference between the *ZTD* and the *ZHD* values, the error in the calculation of the *ZHD* values has higher impact on the *ZWD* values and finally on the PWV values. Therefore, in this section we compute the PWV values by using different *ZHD* models and assess their differences. For this set of experiment, GIPSY is set to use its default mapping function (NMF mapping function).

3.3.1 Time Series Plots

Fig. 3.4 shows the time series plots of PWV values calculated using different *ZHD* models; Static, GPT, GPT2 and VMF1 for three different stations; a station from the tropical region in (a), a station from the sub-tropical region in (b) and a station from the temperate region in (c). The station information are mentioned in the individual figure title and the details about different curves are given in the figure caption. The time series is plotted using GPS-PWV values from year 2012.

Figure 3.4: Time series plot of PWV values (Year 2012) calculated by using different ZHD models for three randomly chosen IGS GPS stations; Station No. 1, 10 and 16 in (a), (b) and (c) respectively. The small figures corresponding to each plot, show the zoomed-in view of the portion highlighted by the dotted box. The details about the stations are mentioned in the figure title. The x-axis indicates the day number and y-axis indicates the PWV values. Curves in red, blue, green and black represent PWV values when ZHD models Static, GPT, GPT2 and VMF1 are used respectively [Best viewed in color].

Fig. 3.4(a) shows the time series plot for a tropical station. Here, the time series shows that there is not much difference between the different curves corresponding to the different ZHD models. The zoomed-in graph also does not show the differences among the different curves. Fig. 3.4(b) shows the full-year time series plot for a sub-tropical station [Note: there are missing values for VMF1 ZHD models after Day Number : 255]. Unlike in the tropical station, here the time series shows that the different curves have a similar trend but have some differences in their absolute numbers. The zoomed-in graph clearly shows that the GPS-PWV values calculated by using the Static ZHD model are higher than the PWV values calculated using

the other models. The PWV values calculated using the GPT, the GPT2 and the VMF1 models have not much differences between each other. Similarly, Fig. 3.4(c) shows the time series plot for a temperate station. The plot for the temperate station shows the differences between the PWV values calculated using the different *ZHD* models. Similar to the sub-tropical station, the zoomed-in graph of the temperate region clearly shows that the GPS-PWV values calculated using the Static model are higher than the PWV values calculated using the other models. Interestingly, unlike the sub-tropical station, here it can be noticed that the PWV values calculated using the GPT or the GPT2 models have similar values but are higher than the PWV values calculated using the VMF1 *ZHD* model.

Residual Plot of Differences

Figure 3.5: Histogram of the difference between PWV values (mm) when using Static *ZHD* model and other models as indicated by the legend for (a) tropical region (b) sub-tropical region and (c) temperate region. The black line marks the position of 0 difference.

In this section we analyze the difference between the GPS-PWV values calculated using the default *ZHD* model of GIPSY software; Static model and the GPS-PWV values calculated using other models. Fig. 3.5 shows the histogram of the differences for the tropical, the sub-tropical and the temperate regions in (a), (b) and (c) respectively. The results from the polar stations are included in the results of the temperate region. Different markers in the figure represent the differences when using the Static model and the other models, as indicated by the figure legend. To

plot this figure, GPS-PWV values of year 2012 from all the stations (totally 26) as shown by Fig. 3.1 were used.

When the individual residual plots are analyzed, it can be seen that all the differences i.e. difference in PWV values when using Static *ZHD* and GPT *ZHD*; difference in PWV values when using Static *ZHD* and GPT2 *ZHD* and difference in PWV values when using Static *ZHD* and VMF1 *ZHD* tend to be skewed towards the right. These observations indicate that the PWV values are higher when the Static *ZHD* values are used, compared to the cases when GPT, GPT2 and VMF1 *ZHD* values are used. The comparative study of these graphs show that for the tropical region, the differences are less as the peaks center around near to the zero. For the sub-tropical region, all the three differences have similar trend and are higher than the tropical region. For the temperate region, the differences for the GPT and the GPT2 model are similar and are lesser compared to the VMF1 model. These results correspond to our earlier findings from the typical time series plots, whereby we found that there is not much difference in PWV values when either of the models are used for the tropical region. The PWV values when the GPT, the GPT2 or the VMF1 models are used have similar trend for the sub-tropical region. Whereas for the temperate region, the PWV values when GPT and the GPT2 models are used are higher compared to the PWV values when the VMF1 model is used.

Here it can be observed that the differences in PWV values when the Static model is used and when the other models are implemented, can go higher than 5 mm, mainly in the temperate region. The GPT, the GPT2 and the VMF1 models include the latitude and day-of-year factor whereas the static model considers only the height of the station. For the temperate and the sub-tropical regions, there is a higher fluctuation in the seasons compared to the tropical regions. Therefore the the differences are higher in the temperate region and the sub-tropical region as compared to the tropical region.

In summary, from the observations for the different mapping functions, it is clear that the use of any mapping function, either the default or the others, does not impact on the GPS-PWV values. Whereas from the observations for the different *ZHD* models, it is seen that the use of the different *ZHD* models can result in

a higher difference compared to the default *ZHD* model. Therefore, in our next section, we compare the GPS-PWV values calculated using different *ZHD* models to the ground truth values, which are the PWV values calculated from the co-located radiosonde stations and propose a suitable *ZHD* model for different regions.

3.4 GPS-PWV Vs Radiosonde-PWV

PWV values are calculated for 26 radiosonde stations co-located with all the IGS GPS stations shown by Fig. 3.1. For both radiosonde and IGS GPS station pairs, the PWV values are calculated for years 2012-2015. Since the radiosonde data is available only twice a day, the GPS-PWV values are sampled according to the radiosonde launch timings. We calculate the GPS-PWV values by varying the *ZHD* models and compare it against the PWV values from the co-located radiosonde station. For this set of experiment, GIPSY is set to use the default mapping function (NMF mapping function).

3.4.1 Root Mean Square Errors (RMSE)

For the comparison purpose, RMSE of the difference between the GPS-PWV and Rad-PWV values are calculated. Fig. 3.6 shows the RMSE of difference between the Rad-PWV values and the GPS-PWV values when different *ZHD* models are used. The RMSE values shown in Fig. 3.6 are the average values calculated over 4 years for all the individual stations (ref. Fig. 3.1). The description about different markers are given in the figure caption. The figure is divided into sections based on the absolute latitude values.

For the tropical region, it can be observed that for most of the stations, the RMSE values when different *ZHD* models are used are almost the same. This is expected as our time series results in Fig. 3.4 and residual plots in Fig. 3.5, showed that the PWV values are almost the same when different *ZHD* models are implemented. It is interesting to observe that for few tropical stations, the RMSE values are better when the Static *ZHD* model is used compared to the other models.

Figure 3.6: Average Root Mean Square Error (mm) of difference between GPS-PWV and Rad-PWV for all the stations (ref. Fig. 3.1) for years 2012-2015. Values in star, circle, triangle and square markers represent the RMS Errors when Static, GPT, GPT2 and VMF1 *ZHD* models are used respectively. The absolute latitude values are plotted. The broken lines separate the tropical, sub-tropical, temperate and the polar regions. [Best Viewed in Color]

For the sub-tropical, the temperate and the polar regions, a gradual shift in the RMSE values can be observed. It can be seen that the RMSE values when the Static *ZHD* model is used are higher compared to other models for the sub-tropical, the temperate and the polar regions. The performance of the GPT or the GPT2 models are almost the same for all the sub-tropical stations. In the sub-tropical region, for few stations, the RMSE values when the GPT or the GPT2 models are implemented are similar to the RMSE values when the VMF1 model is implemented and for few other stations, the implementation of the GPT or the GPT2 models show better performance compared to the VMF1 model. For the temperate stations and a few stations from the polar region, it can be observed that the performance of the VMF1 *ZHD* model is always best compared to the other models.

In summary, this comparative study shows that the performance of the Static model is comparable to the performance of the other models in the tropical region and is even the best for most tropical stations. The performance of the Static model slowly degrades as we move from the tropical to the other regions. This is mainly

because of the seasonal fluctuations in the sub-tropical and the temperate regions which cannot be captured by the Static model. But the Static model works well for the stations in the tropical region where seasonal fluctuations are small. In the sub-tropical region, the performance of the GPT or the GPT2 model are better most of the time and in the temperate region and beyond, the performance of the VMF1 model outperforms others.

Therefore, we propose to use the Static *ZHD* model for the tropical region, the GPT or the GPT2 *ZHD* model for the sub-tropical region, and the VMF1 *ZHD* model for the temperate and the polar regions. Implementation of the Static, the GPT and/or the GPT2 *ZHD* models is easy and faster compared to the VMF1 *ZHD* model, as the VMF1 model is based on real observation data and the dataset needs to be updated frequently.

3.4.2 Density Plots

We calculate the GPS-PWV values for different regions using the suggested *ZHD* models with NMF mapping function and plot it against the Rad-PWV values as shown in Fig. 3.7 (a), (b) and (c) for the tropical, the sub-tropical and the temperate regions respectively. The results for the temperate region include the results from the polar stations. Four years (2012-2015) of data from all the stations (ref. Fig. 3.1) is shown in the plot. Here the scatter plot is represented in the form of a density plot to clearly show the data concentration for different regions. The color bars in Fig. 3.7 represent the number of data points in each bin (1 mm).

The scatter plots between the GPS-PWV and the Rad-PWV data for different regions show a very good correlation between the GPS-PWV and the Rad-PWV values. The overall correlation coefficients between the GPS-PWV and the Rad-PWV values are 0.91, 0.94 and 0.90 and the overall RMSE of the difference between the GPS-PWV and the Rad-PWV values are 6.3 mm, 5.1 mm and 3.4 mm for the tropical, the sub-tropical and the temperate regions respectively. The average RMSE for all the regions together is 4.9 mm. Comparable results can be found in the literature. In [22], 6 years of data from the temperate stations of Turkey has

Figure 3.7: Density plot of GPS-PWV Vs Rad-PWV for (a) tropical region (b) sub-tropical region and (c) temperate regions for data from years 2012-2015. Color bars represent the total number of data points in each bin (1 mm). *[Best Viewed in color]*

been used and results show a RMSE of difference between the GPS-PWV and the Rad-PWV values to be up to 3 mm. Similarly, in [24], 6 months of data from 10 temperate stations of Japan has been used and results show a RMSE of difference between the GPS-PWV and the Rad-PWV values to be 3.7 mm. In [25], 2 months of data from the sub-tropical stations of Korea has been used and the results show a RMSE of difference between the GPS-PWV and the Rad-PWV values to be 5.7 mm. Not much results have been reported for the tropical stations.

From the density plots, it can be clearly observed that for the tropical region the data points are more scattered compared to the sub-tropical and the temperate regions. This is also reflected by the average RMSE values; average RMSE is the highest for the tropical region and the least for the temperate region. This is because the climate is more localized in the tropical region compared to the other regions. The tropical region experiences the occurrence of cumulus clouds and convective rain events which are frequent in number and cover very small areas [127]. For example, although Singapore is a small tropical island, statistics shows that the West coast of Singapore is wetter than the east coast of Singapore [5]. For Singapore, the IGS-GPS station is in the west and the radiosonde station is in the East with a distance of around 24 km. Therefore, there might be cases whereby the PWV data at one of the stations is affected by a weather event such as rain and the other is not. This

affects the overall RMSE of difference between the GPS-PWV and the Rad-PWV values. In the other hand, the rainfall in the temperate region is mostly stratiform rain which covers a larger area and hence the probability of two co-located stations (with a distance of 50 km or less), to experience the similar weather conditions at the same time is higher. The difference is also affected by the location of stations towards the sea or to towards the mainland. The PWV values nearer to the coastal area are generally higher compared to the inland areas.

Here, the results presented in this section verify that the GPS-PWV values derived for this thesis have a good accuracy when validated against the ground truth, radiosonde-PWV values.

3.5 GPS-PWV and Radiosonde-PWV Values at Varying Cut-Off Angles

As discussed in Chapter 2.4.7, we know that the GPS-PWV values are the volumetric average of multiple slant path PWV values. The number of slant paths depends on the number of satellite observations present inside the GPS cone. The GPS cone is defined by the chosen elevation cut-off angle. As the cut-off angle increases, the cone becomes narrower and the number of GPS satellite observations inside the cone becomes smaller as shown by Fig. 3.8 (a). The chosen elevation angle is defined as the cut-off angle. There is a common practice of taking 7° to 10° of elevation cut-off angle for PWV calculations [48].

Fig. 3.8 (b) shows the number of GPS observations available when a particular elevation cut-off angle is chosen. The x-axis of the figure is the elevation cut-off angle and y-axis is the number of observations available for every 5 min reading within the particular cut-off angle. The boxplot clearly shows the range of the number of observations that is available for a particular cut-off angle. It can be seen that the number of observation decreases with the increase in the cutoff angle. That means the GPS-PWV at lower cut-off angle represents the average of many number of slant paths, at maximum reaching up to 12 paths. At higher cut-off angle the GPS-PWV may represent only a single slant path PWV value. At cut-off angle higher than 70° , there is almost no observations. Thus in the following analysis, a cut-off range of 5° to 50° is presented as the PWV values calculated within this range represent the average of a substantial (at least 2 paths per reading) number of slant path PWV values.

3.5.1 GPS-PWV Values at Different Cut-Off Angles

To calculate the PWV values at a particular cut-off angle we follow the following steps. Firstly, the *ZWD* values are calculated using GIPSY software for a minimum cut-off angle (5°). Secondly, *residuals.pl* function GIPSY software is called for the post-processing. Using the *residuals.pl* function, we can get the information on the

Figure 3.8: (a)Schematic diagram of elevation cut-off angles for GPS satellite observations. (b)Number of GPS satellites observed at different Elevation Cut-Off Angles.

number of slant paths that are used in processing each of the ZWD values. For each ZWD value, the information on the slant path includes, the number of GPS satellites, the elevation angle (e) and the azimuth angle (ϕ) of each GPS satellite, the East and the North delay terms (G_N & G_E) and the value of residuals (ϵ) for each GPS slant path. Thirdly, the residual values are corrected using Multipath Stacking (MPS) algorithms. Fourthly, the corrected residuals with other details are used in eq. (2.16) to calculate the SWD values. These SWD are then used in eq. (2.12) to calculate the slant water vapor (SWV). Finally, the SWV values which are from the cut-off elevation angles above the chosen cut-off elevation angles are projected to the zenith direction using NMF mapping function. The projected SWV values are then averaged to give the PWV value of certain cut-off angle. In the following the MPS

Figure 3.9: Residual values plotted against (a) Elevation and (b) Azimuth angles for the data from 31-Dec-2012 for IGS GPS station, NTUS. The blue dots for both the figures represent the uncorrected residuals and the red dots represent the corrected residuals. [Best viewed in color]

algorithms are described and the results for the slant water vapor are reported.

Multipath Stacking Algorithm

The residuals which are caused by water vapor, changes at different time stamps but the residuals from multipaths do not vary significantly with time. Multipaths are generally caused by different man-made or natural structures, which generally repeat on a daily basis according to the direction of the incident ray [128] and can be expressed as a function of the azimuth and the elevation angle of the GPS slant path. Therefore, the residual values are stacked according to the azimuth and the elevation angles for certain period of time and a mutipath map is created by averaging the residuals of each cell.

For each observed residual value, the mutipath map is called and the average residual value corresponding to the observed elevation and the azimuth angle is subtracted from the observed residual value to correct the multipath error. This is known as the mutipath stacking algorithm. Researchers use the multipath stacking algorithm with some differences in terms of the number of days considered to calculate

the average residual and the step size for the elevation and the azimuth angle to create the multipath maps [59, 128, 129]. For this thesis, we use a multipath stacking map created with a resolution of 1° for both elevation and azimuth angles. The residual values are averaged for every 20 days and the multipath stacking map is updated.

Fig. 3.9 shows residual values plotted against the elevation angles and azimuth angles of the GPS satellites. Both the figures show the corrected and uncorrected residuals as described by the figure caption. After correction, the range of residual values becomes narrower. It is interesting to observe the pattern in residual values w.r.t the elevation angles. From Fig. 3.9 (a), it can be observed that the residual values are higher at the lower elevation angles. This indicates that the slant wet delay at the lower elevation angles are higher. This is expected because the satellite signal travels through a relatively longer distance at the lower elevation angle therefore the delay value increases.

Zenith and Slant Wet Delays

Figure 3.10: Zenith wet delay and projected slant wet delay on 12-Dec-2018 for the NTUS IGS GPS station. Red dots represent the average zenith wet delay values and the blue dots represent the slant delays projected on the zenith direction. [Best viewed in color]

Using the corrected residuals the slant wet delay values can be calculated. To

compare the slant and the zenith wet delay values, in Fig. 3.10 the zenith wet delays are plotted together with the slant wet delay projected in the zenith direction. The zenith wet delay values are represented by the red dots and the slant wet delays projected in the zenith direction are represented by the blue dots. The result is shown from 31-Dec-2012 for IGS GPS station, NTUS. From the figure, it can be clearly observed that the zenith wet delays are an average of a number of slant wet delay values projected in the zenith direction. Therefore, the zenith wet delay values are an average over a cone and the size of the cone depends on the chosen elevation cut-off angle.

Slant Water Vapor

Once the SWD values are calculated the slant water vapor (SWV) values can be calculated by multiplying the SWD values by a conversion factor PI . The ZWD values in the eq. (2.12) can simply be replaced by SWD to calculate the slant water vapor. Fig. 3.11 shows the slant water vapor values for two different days for IGS GPS station, NTUS.

Figure 3.11: Slant water vapor (mm) for IGS GPS station NTUS. Different colors represent the different range of slant water vapor values as shown by the figure legend. [Best viewed in color]

In Fig. 3.11 sky-plots of the GPS satellites are shown. For both the sky-plots, the

circles represent the zenith angle (Zenith angle = 90° - elevation angle). Therefore, the center of the plot represents the zenith angle of 0° (and elevation angle of 90°), which means at this point, the satellite is in the zenith direction w.r.t the station. For both the plots, the radii represent the different azimuth angles from 0° - 360° . Each point in the sky-plot represent the SWV at the specific elevation and the azimuth angle of the satellite. The plots show the SWV on the satellite trails for a specific number of satellites. For the purpose of clarity in the figure, all the satellite trails are not displayed. Different colors on the satellite trails represent different ranges of SWV values. These sky plots give an overall idea about the distribution of water vapor. This can be useful in identifying the areas with higher water vapor which can probably help in forecasting cloud and rain occurrences. From Fig. 3.11, it is most likely that it was a cloudy/rainy day on the 4th January 2014 while a dry day with some thin clouds on the 20th February 2014. This is clearly indicated by the high SWV on the 4th of January compared to the 20th February.

To calculate the PWV values for a particular elevation cut-off angle, the slant water vapor values are projected in zenith direction and all the SWV values corresponding to the elevation angles greater than or equal to the given cut-off angle are averaged.

3.5.2 Methodology

For the purpose of comparing the GPS-PWV with the radiosonde-PWV values at various cut-off angles, we use the data from the GPS stations and the radiosonde station from Singapore. In addition to the IGS GPS stations, we use data from GPS stations under Singapore Satellite Positioning Reference Network (SiReNT). Details about different SiReNT GPS stations are mentioned in Appendix A in Table A.2.

Based on the common data availability, data from SiReNT GPS stations SRPT, SNPT, SLYG and SNYU along with the data from the radiosonde station of Singapore (Rad Id : 48698) for the year 2015 are used for this study. The location of these GPS stations and the radiosonde station are shown in the Fig. 3.12. Different markers are explained in the figure caption.

The GPS-PWV values are calculated for a range of the elevation angles for all the GPS stations as marked in Fig. 3.12. Similarly, the radiosonde-PWV values are calculated for the single radiosonde station. The radiosonde data for Singapore is available twice a day; 00 UTC, referred to as AM, and 10 UTC, referred to as PM. Therefore, the GPS-PWV values are extracted at the radiosonde timings for the comparison purpose. Correlation coefficient and RMSE are the two evaluation metrics that have been used to compare the GPS-PWV and radiosonde-PWV values at different elevation cut-off angles.

Figure 3.12: Map of Singapore showing 4 GPS Stations (red marker) and 1 Radiosonde Station (green marker). Marker 1: Radiosonde Station, GPS Station Names; Marker 2: SRPT, Marker 3: SNPT, Marker 4: SLYG and Marker 5: SNYU. The map is generated from an online map customizer site (<https://www.mapcustomizer.com/>). [Best viewed in color]

The GPS-PWV calculated at different elevation cut-off angles from 5° to 50° for the 4 GPS stations are compared against the radiosonde-PWV values. The result is discussed in terms of the correlation coefficient (%) and RMSE (mm). Fig. 3.13 shows the average correlation coefficient and average RMSE for both the AM and PM values over a period of one year (2015). The legend in the figure indicates the distance between a particular GPS station and the radiosonde station. The smallest distance is 6 km and the largest is 23 km.

3.5.3 Results and Discussion

Trend in Correlation Coefficient:

From the correlation coefficient plot in Fig. 3.13, it can be observed that for all the GPS stations, the average correlation coefficient between radiosonde-PWV and GPS-PWV decreases with the increasing elevation cut of angles. The effect of the cut off angles is more prominent for stations far from the radiosonde station. The correlation coefficients for GPS stations namely SNPT, SLYG and SRPT (ref. Fig. 3.12) are quite similar. This is because the radiosonde balloon has been observed to have a drift of up to 15 km away from the radiosonde station. Therefore stations within 15 km of the radiosonde station are expected to have a similar trend. The 3 stations SNPT, SLYG and SRPT are at a distance of 6, 10 and 16 km respectively, and so similar trends are observed. The only GPS station SNYU, which is more than 15 km away from the radiosonde station (23 km), gives the smallest correlation coefficient. In general, as distance increases, correlation coefficient decreases as shown in Fig 3.13.

Figure 3.13: (a) Correlation Coefficient in % and (b) RMSE in mm between GPS-PWV and radiosonde-PWV.

Here it is interesting to discuss the results from the literature [130], which com-

compares the GPS-PWV at different elevation cut-off angles to the radiosonde-PWV using data from Sweden. The results show that the correlation coefficient varies by a significant amount when the cut-off angle changes, which is higher than that is observed from Singapore data. One of the reasons is that the distance between the radiosonde station and the GPS station is mostly greater than 50 km in [130]. Another important reason is that, unlike the stations in Singapore, the stations in [130] have higher variation in altitude levels. Additionally, Singapore has the tropical climate, thus not much variation in water vapor content is observed at different stations for the same period of time.

Trend in RMS Error:

The RMSE between the GPS-PWV and radiosonde-PWV values for different stations at varying elevation cut-off angles are shown in Fig. 3.13 (b). The trend in RMSE is similar to the correlation coefficient and can be described in a similar manner. The RMSE for GPS stations nearer to the radiosonde station is smaller compared to the farther ones. It can be observed that the RMSE in the lowest elevation angle i.e. 5° is higher. This is because of the effect of multipath, which is high in the lower elevation cut-off angles as we discussed in Chapter 2.4.7 and is also reported in [130].

Effect of Seasonal Parameters:

In this section, the effect of different seasonal and weather parameters on the evaluation metrics is discussed. The correlation coefficients (in %) between radiosonde-PWV and GPS-PWV at a particular cut-off angle of 20° is listed in Table 3.1.

Time	SNPT 6 km	SLYG 10 km	SRPT 16 km	SNYU 23 km
AM	95.6	95.7	95.6	94.3
PM	78.3	77.8	76.4	76.1

Table 3.1: Correlation coefficients (%) at different times of day; AM and PM

It should be noted that a significant change is observed when the coefficients from

AM and PM are compared. The correlation coefficient for PM are generally smaller than the AM for a given elevation angle. This is because, Singapore generally has Monsoon season with many rainfall events occurring in the evening. These evening rainfall events are mostly the convective rain events which occur for small amount of time, covering smaller area [127, 131]. Hence, climate during the evening time is more localized in Singapore and is reflected upon the correlation coefficients.

3.6 Comparison of GPS-Derived PWV and MODIS-Derived PWV Values

In this section, we calculate the PWV values from the satellite instrument (MODIS) measurements and compare with the GPS-PWV values.

As described in Chapter 2.3.3, MODIS uses the near-infrared algorithm to retrieve the PWV data. The derived PWV values are stored as the MODIS Level-2 PWV products of collection-5 from Terra (MOD05_L2) and from Aqua (MYD05_L2). The MODIS level-2 collection-5 also includes the cloud mask products. The PWV and cloud products are generated at a 1 km by 1 km spatial resolution with a temporal resolution of 2 data points (TERRA and AQUA) per day. The PWV and the cloud mask products are downloaded from MODIS website [32] for 3 IGS stations namely *NTUS*, *NRMD* and *HYDE* from the tropical region, 3 IGS stations namely *AIRA*, *ALIC* and *TSKB* from the sub-tropical region and 3 IGS stations namely *ANKR*, *GLSV* and *HOB2* from the temperate region for the year 2015. The station coordinates of these stations are in Table A.1 of appendix A. The PWV values are calculated for these IGS GPS stations using the default mapping function (NMF), VMF1 *ZHD* model and 10° elevation cut-off angle.

We know from Chapter 2.3.3 that the MODIS retrieved cloud mask products have 4 levels : 192, 128, 64 and 0. 192 indicates that the MODIS pixel has a probability of 99% of being cloud free, 128 indicates a probability of 95 % of being cloud free, 64 indicates a probability of 66 % of being cloud free and 0 indicates 0 % of being cloud free (i.e. fully cloudy condition). In this section, we calculate the GPS-PWV

values and compare to the MODIS-PWV values at different cloud mask levels. The GPS-PWV values are sampled at the exact time when MODIS-PWV values are available.

Figure 3.14: Scatter plot of MODIS-PWV and GPS-PWV values from the tropical, the sub-tropical and the temperate regions. The description of the markers are given in the legend and along the text. Solid black line indicates the $x=y$ axis. [Best viewed in color]

Fig. 3.14 shows the scatter plot between MODIS-PWV and GPS-PWV for stations from tropical, sub-tropical and temperate regions. In Fig. 3.14, markers in red dots represent the cloud mask values of 0 (0% cloud free), blue stars represent the cloud mask values of 64 (66% cloud free), magenta circles represent cloud mask

values of 128 (95% cloud free) and green triangles represent cloud mask values of 192 (99 % cloud free). Here, from all the plots, it can be seen that the MODIS-PWV values correlate well with the GPS-PWV values when the MODIS cloud masks values are 192 with a probability of 99 % being cloud free. The average root mean square errors between GPS-PWV and MODIS-PWV data from 3 stations from each region when the cloud mask values are 192 (99 % being cloud free), are 7.80 mm, 5.47 mm and 4.19 mm for the tropical, the sub-tropical and the temperate regions respectively. And the average correlation coefficients at the cloud mask value of 192 (99 % being cloud free) are 0.85, 0.94 and 0.85 for the tropical, the sub-tropical and the temperate regions respectively as listed in Table. 3.2. The Table. 3.2 also shows the correlation coefficients for fully cloudy conditions (0 % being cloud free). From these observations, it is clear that the MODIS PWV values are accurate when the sky condition is almost clear [21].

Region	% of being cloud free	
	0 %	99 %
Tropical	0.16	0.85
Sub-Tropical	0.50	0.94
Temperate	0.47	0.85

Table 3.2: Average correlation coefficients at 0 & 99 % of cloud free conditions

Another important observation is that, the MODIS overestimates the PWV values in clear sky conditions [39]. It can be observed that the overestimation is more prominent for higher range of PWV values. MODIS overestimates the PWV values more in the tropical region which generally has higher water vapor content compared to sub-tropical and temperate regions. The presence of higher PWV values in the tropical region is clearly viewed from the density plot in Fig. 3.7.

One of the possible reasonings for such discrepancies between MODIS-PWV and GPS-PWV is that the fundamental working principles of the two measurement systems are different. GPS measurements are performed at every 5 minutes and are an average of at least 4 GPS satellite paths. In contrast, one scan of the MODIS mirror takes 1.4771 s to observe a swath and is the instantaneous observation over a 1 km

pixel.

In this section, we compared the MODIS-PWV and the GPS-PWV for 9 stations from different regions. It was found that the MODIS overestimates the PWV values especially for conditions with higher water vapor content. Possible reasons for such observations were analyzed. As a future work, the PWV values from MODIS can be calculated considering different pixel size (different area) to compare with GPS PWV values.

3.7 Summary

In this chapter, a detailed study over the effect of the different mapping functions; NMF, GMF, GPT2 and VMF1 on the final PWV values was carried out. It was found that there is not much difference in using the GIPSY software's default NMF and other mapping functions. Similarly, different *ZHD* models were used and their effect on the final PWV values were assessed. The PWV values calculated using different *ZHD* models were compared against the ground truth which is the radiosonde-PWV values. It was found that in the Static *ZHD* model, the GPT or the GPT2 model, and the VMF1 model have the least RMSE values in the tropical, the sub-tropical, and the temperate and the polar regions respectively. Therefore, it is proposed to use the Static model, the GPT or the GPT2 model, and the VMF1 model for the respective regions. The RMSE errors when the suggested models are used are 6.3 mm, 5.1 mm and 3.4 mm for the tropical, the sub-tropical and the temperate regions respectively.

The GPS-PWV values were also calculated for different elevation cut-off angles and were compared against radiosonde-PWV values for GPS and radiosonde stations from Singapore. It was found that the correlation coefficients/RMSE decrease/increase as the distance between the radiosonde station and the GPS stations increases. It was found that the effect was more prominent in the PM timings compared to the AM timings in Singapore as in the evening (PM), the climate in Singapore is localized with frequent local convective rain events.

The GPS-PWV values were also compared to the MODIS-PWV values. The

comparison was done under different cloud mask values. It was concluded that the GPS-PWV and the MODIS-PWV values correlate better during the cloud-free conditions with the RMSE values of 7.8 mm, 5.4 mm and 4.1 mm for the tropical, the sub-tropical and the temperate stations respectively.

Chapter 4

Conversion of Zenith Wet Delays to PWV

4.1 Introduction

In Chapter 2, it was reviewed that to derive PWV values from ZWD using eq. (2.12), a constant conversion factor PI is required. The PI value is calculated using eq. (2.13), in which the mean weighted temperature (T_m) is generally calculated using eq. (2.14), which uses the radiosonde data. But due to the low temporal resolution of radiosonde data, T_m is generally predicted using surface temperature (T_s) data. The relation between T_m and T_s is found to be site-specific and varies from one region to another. Thus, a new ($T_m - T_s$) relationship has to be derived for different observation sites and this is often time consuming. There are different empirical models that can be used to derive the T_m values, but the literature points out that some of these models also require T_s values and hence they cannot be used when ground-truth temperature values are not available [103]. Therefore, it is of great interest for us to propose an alternative and efficient way for retrieving the PWV values accurately from GPS signal delays.

4.1.1 Outline of The Contribution

The main contributions of this chapter are listed as below:

- Study of the different existing T_m - T_s relationships using data from the tropical, the sub-tropical and the temperate stations.
- In depth study of the PI values for different regions and proposal of a global PI model for the retrieval of PWV values from the ZWD values.

The rest of this chapter is organized as follows. Firstly, we present an overview of some existing (T_m - T_s) relationships and the relationships that we have studied using 4 years of data from 174 different stations consisting of tropical, sub-tropical and temperate regions. Secondly, we study the PI values with respect to the parameters like station latitude, station height and day-of-year and propose a model to calculate the PI values. Finally, we compare the results from the proposed model to other databases and validate the proposed model.

Database:

Radiosonde Data: Out of many radiosonde stations, Fig. 4.1 shows all those stations whose data have been used for this Chapter. Fig. 4.1 shows a total of 174 radiosonde stations out of which 59 are from the temperate region, 62 are from the sub-tropical region and 53 are from the tropical region. The labels are described in the caption of the Fig. 4.1. Four years of data (2012 - 2015) have been processed for all the 174 radiosonde stations. As we know, a radiosonde from a particular station is launched only twice a day; at 00 UTC (AM) and at 12 UTC (PM), in a year, there should be a total of 730 radiosonde profiles. But in general it is found that most of the stations have very poor data availability at PM as compared to the AM. For example, for radiosonde station at Singapore (with radiosonde station Id: 48698), there are 358 valid observations for AM and only 218 observations for the PM for year 2014. For many other stations, the availability of data for both AM and PM are missing for considerable number of days. Therefore, only those stations which have good data availability were selected.

Figure 4.1: Location of 174 radiosonde stations. Markers numbered from 1 – 59 (in red color) represent the temperate stations, numbered from 60 – 121 (in green color) represent the sub-tropical stations and those from 122 – 174 (in pink) represent the tropical stations. The map is generated from an online map customizer site (<https://www.mapcustomizer.com/>). [Best viewed in color]

METAR Data: METAR is a format for reporting the weather information. We use METAR reports to retrieve the surface temperature information T_s for different locations. These T_s values are later used in modeling station specific T_m - T_s relationships. METAR reports are also provided by the Wyoming University database for different locations [132]. The METAR reports show that the surface observation is made for different stations at different timings. The temporal resolution of T_s values reported in METAR is generally 30 minutes. Four years; 2012-2015 of T_s values from METAR reports for all the stations as shown by Fig. 4.1 are used in this thesis.

4.2 Mean Weighted Temperature (T_m) and Surface Temperature Relationships (T_s)

4.2.1 Review of Existing ($T_m - T_s$) Models

Ref	Coordinates	Region	a	b
[95]	45.95° N, 13.64° E	Temperate	39.94	0.83
[4, 96]	30.53° N, 117.12° E	Sub-Tropical	44.05	0.81
[97]	25.77° N, 113.01° E	Sub-Tropical	89.33	0.65
[88]	4.21° N, 101.98° E	Tropical	182	0.35
[133]	1.35° N, 103.68° E	Tropical	182.56	0.34

Table 4.1: ($T_m - T_s$) relationship for specific stations from literature

As discussed in the literature, the most commonly used ($T_m - T_s$) relationship was reported by Bevis et al. [50] in eq. (2.15). It was noted that eq. (2.15) was derived based on analysis of radiosonde profiles from temperate region only. Therefore, instead of using eq. (2.15), some researchers were aware that the coefficients a and b in eq. (2.15) are region and season specific, and thus proposed their own ($T_m - T_s$) relationships [88, 95, 96, 133] as summarized in Table 4.1.

As shown in Table 4.1, for the temperate city of Nova Gorica, Slovenia (45.95° N, 13.64° E) [95], the coefficients a and b are found to be 39.94 and 0.83; while for the sub-tropical city of Anqing, China (30.53° N, 117.12° E) [96], they are found to be 44.05 and 0.81 and for sub-tropical region from Hunan province of China, the coefficients are reported to be 89.33 and 0.35 [91]. Moreover, for the tropical regions of Malaysia (4.21°N, 101.98° E) [88] and Singapore (1.35° N, 103.68° E) [133], the values of a and b are found to be close to each other, i.e., 182 and 182.5, and 0.35 and 0.34, respectively. As can be seen from Table 4.1, when latitude increases, the coefficient a decreases and b increases. It is also clear that the coefficients are distinctly different from those of Bevis equation, eq. (2.15). Therefore, Bevis equation is not global and station specific $T_m - T_s$ relationship needs to be defined [134].

4.2.2 Derivation of New ($T_m - T_s$) Relationships

For our analysis, we have also derived the (T_m - T_s) relationships for different regions. T_m data from 174 radiosonde stations and T_s data from collocated METAR reports are used in the derivation of the station specific models. Fig. 4.1 shows the 174 stations from which the radiosonde data and the METAR reports have been used.

Fig.4.2 presents the scatter plot of T_m vs T_s and the linear regression between the two for the temperate, sub-tropical and tropical regions. The relationship for all the regions are in the form of eq. (2.15). The estimated coefficients a and b with the total number of stations used for derivation and the latitude ranges are summarized in Table 4.2.

From Table 4.2, it can be seen that the gradient term, b , for temperate region (0.73) is higher and decreases in the sub-tropical (0.60) and the tropical (0.52) regions. This trend is consistent to those derived by other literature as shown in Table 4.1. This indicates that the correlation between T_m and T_s is highest for temperate region and lowest for tropical region, and also verified by the estimated correlation coefficient values which are 0.90, 0.77 and 0.5 for temperate, sub-tropical and tropical regions respectively. This result can also be distinctly analyzed from Fig.4.2, which shows data from all region at once. Such variations amongst the 3 regions are mainly due to seasonal changes. For example, temperate regions have four distinct seasons, spring, summer, autumn and winter, and therefore, the surface temperature fluctuates over a large range from 230 K to 320 K. For the tropical region, temperatures remain high all the year round, and therefore, the surface temperature fluctuates over a much smaller range of 289 K to 310 K as compared to the temperate region.

From both Table 4.1 and Table 4.2, it can be clearly noted that the relation between T_m and T_s is site-specific which varies from one region to another and the coefficients of only temperate region matches well with that of the Bevis equation 2.15. Thus, it can be concluded that the Bevis equation is only suitable for stations in the temperate region and a new ($T_m - T_s$) relationship needs to be derived for each station/region, which becomes complex and tedious especially when the data sources for T_m and T_s are not collocated.

Figure 4.2: T_m vs T_s for stations from temperate region (in red dots), from sub-tropical region (in blue circles) and from tropical region (in green triangles) with linearly regressed equations represented in black, yellow and magenta for the respective regions. [Best viewed in color]

Therefore, it is significant to develop an alternative methodology with good accuracy that can directly predict the PI values for different stations, instead of using long term site specific parameters like temperature which might be difficult to come by.

Region	Number of Stations	latitude Range (Absolute Values)	a	b
Temperate	59	$36.41^\circ - 90^\circ$	67.12	0.73
Sub-Tropical	62	$23.80^\circ - 35.33^\circ$	106.36	0.60
Tropical	53	$0.05^\circ - 22.31^\circ$	129.13	0.52

Table 4.2: Derived ($T_m - T_s$) relationships for different regions

4.3 Modeling and Proposal of a New *PI* Function

Following the discussion in the previous section, an alternative method to predict the *PI* values is presented in this section.

Four years (2012 – 2015) of *PI* values calculated from all the 174 radiosonde stations (shown in Fig.4.1) using eq. (2.13) and 2.14 are studied against the independent parameters like station coordinates, day of year (*DoY*), and station altitude. A temperature independent model is then proposed. Predicted *PI* values with the new proposed model will be used in eq. (2.12) to derive the PWV values.

4.3.1 Trend in *PI* Values with Day of Year

Fig. 4.3 presents 4 years of daily averaged *PI* values plotted against the *DoY* for all 174 stations. The red and green circles are for the northern (31 stations) and the southern (22 stations) tropical stations, blue and cyan dots are for the northern (42 stations) and the southern (20 stations) sub-tropical stations and black and magenta crosses are for the northern (49 stations) and the southern (10 stations) temperate stations respectively. In general, radiosonde stations are sparsely distributed in the southern hemisphere. Most of the southern area on earth is covered by sea. Thus the number of southern temperate stations compared to the northern hemisphere is small. Almost all the available stations from southern temperate regions have been included in the study. It can be noted from Fig. 4.1 that the chosen stations are well distributed.

From Fig. 4.3, it can be seen that the *PI* values for tropical stations are almost constant throughout all years [135]. However, the *PI* values for other two regions show sinusoidal trends with the change of day number (*DoY*). The trend of change in *PI* values for stations from north and south alternates for the respective region due to the alternate seasons. The amplitude of the sinusoid is higher for stations from northern hemisphere compared to that for stations from the southern hemisphere. This phenomenon is especially observed for stations from southern temperate region. As previously mentioned, the number of stations considered from southern temperate region is limited and the chosen stations are all near to the sea (from coastal areas).

Figure 4.3: Plot of daily mean PI values against Day of Year (DoY), for 4 years of data from 2012 to 2015. The red and green circles are for the northern and the southern tropical stations, blue and cyan dots are for the northern and the southern sub-tropical stations and black and magenta crosses are for the northern and the southern temperate stations respectively. The solid red line represents the best fit line for respective regions. [Best viewed in color]

Thus the variation in PI values does not follow very distinct pattern as that of northern stations.

$$PI = c \cdot \cos\left(\frac{(DoY - 28) \cdot 2\pi}{365.25}\right) + d \quad (4.1)$$

The sinusoidal trend of PI values with DoY can be mathematically modeled with eq. (4.1). Eq 4.1 was modeled by fitting a sinusoidal curve as shown by the red lines in Fig. 4.3. Eq. (4.1) follows the same pattern as the Global Pressure Temperature (GPT) model to calculate surface temperature and the mean weighted temperature in [100]. Through the regression technique, the values of c and d in eq. (4.1) for different regions can be estimated and are tabulated in Table 4.3 and the best fit line is plotted in solid red in Fig. 4.3.

From Table 4.3, it is found that values of d decreases as we move from the tropical stations to the temperate stations for both the northern and the southern regions and

Region	c		d	
	north	south	north	south
Temperate	-0.0066	0.0028	0.152	0.145
Sub-Tropical	-0.0038	0.0019	0.160	0.161
Tropical	-0.00016	0.00010	0.164	0.163

Table 4.3: c and d values for different regions.

the absolute value of parameter c increases from tropical stations to the temperate stations. Physically, parameter d , represents the long term annual statistics of PI values which can be derived based on the station location. Parameter c indicates the seasonal variations, which is the highest in the temperate region, low in the sub-tropical region and the least in the tropical region. This trend can be seen from c values in Table 4.3. Note that the trend is more dominant in the northern hemisphere as compared to the southern hemisphere. This parameter c is therefore location dependent as well as the fact whether the station is from the northern or southern hemisphere.

4.3.2 Study of Latitude Dependency

Modeling Parameter d

Four years' of data is used to derive the long term annual median PI values to model the parameter d . Fig. 4.4 shows the boxplot of PI values for all the 174 stations. The red colored boxes are for stations from negative latitude and black for the stations from positive latitude region. From Fig. 4.4, it can be seen that when the absolute value of latitude increases, the median PI values decrease whilst its variation increases. For stations from tropical region (with absolute latitude between 0.05° and 22.31°), the median PI value is the highest at around 0.162 and its variation is also the smallest. A significant variation in PI values is observed as the stations move farther from the equator. Such differences can again be linked to the seasonal changes that occur in sub-tropical and temperate regions, which cause larger variation in temperature and hence effects are observed in the PI values.

Moreover, as can be inferred from Fig. 4.3, the median PI values of a given

Figure 4.4: Boxplot of PI values for 174 radiosonde stations for 4 years (2012-2015). Red color represents the stations from negative latitude and black from positive latitude. [Best viewed in color]

station could be predicted using its latitude information. The median PI values is found to have a good and clear correlation with the latitude of the station as shown in Fig. 4.5(a). The median PI values for stations from negative latitude are represented by red crosses and from positive latitude are represented by black dots. The solid blue line is the best fit curve as shown in eq. (4.2). The best fit line was achieved by the least-squares method such that the difference between the modeled PI values and actual PI values are minimum. Given L_a , the latitude value of any given station, eq. (4.2) calculates the median PI value for that station.

$$d = 0.165 - (1.7 \cdot 10^{-5})|L_a|^{1.65} \quad (4.2)$$

While modeling PI values, it was observed that the station altitude (H) might also affect the accuracy of the modeled PI values. Therefore, the effects of station altitude, H , on the predicted PI values is also assessed by using a residual parameter which is the difference between the predicted median PI value, d using eq. (4.2) and actual median PI value. Fig. 4.5(b) shows the residual plot. It clearly shows a linearly decreasing trend of residual with reference to the station altitudes. The solid blue line in Fig. 4.5(b) is the linearly regressed equation which gives the altitude correction factor defined as $f = -2.38 \cdot 10^{-6}H$. It is found that the change in PI

Figure 4.5: (a) Study of median PI values with respect to absolute latitude values. (b) Altitude correction for the estimated PI values in (a). (c) Variation in parameter c with respect to the absolute latitude values. [Best viewed in color]

values is independent of altitude when stations are from lower altitude region where $H < 1000$ m. The effect of H becomes significant when the stations are from higher altitude region where $H > 1000$ m. Therefore, it is advisable to use the altitude correction factor for stations from higher altitude region where $H > 1000$ m.

Modeling Parameter c

It is noted that the parameter c reflects the seasonal variations and specific values of c (as shown in Table 4.3) show latitude and hemisphere dependency of PI values. The plot in Fig. 4.5(c) shows the relation of parameter c with respect to the latitude. The red crosses are for stations from negative latitude and the blue dots are for stations from positive latitude. The solid red and blue lines represent the best fit curve for parameter c as shown by eq. (4.3).

$$c = -1 \cdot \text{sgn}(L_a) \cdot 1.7 \cdot 10^{-5} \cdot |L_a|^{h_{fac}} - 0.0001 \quad (4.3)$$

Fig. 4.5(c) and typical values of parameter c from Table 4.3, show that the parameter c also captures the phenomenon of higher PI values for southern stations compared to the northern stations with alternating signs.

In eq. (4.3), a new factor, named as h_{fac} , is introduced to account for the differences between northern and southern hemispheres. The regression results show that h_{fac} is 1.25 for stations from southern hemisphere and 1.48 for stations from northern hemisphere. A term $\text{sgn}(L_a)$ denotes whether the station is from the northern hemisphere or the southern hemisphere. $\text{sgn}(L_a)$ is 1 for stations from the northern hemisphere and -1 for stations from the southern hemisphere.

4.3.3 Proposal of A Simplified Model

Therefore, to calculate the PI values for a given station at a specific day, we need the station latitude information in degrees, L_a , altitude of the station in meters, H , day of the year for which the PI values are to be calculated, DoY . Depending on whether L_a values are positive (from north) or negative (from south), h_{fac} can be selected accordingly. h_{fac} for the northern hemisphere is 1.48 with $\text{sgn}(L_a) = 1$ and for the southern hemisphere is 1.25 with $\text{sgn}(L_a) = -1$. Using these information, the proposed PI model is given in eq. (4.4).

$$PI = [-1 \cdot \text{sgn}(L_a) \cdot 1.7 \cdot 10^{-5} |L_a|^{h_{fac}} - 0.0001] \cdot \cos\left(\frac{(DoY - 28)2\pi}{365.25}\right) + [0.165 - (1.7 \cdot 10^{-5})|L_a|^{1.65}] + f \quad (4.4)$$

Where, f is negligible for stations from low altitude ($H < 1000$ m) region and $f = -2.38 \cdot 10^{-6} \cdot H$ for stations from high altitude ($H > 1000$ m) region.

4.4 Performance Evaluation and Validation

4.4.1 Performance Comparison

For assessing the performance of our proposed temperature independent model, PI values calculated from our proposed method in eq. (4.4) are compared with two different temperature dependent methods: 1) PI values calculated using T_m derived from radiosonde data and 2) PI values calculated using T_m database published by GGOS Atmosphere. GGOS database uses data from numerical weather models from the European Centre for Medium-Range Weather Forecasts by operational analysis [98].

Four years of PI values calculated for all 174 locations (shown in Fig. 4.1) using both radiosonde and GGOS data are used for this purpose. In the following, PI values calculated using eq. (4.4) is named as $Prop-PI$, PI values from radiosonde data calculated using eq. (2.3) is named as $Rad-PI$, and PI values calculated from GGOS T_m database using eq. (2.13) and eq. (2.14) is named as $GGOS-PI$.

The relative error in PWV, from eq. (2.12), due to error in PI values can be calculated using eq. (4.5) [94, 136].

$$\frac{\Delta PWV}{PWV} = \frac{\Delta PI}{PI} \quad (4.5)$$

On the basis of eq. (4.5), for PI values ranging from 0.120.17 (Fig. 4.2), the 1% and 2% accuracies in PWV requires errors in PI values of less than 0.0014 and 0.003 on average, respectively. Fig. 4.6 (a) and (b) show the histograms of difference between $Prop-PI$ and $GGOS-PI$ and difference between $Prop-PI$ and $Rad-PI$ respectively. Almost 67 % of the differences between $Prop-PI$ and $GGOS-PI$ and 58 % of the differences between $Prop-PI$ and $Rad-PI$ are within $\pm 0.003 PI$ values. Therefore, the results suggest that the predicted PI values from the proposed

method can be used to predict PWV values with high accuracy as compared to the existing methodologies. This is also reflected in the next section.

4.4.2 Further Validation of the Proposed Model

The proposed model is further validated using different databases; the IGS GPS GGOS database, IGS GPS NASA database, Very-Long-Baseline Interferometry (VLBI) wet delay data, and the radiosonde data.

1. GPS GGOS Database

Two years (2013-2014) of ZWD values and corresponding T_m values for 384 IGS stations are downloaded from GGOS Atmosphere website [68]. The PWV values are then derived for each IGS site using T_m and GGOS ZWD (eq. (2.12-2.13)) and is named as $PWV - GGOS - T_m$. The IGS site location information (altitude and latitude) and DoY is used in eq. 4.4 for the proposed model and PWV values are calculated using predicted PI values and GGOS ZWD (eq. (2.12)), named as $PWV - GGOS - Prop$.

Fig. 4.7(a) shows the histogram plot of difference between $PWV - GGOS - Prop$ and $PWV - GGOS - T_m$. It is calculated that almost 98% of difference is within ± 1 mm, and the root mean square (RMS) of difference is around 0.31 mm. The RMS values suggest that the PWV values calculated using the proposed method works with similar accuracy as compared to the existing temperature dependent methodology [137].

2. GPS NASA Database

Two years (2013 – 2014) of RINEX files are downloaded from NASA’s website [44] for IGS stations shown in Table A.1 and ZWD is processed using GIPSY OASIS software. All the chosen stations are co-located to the corresponding radiosonde stations used in the derivation of the model. For each IGS station, corresponding ($T_m - T_s$) relationship is selected from Table 4.2 to predict the T_m values. The predicted T_m values are then used in eq. (2.12-2.13) to derive the PWV values which is named

Figure 4.6: (a) Histogram of the difference between $Prop - PI$ and $GGOS - PI$ (b) Histogram of difference between $Prop - PI$ and $Rad - PI$. With bin size of 0.006 PI values. (4 years of data from 174 stations (Fig. 4.1))

as $PWV - IGS - T_m$. Each station location information (altitude and latitude) and DoY is then used in the proposed model, eq. (4.4), to get the corresponding PI values. Finally the predicted PI values and the calculated ZWD values are used in eq. (2.12) and PWV values are calculated, which is named as $PWV - IGS - Prop$.

Fig. 4.7(b) shows the histogram plot of difference between $PWV - IGS - Prop$ and $PWV - IGS - T_m$. The distribution of histogram is similar to Fig. 4.7(a) with almost 99% of the differences within ± 1 mm with an RMS value of 0.33 mm. This result again strongly suggest that the proposed model can calculate PWV values as accurately as using the traditional $T_m - T_s$ relationship approach [137].

3. VLBI Database

The proposed model is further validated using the VLBI database. The CDDIS archive of VLBI products consists of solutions derived by IVS analysis centers from

Figure 4.7: (a). Histogram of difference between $PWV - GGOS - Prop$ and $PWV - GGOS - T_m$. (2 years of data from 384 IGS stations) (b) Histogram of difference between $PWV - IGS - Prop$ and $PWV - IGS - T_m$. (2 years of data from IGS GPS stations (ref. Table A.1)). (c) Histogram of difference between $PWV - VLBI - Prop$ and $PWV - VLBI - T_m$. (2 years of data from 15 VLBI stations). With bin size of 1 mm.

the analysis of VLBI experiment files [138]. These products include the troposphere parameters like ZTD and ZWD . These tropospheric parameters with meteorologically important surface parameters such as temperature, pressure and humidity, is downloaded from IVS data center [139]. In this analysis, 2 years (2013 – 2014) of ZWD values and corresponding surface temperature values for 15 VLBI stations (Algo, Bada, Fort, Hart, Hoba, Kath, koke, Mate, Nyal, Onsa, TSKB, West, Wett, Yebe, Zele) are used.

For each VLBI station, corresponding $(T_m - T_s)$ relationship is selected from Table 4.2 and T_m values are then predicted. The predicted T_m values and VLBI wet delay values are then used in eq. (2.12-2.13) to calculate the PWV values which is named as $PWV - VLBI - T_m$. Each station location information (altitude and latitude) and DoY is then used in the propose model, eq. (4.4), to get the corresponding PI values, which are used in eq. (2.12) with VLBI ZWD to calculate the PWV values, and named as $PWV - VLBI - Prop$.

Fig. 4.7(c) shows the histogram plot of difference between $PWV - VLBI - Prop$ and $PWV - VLBI - T_m$. The distribution shows that a large portion (almost

93%) of the differences is very small (within ± 1 mm) with an RMS value of 0.38 mm. Therefore, similar conclusions as presented in earlier subsection can be drawn. Through validations using different datasets, it is shown that the proposed model works well with high accuracy.

4.4.3 Independent Station Validation

To have a comprehensive comparison, a GPS station from Japan (IGS GPS station ID: TSKB), which is collocated with the VLBI station (VLBI station ID: TS) and radiosonde station (radiosonde station no.: 47646) is used for independent validation purpose. The radiosonde data from this station was deliberately left out (not one of the 174 stations) during the derivation of the proposed equation

Based on the good data availability from all databases, two years (2011 and 2012) of *ZWD* data are downloaded from the respective sites of GPS station (using NASA database and GGOS database) and VLBI station. Corresponding PWV values are then calculated using the proposed model. PWV values are also calculated for the collocated radiosonde station. PWV calculated for GPS station; using NASA database is named as *IGS – PWV*; using GGOS database is named as *GGOS – PWV*; using VLBI database is named as *VLBI – PWV* and using radiosonde database is named as *radiosondePWV*.

Fig. 4.8 shows the scatter plot of PWV values from different databases against *DoY*. The green circles represent the *IGS – PWV*, black dots represent the *GGOS – PWV*, red stars represent the *VLBI – PWV* and blue triangles represent the radiosonde PWV. From Fig. 4.8, it is observed that the PWV values calculated using GPS data and VLBI data which uses the proposed model follows the same pattern over different years and matches well with the radiosonde PWV. Although the pattern is the same, absolute values of PWV from GPS, VLBI and radiosonde have some differences since different instruments were used for the measurement of the delays and for calculation of the PWV. Furthermore, all these stations are not exactly at the same location. There is approximately a 7 km distance between the GPS and radiosonde stations.

Figure 4.8: (a). Plot of PWV Vs *DoY* for different years. Green circle represent PWV from GPS station using IGS database, black dots represent PWV from GPS station calculated from GGOS database, red stars represent PWV from VLBI data and blue triangles represent PWV from radiosonde data. 2 years of data from independent station of Japan, TSKB (36.06°N , 140.05°E). [Best viewed in color]

Moreover, it is also observed that the PWV from VLBI and the PWV from GPS have the same pattern but the PWV from VLBI is generally less than the PWV from GPS. Similar observations have been reported in the literature [138]. Small differences can be observed between PWV calculated for the same GPS station but using different sources (from NASA and from GGOS). This is due to the error induced by different parameters considered during the processing of the wet delay by using different processing softwares with different settings.

Despite the small differences which can be accounted for in the above, all 4 PWV matches very well. Therefore, by using an independent station for performance evaluation, using three independent databases, GPS, VLBI and radiosonde, it can be concluded that the proposed model is found to be both accurate and reliable. It has an added advantage of being simple to implement and does not require any other databases for its implementation.

4.5 Summary

In this chapter, a simplified, latitude and day-of-year based PI value model is proposed for the retrieval of PWV values from GPS signals. The PI value is conventionally determined using the water-vapor weighted mean temperature T_m which is found to have a linear relationship with the surface temperature T_s . This method is site-specific and its' use is limited by the poor temporal and spatial resolution of radiosonde data.

In the analysis of different regions: temperate, sub-tropical and tropical using data of four years from 174 different stations, it was concluded that the median PI value varies with latitude and variations about the median PI value are the result of seasonal changes that a particular station goes through. Stations from the tropics have very little variation whereas stations from the temperate has the highest variation. PI value residuals were studied with respect to station altitude and was concluded that the altitude of a station plays a significant role when the stations are from higher altitude compared to lower altitude. For simplified model it is recommended that the altitude factor should only be used when the altitude of a given station is higher than 1000 m. Thus, by examining the relationship of PI with respect to its geographic coordinates and considering the seasonal effects, a latitude and DoY based PI model was proposed.

The accuracy of the proposed PI model for estimating the GPS-based PWV is compared against the PI values calculated using T_m data from radiosonde and from GGOS database using 4 years of data from 174 stations. Results show that the PI values estimated using the proposed model have similar accuracy to the temperature dependent model. PWV values were calculated using temperature dependent models and compared against the proposed model using 3 databases; GPS, VLBI and radiosonde. It is shown that the error in PWV is small and lies within $\pm 1\text{mm}$. This accuracy is high enough for different geodetic applications [137].

The proposed model has the advantage of being computationally efficient and simple to use without requiring high temporal and spatial resolution resources, and thus can be applied universally.

Chapter 5

GPS-Derived PWV Values for Long-Term Rainfall Prediction

5.1 Introduction

One of the most important applications of PWV values are in analyzing the cloud and rain events. Rainfall depends on a myriad of factors and moisture content of atmosphere is one the important factors. For initiation of a rainfall event, the moisture content of the atmosphere should be high enough that it exceeds the saturation threshold value. The saturation depends on temperature values. For lower temperature, saturation values are smaller and for higher temperature the saturation values are higher. This is because hotter air can hold more water vapour and thus saturation point is higher. Therefore, more moisture content is needed in a warm environment for a rainfall event to initiate. This hints us that the PWV values can directly link to rainfall events.

With the rapid deployment of GPS CORS stations, GPS-derived PWV values are extensively used in many applications including rainfall prediction. With an acceptable accuracy of GPS-derived PWV values and an added advantage of having higher spatio-temporal resolution and being operational under all weather conditions, there is a growing interest among remote sensing analysts to study the usefulness of

GPS-derived PWV values in rainfall prediction.

In Chapter 2 a number of recent papers that discuss about the possible usefulness of GPS-derived PWV values in rainfall prediction were reviewed [1, 3, 4, 15]. In literature there are papers that propose algorithms for rainfall prediction from few minutes to upto 6 hours. In [3] and [4], data from the temperate and the sub-tropical stations are used and 6-hourly rainfall prediction algorithm is proposed. In [1], data from the sub-tropical station are used to predict typhoons in 10-30 minutes in advance. In this thesis we consider the 6-hourly prediction as the long-term prediction and we consider the prediction window of upto few minutes as rainfall nowcasting or short-term prediction.

The lead time to predict a rain event can vary based on applications, interests and location. Long term predictions including 2 days ahead of predictions are useful for long term planning of big events. Long term rainfall forecasting algorithms are mostly suitable for temperate climate where mostly stratiform rain events that last for a long duration are common. For a tropical climate, mostly convective rain events are common which lasts for a shorter duration. Within a duration of 6 hours there can be a multiple number of rain events. Likewise, short-term predictions of 5-30 minutes with higher accuracies are useful in cases where rainfall alarm in few minutes is enough to take preventive measures. The GPS-PWV values have a time resolution of 5 minutes. The rainfall nowcasting algorithm defined in this thesis aims at predicting rainfall event in next 5 minutes of time.

In literature, there are algorithms to forecast rainfall events solely based on GPS-PWV values from temperate [3] and subtropical [4, 15] regions. Less works have been reported for the tropical region. The objectives of this chapter and the subsequent chapters are to study the usefulness of GPS-PWV in rainfall prediction using data from the tropical region.

5.1.1 Outline of the Contribution

The main contributions of this chapter are listed as below:

- Study of the relationship between the GPS-PWV values and rainfall events for

data from the tropical region.

- Detailed review of the existing GPS-PWV based long-term rainfall prediction algorithm from the temperate and the sub-tropical regions and implementation and modification of the algorithms using the data from the tropical region.

The rest of this chapter is organized as follows. Firstly the time series plots of GPS-derived PWV and rainfall data are studied. Secondly, the existing long term rainfall forecasting models are reviewed. Thirdly, data from the tropical region are applied on the existing models and finally the algorithm suitable for the tropical region is discussed.

Database:

GPS PWV Data: GPS-PWV data are processed for 4 IGS GPS stations namely NTUS, SALU, RECF and IISC and one SiReNT GPS station namely SNUS. 6 years (2010-2015) of data from NTUS, 1 year (2016) of data from SNUS, 2 years (2016-2017) of data from SALU, 1 year (2017) of data from RECF and 1 year (2010) of data from IISC are processed for this Chapter. The details about these IGS GPS stations and SiReNT station are mentioned in Table. A.1 and Table. A.2 respectively of appendix A.

Weather Station Data: We use the data from the weather stations located at different regions. Table. A.3 of Appendix A gives details about the different weather stations, their location, weather parameters that we can access from the individual weather station, their time resolution and their availability.

We use data from 5 weather stations located in Singapore, Brazil and India. In Singapore, the two weather stations are from Nanyang Technological University (NTU) and at National University of Singapore (NUS). Fig. 5.1 shows the weather station which is installed at the NTU rooftop. From the weather station installed at NTU, we use the weather parameters like temperature ($T, ^\circ\text{C}$), relative humidity ($RH, \%$), dew point temperature ($DPT, ^\circ\text{C}$), solar radiation ($SR, \text{W/m}^2$) and

rainfall rate (RR, mm/hr). The weather station data at NTU are recorded at an interval of 1 minute and the rain data is recorded by the tipping-bucket rain gauge with a resolution of 0.2 mm/tip. The weather station data at NUS are recorded at an interval of 5 minutes and the rain data are recorded by the tipping-bucket rain gauge with a resolution of 0.2 mm/tip. The weather station data from NUS are accessible from the given link [140].

Figure 5.1: Weather station located at NTU rooftop (1.3 N, 103.68E)

From Brazil, we use rainfall rate data recorded by stations located at São Luis and Recife. The station coordinates details are in Table. A.3. The rainfall rate data are made available by the Cemaden's observational network for natural disaster risk monitoring, Brazil [141]. The rain gauge has a resolution of 0.1 mm/tip and the data are informed in every 10 minutes slot if the rain is registered else the data are recorded in an hour's interval [141].

We also use the rainfall data from India. The station is located at Bangalore. The rainfall data was purchased from National Data Center, India Meteorological Department (NDC-IMD). Hourly rainfall data is made available from NDC-IMD.

For this Chapter, the weather station rain data is processed for the same time period as that of the GPS-PWV data.

5.2 Time Series of GPS-Derived PWV and Rain

In this section, time series for rainfall and GPS-derived PWV values are presented. Time series is an important step as it visually shows correlation between two or more parameters. It also helps to explain the variations in different parameters which might be dependent on time. It is a common practice amongst remote sensing analysts to observe the time series plots as a first step towards major findings [3,4,48]. For this time series analysis, we use the PWV values derived from the IGS-GPS station of Singapore (NTUS). We use the rain fall data recorded by the co-located weather station at NTU. The details about these stations are mentioned in Table A.1 and Table A.3 respectively.

Fig. 5.2 shows the time series plot of PWV (mm) and rainfall rate (mm/hr). PWV and rainfall data for four rainy days; 5th May 2010, 3rd Oct 2011, 21st May 2012 and 29th Oct 2013 and a non-rainy day, 24th Oct 2013 are shown in Fig. 5.2. The x-axis of all the subplots represents the time of the day in hours. The left y-axis for all the subplots represents the PWV in mm and the right y-axis represents the rainfall rate in mm/hr. The time series plots for rainy days clearly show the general trend in PWV with respect to a rainfall event. It can be observed that before each rain event, there is a slow rise in PWV values. The rise in PWV values before the rain event is a clear indication that a certain amount of water vapor needs to be accumulated in the atmosphere before saturation.

It is interesting to note the trend in PWV values before the start of a rain event. The PWV values slowly increase, generally reach a peak value and then drop slightly before the start of a rain event and it continues to drop throughout the rain period. In Fig. 5.2(a), a significant rise in the PWV values can be noted at around 09 : 00 hours, the PWV values reach a peak at around 12 : 00 hours and decrease slightly 10 minutes before the start of the rain event. The PWV values then decreases throughout the rain. Similar observations are made for other rain events. The PWV values increase within 2-3 hours before the start of a rain event. In [15], for a sub-tropical station from China, it was found that the PWV values increase within 4 hours before the start of a rain event. Therefore, from this time series observation it

Figure 5.2: Time series plot of PWV (mm) for non-rainy days and rainy-days with rainfall rate (mm/hr) for IGS GPS station (NTUS). Time series for rainy-days are shown for 4 different days 5th May 2010, 3rd Oct 2011, 21st May 2012 and 29th Oct 2013. Time series for a non-rainy day is shown for 24th Oct 2013. Here X-axis represents the hours of the day. The red dashed line is the constant PWV value at 55 mm.

is clear that a higher amount of PWV values is an essential condition for a rainfall event to occur.

To illustrate more on this point we divide the PWV database according to the rain data. All the PWV values during and up to 3 hours before the start of the rain are separated as rain class and the remaining PWV values are separated as no-rain class. Fig. 5.3 shows the histogram of PWV values from these two classes. Fig. 5.3(a) shows the plot for rain class and Fig. 5.3(b) shows the plot for the no-rain class. Different colors on the plot represent the PWV values for different years as mentioned in the figure caption. The x-axis of the plots represent the PWV data (mm) and the y-axis represent the total number of occurrences. It should be noted that rain is a rare event as compared to the non-rain cases. Thus, the number of occurrences for non-rain cases are very high compared to the rain cases.

When Fig. 5.3(a) is compared to the Fig. 5.3(b), a clear shift can be observed between the PWV values for rain and no-rain classes. The PWV values for rain cases are higher than the PWV values for the non-rain cases. This is in-line with our time series discussion that the PWV values are important indicator of rain and higher PWV values are essential for rainfall initiation. The histogram plots clearly show that a PWV threshold value can be useful to classify the rain and the no-rain classes. Based on data from different years, a constant PWV threshold value of 55 mm can be chosen to differentiate rain and no-rain cases for this location (Singapore). This value is chosen as most of the PWV values for rain cases are greater than it and most of the PWV values for non-rain cases are smaller than it. The constant threshold is also plotted in the time series plot of Fig. 5.2. The red line in the plot shows the constant PWV threshold value of 55 mm. It can be noted that for all the rain events, the PWV values go beyond 55 mm whereas for a non-rainy day, the PWV values are in general below 55 mm. Although, the condition satisfies for most of the cases, there are overlapping cases between the two classes as shown by the histogram plot. These overlapping cases are for those conditions, where by PWV values are higher but no rainfall events occur. Many literature have rightly pointed out that for a rainfall initiation higher PWV values are essential but all instances with high PWV values do not ensure the occurrence of a rainfall event [48]. This indicates that the PWV values have potential to correctly predict a rainfall event but might give some amount of false predictions as well.

Figure 5.3: (a) Histogram of PWV values during rain (b) Histogram of PWV values for no rain conditions. The red line with diamond markers represent year 2010, the blue line with star markers represent year 2011, the green line with circle markers represent year 2012 and the black line with square markers represent year 2013.[Best viewed in color]

In [48], a constant PWV threshold value of 50 mm is used for a sub-tropical station for classifying the rainfall conditions, which is slightly lower than what we

have analyzed for a tropical station from Singapore. In [31], different values of PWV thresholds under different rainfall strengths are used. The average PWV thresholds are reported to be 41.8 mm, 52.9 mm, 62.5 mm and 64.4 mm for no rain, light rain, moderate rain and heavy rain conditions, respectively. Therefore, although a constant threshold value exists to separate rain and non-rain cases, the constant value changes based on different locations and also based on different seasons. Since there is a large number of non-rain cases, a slight change in the threshold value might cause a large number of PWV values to exceed the threshold value during non-rain cases which gives rise to a higher number of false classifications. Therefore, for a particular station the value must be optimized taking the seasonal factors into consideration. Different than the existing literature, in this thesis, we will propose a method to derive the season dependent PWV threshold values for rainfall prediction in tropical region in Chapter 6.

5.3 Review and Implementation of Existing Long-Term Rainfall Prediction Algorithms

It was observed that the GPS-derived PWV values can be a useful parameter in rainfall prediction. There are a few algorithms in literature that use the GPS-derived PWV values for a long-term rainfall prediction. In this section we implement two algorithms; developed using data from the temperate region [3] and using data from the sub-tropical region [4]. Both of these algorithms use the PWV data from a 6 hour time period to predict the rain within the next 6 hours. Fig. 5.4 shows a simple schematic diagram which explains the algorithm. The PWV values from window A, which has 6 hours of data is used to predict the rain within window A1, which is within next 6 hours of window A. The next prediction window, B is 1 hour after the first window, A. Based on the analysis of the PWV values from windows A, B, C and so on rain is predicted for windows A1, B1, C1 and so on.

Figure 5.4: Schematic diagram for a long term rainfall prediction. 6 hour PWV values from box A is used to predict rain within next 6 hours as indicated by box A1.

Figure 5.5: Flowchart of general rainfall prediction algorithm.

Evaluation Metrics:

Generally, if rain is predicted by window A and actual rain data from window A1 also shows the presence of rain then this is classified as a *true detection* case whereas if the

rain is predicted by the window A but the actual rain data from window A1 shows absence of rain then this is classified as a *false alarm*. Similarly, if window A1 shows rain but the algorithm fails to predict it, then it is classified as *miss-detection*. Fig. 5.5 shows all these cases in a simple flowchart form. The flowchart is self explanatory. Here, TP , FP , TN and FN are true positive, false positive, true negative and false negative values. These values are the results of different combinations of actual and predicted cases as shown in Fig. 5.5. Here the *true detection* (TD) is calculated as $TD = TP/(TP + FN)$ and *false alarm* (FA) is calculated as $FA = FP/(TN + FP)$. It is aimed to achieve a high true detection rate with as low false alarms as possible.

The different rainfall algorithms that have been proposed for the temperate and the sub-tropical regions use different conditions for the rainfall prediction. These conditions are discussed in detail and are applied on the data from the tropical region. The algorithms are applied on the data from the tropical stations NTUS, SNUS, SALU, RECF and IISC. The details about these IGS-GPS stations and the co-located weather stations are given in Table A.1 and Table A.3 respectively along with their data availability. Here it is important to note that the existing algorithms from the literature use the hourly PWV data (PWV data at end of each hour) and hourly rainfall rate. Therefore for a fair comparison, we sample the 5 minute PWV data into the hourly PWV data by taking the PWV data at end of each hour and we integrate the rainfall values to calculate the hourly rainfall rate. If the hourly rainfall rate is greater than 0, it indicates a rain event.

5.3.1 Rain Prediction Algorithm from the Temperate Region

Benevides et. al. proposed a rainfall prediction algorithm using data from a temperate region of Lisbon [3]. The algorithm uses the PWV values from a 6 hour window to predict rain within next 6 hours as described by Fig. 5.4. It uses the maximum rate of change of PWV (mm/hr) values as the condition for rainfall prediction. The algorithm is implemented on one year (2012) of data from Lisbon. A true detection rate of 75 % and a false alarm rate of 60 % are reported when the rate of change of PWV values exceeds a threshold of 1.5 mm/hr [3].

Region	Station	Year	Maximum PWV Rate (mm/hr)	<i>TD</i> (%)	<i>FA</i> (%)
Tropical	NTUS	2010	0.3	78.4	63.2
		2011	0.3	79.7	60.1
		2012	0.3	77.4	61.0
		2013	0.3	79.3	59.3
		2014	0.3	82.7	62.1
		2015	0.3	82.3	62.7
Tropical	SNUS	2016	0.3	74.3	66.5
Tropical	SALU	2016	0.4	76.1	64.4
		2017	0.3	76.7	66.7
Tropical	RECF	2017	0.3	74.4	63.3
Tropical	IISC	2010	0.3	76.8	60.7
Sub-Tropical [4]	LJSL	2015	0.6	76.2	65.2
Sub-Tropical [4]	ZHOS	2015	0.8	71.8	66.1
Sub-Tropical [4]	ZJPH	2015	0.6	74.3	68.3
Sub-Tropical [4]	ZJXC	2015	0.6	79.4	65.2
Sub-Tropical [4]	ZJYH	2015	0.6	71.6	66.1
Temperate [3]	Lisbon	2012	1.5	75	66.22

Table 5.1: *TD* and *FA* rates when using threshold of maximum rate of change of PWV values (mm/hr)

The algorithm is applied on the data from the tropical stations and the results are reported in the Table 5.1. The table reports the *TD* and *FA* results when the rate of change of PWV values exceed the threshold values mentioned in the table. The table shows that a similar range of *TD* and *FA* results can be achieved for the data from the tropical region at a lower threshold values of PWV increment. For the temperate station the results were obtained when using a threshold value of 1.5 mm/hr where as for the tropical stations, the threshold value is around 0.3-0.4 mm/hr.

Discussion:

The daily variations of the PWV values in the temperate region is higher than that in the tropical region. The time series plot in Fig. 5.2 shows that the PWV values

vary by 10 mm even for a rainy day in the tropical region, whereas from the similar time series plot for the temperate region [3], it was observed that the daily variation in PWV values can even be higher than 20 mm. Thus, given the same time frame of 6 hours, maximum PWV variation for the temperate region is greater than that of the tropical region. Therefore, the threshold values for the temperate region is higher than the tropical region.

5.3.2 Rain Prediction Algorithm from Sub-Tropical Region

This section discusses about the rainfall prediction algorithm implemented using the data from the sub-tropical region. Yao et al. [4] firstly used the same criteria as that of Benevides et. al algorithm [3] i.e. to use the maximum rate of change of PWV values for rainfall prediction and then proposed two additional criteria to improve the true detection rates for rainfall prediction in sub-tropical region of China. The proposed algorithm is called the three-factor algorithm which uses three conditions; a) Maximum rate of change of PWV values (mm/hr), b) Maximum variation in PWV values (mm) and c) Monthly PWV values (mm). These criteria are explained in detail as below:

- Maximum Rate of change of PWV is defined as; rate of change of PWV = PWV variation / interval epoch. Interval epoch = time1 (corresponding to the maximum PWV) - time2 (corresponding to the minimum PWV) [mm/hr].
- Maximum PWV variation is defined as; Maximum PWV variation = (the maximum PWV before a rainfall) - (the minimum PWV before the adjacent maximum PWV).
- Monthly PWV indicates the average PWV values of a month [mm].

Here out of the three criteria, the rate of change of PWV is taken as a major factor and the monthly PWV thresholds and the PWV variation thresholds are taken as the auxiliary factors to reduce the omission of a rainfall event and hence to improve the true detection rate. Fig. 5.6 explains the flow of the three-factor algorithm. The

Figure 5.6: flowchart to illustrate the three-factor algorithm proposed for the sub-tropical region.

major and the auxiliary factors are also indicated in the flowchart. As explained earlier, TP , FP , TN and FN are the results of different combinations of actual and predicted cases. The true detection (TD) and false alarm (FA) rates are calculated in the similar manner.

The algorithm was applied on the data from 5 sub-tropical stations of China (28.2° N - 36.8° N) and it was found that after the application of the three-factor algorithm the true detection rate was improved by 7 % where as the false alarm rate was found to be 66% which is comparable to the results of the Benevides et. al algorithm. Thus the additional two factors are to improve the true detection rate only.

In the following, the individual factors are assessed for the data from the tropical region and are compared to the results from the sub-tropical region.

a) Maximum Rate of Change of PWV Values (mm/hr)

This is the same criteria as described in the earlier section. When only this criteria was applied for the stations from the sub-tropical region, the results are similar to the temperate and the tropical regions. Table 5.1 reports the results for the sub-tropical stations. For a station, LJSJ when this criteria is applied a true detection rate of 76.2 % and a false alarm rate of around 65.2 % are obtained at a threshold value of 0.6 mm/hr [4]. Most of the sub-tropical stations have the threshold values of 0.6 mm/hr. Here the threshold value for the rate of change of PWV is higher than the tropical region but is lower than the temperate region. This follows our earlier discussion that the higher rate of threshold in temperate is because the PWV values have higher fluctuations in temperate region, which is lesser in the sub-tropical and the least in the tropical region.

b) Maximum Variation in PWV values (mm)

Maximum Variation in PWV values (mm) is the first auxiliary factor that is used for the improvement of the TD rates. In [4] only the true detection rates are reported when this criteria is applied for the rainfall prediction using data from the sub-tropical region. For comparison purposes, we apply this criteria (maximum variation in PWV values) on the data from the tropical region and report the true detection rates as tabulated in Table 5.2.

From Table 5.2, higher differences are observed when the TD rates are compared

Region	Station	Year	PWV Variation Threshold Values (mm)					
			0	1	2	3	4	5
Tropical	NTUS	2010	84.0	73.0	55.6	38.2	22.3	11.3
		2011	86.3	73.5	54.9	32.6	17.7	7.6
		2012	85.4	73.3	56.0	37.6	23.2	11.9
		2013	85.3	74.1	54.3	35.0	19.9	11.4
		2014	85.7	75.1	55.7	35.1	20.1	10.9
		2015	87.4	76.0	60.6	38.1	21.3	12.4
Tropical	SNUS	2016	80.9	68.1	47.8	29.5	18.5	9.9
Tropical	SALU	2016	80.9	73.2	59.6	40.6	25.8	16.1
		2017	80.4	71.7	58.8	38.9	25.5	15.5
Tropical	RECF	2017	78.3	69.1	53.8	38.4	27.1	18.9
Tropical	IISC	2010	82.0	67.2	49.4	31.6	16.5	7.9
Sub-Tropical [4]	LJSL	2015	92.8	87.7	79.3	70.2	61.8	55.2

Table 5.2: *TD* rates when using threshold of maximum variation in PWV values (mm)

at different threshold values of maximum variation between the tropical and the sub-tropical stations. For the sub-tropical station it is found that 92.8 % of the rainfall events have variations in PWV values that are greater than 0 i.e. for almost all the rainfall events there is a trend of increase in the PWV values within 6 hours before the start of the rain. Whereas for the tropical station NTUS, on average 85.6 % of the rain events have PWV variation that is greater than 0, for the tropical station SNUS, 80.9 % of the rain events have PWV variation that is greater than 0, for the tropical station SALU, on average 80.5 % of the rain events have PWV variation that is greater than 0, for the tropical station RECF, 78.3 % of the rain events have PWV variation that is greater than 0 and for the tropical station IISC, 82.0 % of the rain events have PWV variation that is greater than 0.

Discussion:

Here, these results suggest that for most of the tropical stations almost 20 % of the rain events do not have a trend of increase in the PWV values in the 6 hour period before the start of a rain event. But it is important to note that for this

analysis, the PWV values are sampled at an hour's interval. Each hourly PWV value represents the value at the end of the given hour. Since for the tropical region, significant variations in PWV values lie closer to the start of the rain event, the hourly sampling of the PWV data can miss the significant PWV variations. Whereas for the sub-tropical region the PWV increment is observed from few hours to tens of hours before a rain event. Therefore hourly data is still sufficient to capture the significant PWV variations.

From Table 5.2, it can be noted that when the threshold values for the PWV variation are increased from 0, a significant drop in the TD rates can be observed for the tropical stations whereas the corresponding drop is smaller for the sub-tropical station. This is in-line with our previous argument where we discussed that the PWV values in the tropical region have lesser variation compared to that of the temperate and the sub-tropical region. Therefore for tropical region, the threshold values for the PWV variation should change with finer time resolution.

c) Monthly PWV values (mm)

Monthly averaged PWV value is the second auxiliary factor that is used for the improvement of the TD rates. In this section we use the monthly averaged PWV values for the data from the tropical region and compare the results with that of the sub-tropical region.

Fig. 5.7 shows the TD and FA rates for the tropical stations. The stations and the corresponding years are indicated by each figure title. In Fig. 5.7, the solid line with circle markers represent the TD and the broken line with circle markers represent the FA for the sub-tropical station LJSL [4]. The solid line with square markers represent the TD and broken line with square marker represent the FA for the respective tropical station. The left y-axis shows the TD and FA rates. The upper x-axis shows the monthly averaged PWV values for the respective tropical station and the lower x-axis shows the monthly average PWV values for the sub-tropical station LJSL [4]. Here the monthly PWV values of the respective GPS stations show the variation in PWV values based on its location within the tropical

Figure 5.7: TD and FA rates for for different IGS-GPS stations indicated by individual figure title, when the monthly average PWV values are applied for the rainfall prediction. The details about the figure are given in the corresponding text. [Best viewed in color]

region. It can be seen that the stations NTUS and SNUS from Singapore have the highest PWV values amongst all as they are located in a small tropical island (Singapore). The station SALU from Brazil has higher PWV values compared to the

station RECF from Brazil, as SALU is closer to the coast. The station IISC from India has the lowest PWV values as it is located in the main-land area.

From the plots it can be clearly seen that the use of the monthly averaged PWV values show a good separation between the TD and FA rates for all the tropical stations whereas the monthly averaged PWV values do not contribute much in separating the TD and FA rates for the sub-tropical station. For different tropical stations, when the monthly PWV values are used, the TD rates are always higher than the FA rates whereas for the sub tropical station, the TD and FA rates are nearly the same and for few months even the FA rates can be greater than the TD rates.

Discussion:

When only monthly averaged PWV values are used, both TD and FA are high for the sub-tropical region. Therefore, the proposed algorithm in [4] uses the monthly averaged PWV values as an auxiliary factor to improve the TD rates only. But for the tropical region as the monthly averaged PWV values show promising results in controlling the FA rates, the average PWV values can be meaningful to both improve the TD rate and to decrease the FA rates as well.

Implementation of the three-factor method

As proposed by the algorithm [4], the maximum rate of PWV increment is taken as the major factor and maximum PWV variation and monthly averaged PWV values are taken as the auxiliary factors and are applied to the data from the tropical stations. The results are reported in Table 5.3. From these results, it can be seen that the use of the three-factor method improves the true detection rate for the tropical stations but the improvement is less compared to the sub-tropical stations. As can be seen from the table, for sub-tropical stations, the maximum improvement is around 8.5 % and on average the TD rates improve by 5.7 %. Whereas for the tropical stations, the maximum improvement is around 2.9 % and on average the improvement is less than 2 %. Therefore, for the sub-tropical stations use of both

Region	Station	Year	TD (%)		FA (%)
			Major factor Only (mm/hr)	Three-factors	
Tropical	NTUS	2010	78.4	80.5	63.2
		2011	79.7	81.9	60.1
		2012	77.4	80.3	61.0
		2013	79.3	81.6	59.3
		2014	80.7	82.9	62.1
		2015	82.3	83.6	62.7
Tropical	SNUS	2016	74.3	76.0	66.5
Tropical	SALU	2016	76.1	77.9	64.4
		2017	76.7	77.6	66.7
Tropical	RECF	2017	74.4	75.3	63.3
Tropical	IISC	2010	76.8	78.4	60.7
Sub-Tropical [4]	LJSL	2015	76.2	82.3	65.2
Sub-Tropical [4]	ZHOS	2015	71.8	80.3	66.1
Sub-Tropical [4]	ZJPH	2015	74.3	77.7	68.3
Sub-Tropical [4]	ZJXC	2015	79.4	81.5	65.2
Sub-Tropical [4]	ZJYH	2015	71.6	80.4	66.1

Table 5.3: TD and FA rates when the three factors are implemented for the Tropical and Sub-Tropical stations

the maximum variation and monthly averaged PWV values complements each other and improves the true detection rates. But for the tropical stations, as we analyzed from Table 5.2, with hourly PWV values, the maximum variation does not contribute much in improving the TD rates. Hence when using both the maximum variation and monthly PWV values the improvement in the TD rate is not as high as for the sub-tropical stations.

5.4 Long-Term Rainfall Prediction for Tropical Region

The existing literature pointed out three possible factors derived from GPS-PWV values which can be useful for rainfall prediction; maximum rate of increment of

PWV (mm/hr), maximum variation in PWV (mm) and monthly averaged PWV values (mm). When the factors were implemented individually on the data from the tropical region, it was found that the factors like maximum increment and monthly averaged PWV values ensure good detection rate with lower false alarms compared to the factor of maximum PWV variation. Here, it should be noted that, it was argued in the previous section that the sampling of the PWV values at an hour's interval causes the loss of important information w.r.t rainfall events especially for the tropical region as the PWV values show important fluctuations near to the start of a rain event. Therefore, in this section, we use the GPS-PWV values sampled at 5-minutes interval and analyze the *TD* and *FA* rates.

5.4.1 Optimum Threshold Criteria for The Tropical Region

In this section, the *TD* and *FA* rates are evaluated for the 6 hour rainfall prediction using GPS-PWV data with a temporal resolution of 5 minutes for the tropical region. Fig. 5.8 shows the *TD* and *FA* rates when the individual parameters are used. The graphs are plotted using data from NTUS GPS station from years 2010-2013.

Maximum Rate of Increment of PWV Values (mm/hr):

First column of Fig. 5.8 shows the plots when maximum rate of PWV is used. Previously, when hourly data was used, the maximum rate of PWV was at around 0.3 mm/hr (ref. Table. 5.1) for NTUS station. Now when the 5 min data is used, similar results are obtained at higher threshold values of around 0.5-0.6 mm/hr (ref. Fig. 5.8).

Maximum Variation of PWV Values (mm):

Second column of Fig. 5.8 shows the plots when maximum variation of PWV values are used. Here it can be noted that when 5 minute PWV values are used, the *TD* rates can go beyond 90 % at a PWV variation threshold of greater than 0. Previously, when hourly data were used, the *TD* rates for PWV variation greater than 0 were around 85 % for the given years (ref. Table 5.2). Therefore, these results verify

our earlier arguments that the hourly sampling of PWV values could have missed the important information w.r.t a rain event. Especially this is seen for the tropical station as significant changes in PWV values are noted near to the start of a rain event.

Maximum PWV Values (mm):

Third column of Fig. 5.8 shows the plots when maximum PWV values are used. The first two parameters are the same as that of the three-factor method [4]. However here we change the third parameter. In the previous section (Section 5.3.2), it was found that the monthly averaged PWV factor performs well in the tropical region as compared to the sub-tropical region. This indicated that the absolute PWV threshold values can be significant in predicting a rainfall event in the tropical region. Also, in the Section 5.2 of this Chapter, it was discussed that the PWV values increase within few hours before the start of a rain event and the histogram plot of Fig. 5.3 showed a clear separation between the PWV values for rain and no-rain cases. Therefore, in this section, instead of using a monthly averaged PWV value threshold, a maximum PWV value of the given 6 hour period is used in predicting the rainfall within next 6 hours.

Station	Year	TD (%)	FA (%)
NTUS	2010	90.1	69.9
	2011	88.1	60.3
	2012	88.2	62.9
	2013	89.5	68.0

Table 5.4: TD and FA rates when threshold of maximum PWV (55 mm) values are used on data from the tropical region.

The TD and FA rates plotted in the third column of Fig. 5.8, show a very good separation between the true detection and the false alarm rates. The plots from different years for NTUS GPS station show a turning point at around 55 mm of maximum PWV threshold values. At 55 mm, the TD and FA values are around 90.1 % and 69.9 % for year 2010, around 88.1 % and 60.3 % for year 2011, around

Figure 5.8: TD and FA rates when using different criteria for NTUS GPS station.

88.2 % and 62.9 % for year 2012 and around 89.5 % and 68.0 % for year 2013. The results are tabulated in Table 5.4. These results outperform the results that were reported for hourly GPS-PWV values when either an individual parameter was used (ref. Table 5.1 and Table 5.2) or the three-factor method was used (ref. Table. 5.3). These results are also better compared to the results discussed for GPS-PWV with 5 minutes resolution, when threshold for maximum rate and maximum variation are used. Therefore, these results suggest that the threshold of maximum PWV value can be the main factor for rainfall prediction in the tropical region, unlike the algorithms from the temperate region [3] and the sub-tropical [4] region, which use the threshold of maximum rate of increment of PWV as the main factor.

Results for the Tropical Region :

In the three-factor method [4], the rate of increment of PWV (mm/hr) was used as the main factor and the maximum variation in PWV (mm) and the monthly PWV values (mm) were used as the auxiliary factors. The auxiliary factors were introduced to improve the *TD* rates only but its effects on the *FA* rates were not reported. Therefore, the number of false alarms that might arise in real time when the three-factor method is implemented, is unknown.

For the tropical region, we have analyzed that the threshold of maximum PWV values can be an important factor for rainfall prediction and the threshold of maximum increment of PWV and the threshold of maximum PWV variation show almost similar characteristics. Therefore, for the tropical region, we use the threshold of maximum PWV values and the threshold of maximum rate of increment of PWV values for rainfall prediction. Both the criteria are used at the same time with no segregation as main and auxiliary factors and the evaluation metrics are reported.

Fig. 5.9 shows the the *TD* and *FA* values when the above mentioned threshold criteria are used. The title of each figure shows the IGS GPS station and the corresponding year. The x-axis of all the figures show the changing threshold values of maximum rate of PWV increment (mm/hr). The threshold of maximum PWV values (mm) which have been used to derive the results are mentioned in each plot.

The threshold of maximum PWV values were determined for each station by plotting and analyzing the figures similar to the third column of Fig. 5.8. The maximum PWV value for which the sharp bend was observed was chosen as the threshold value. A threshold value of 55 mm was chosen for the NTUS GPS station (ref. Fig. 5.8). Similarly, the threshold values for other stations were chosen. The chosen threshold values are 55 mm, 56 mm, 50 mm, 40 mm and 35 mm for stations NTUS, SNUS, SALU, RECF and IISC respectively. It can be observed that the threshold values are higher for stations from Singapore (SNUS and NTUS). For stations from Brazil, SALU has higher threshold value compared to RECF. In overall, station IISC has the lowest threshold value. These observations are in-line with our discussion in section 5.3.2, where it was observed that the monthly averaged PWV values vary as per the location of the stations (coastal and mainland).

From Fig. 5.9, a particular threshold value for the maximum increment of PWV values is chosen such that a good detection rate and fairly low false alarm is maintained, Table 5.5 shows the final TD and FA rates for the tropical stations.

Station	Year	Threshold Values		TD (%)	FA (%)
		Max. PWV (mm)	Max. Rate (mm/hr)		
NTUS	2010	55	0.3	80.8	56.3
	2011	55	0.3	79.4	46.9
	2012	55	0.2	80.6	51.1
	2013	55	0.3	79.3	52.0
	2014	55	0.3	79.4	45.8
	2015	55	0.3	80.2	43.5
SNUS	2016	56	0.3	79.3	51.5
SALU	2016	50	0.4	81.0	39.1
	2017	50	0.3	85.6	37.4
RECF	2017	40	0.3	73.2	48.5
IISC	2010	35	0.3	75.0	43.1

Table 5.5: TD and FA rates when threshold of maximum PWV values and threshold of maximum rate of PWV increment are used on data from the tropical region.

The evaluation metrics reported in Table 5.5 show a significant improvement

Figure 5.9: *TD* and *FA* rates when the threshold of maximum PWV values and the threshold of maximum increment in PWV values are used.

over the three-factor method. The TD and FA rates for three-factor method and the proposed method are tabulated in Table 5.6. From the table it can be clearly observed that the proposed method decrease the FA rates significantly with very less effect on the TD rates. There is enhancement in the TD rates for few tropical stations as can be observed from the table. From Table. 5.6, it can be observed that when the proposed method is used, for NTUS station, the FA rates decrease by up to 10 % with a slight decrement of 1.85 % in the TD rates in average. For SNUS station, the FA rates decrease by 15 % and the TD rates are enhanced by 3.3 %. For SALU station, the FA rates decrease by 27.3 % and the TD rates are enhanced by 5.5 % in average. For RECF station, the FA rates decrease by 14.8 % with a slight decrement of 2.1 % in the TD rates. And for IISC station, the FA rates decrease by 17.6 % with a slight decrement of 3.4 % in the TD rates. On average, for all stations, the FA rates decrease by 16.9 % with negligible effect on TD rates. Therefore, for the tropical region, better results are obtained compared to the existing literature [3,4] when the threshold of maximum PWV values and the threshold of maximum increment in PWV values are applied using GPS-PWV values with higher temporal resolution (5 minutes).

Station	Year	Three-factor Method		Proposed Method	
		TD (%)	FA (%)	TD (%)	FA (%)
NTUS	2010	80.5	63.2	80.8	56.3
	2011	81.9	60.1	79.4	46.9
	2012	80.3	61.0	80.6	51.1
	2013	81.6	59.3	79.3	52.0
	2014	82.9	62.1	79.4	45.8
	2015	83.6	62.7	80.2	43.5
SNUS	2016	76.0	66.5	79.3	51.5
SALU	2016	77.9	64.4	81.0	39.1
	2017	77.6	66.7	85.6	37.4
RECF	2017	75.3	63.3	73.2	48.5
IISC	2010	78.4	60.7	75.0	43.1

Table 5.6: Comparison of TD and FA rates when the three-factor method and the proposed methods are implemented for the Tropical stations.

5.4.2 Lead Time Values for Rainfall Prediction for Tropical Region

The rainfall prediction results discussed till now are for prediction within the given 6 hours time. When we talk about prediction within a given time frame, we are not sure of whether we are predicting a rain event which will occur after 5 minutes or after an hour or after 4 hours within the given 6 hours time frame. Therefore in this section, statistical results are presented to show how far ahead of time a rainfall event can be predicted by using the GPS-PWV values. It is interesting to present and discuss these results as it helps to explore the usefulness of GPS-PWV values in predicting a rainfall event with certain lead-time value and also this has not been discussed in any literature by far to our knowledge.

Firstly, a rainfall event is defined. Any number of rainfall events that occur within a duration of 6 hours or less is considered as a single rain event. Therefore, a minimum separation time between two rain events is 6 hours. For eg. the total number of rainfall events for NTUS station with this criteria are 154, 133, 155, 180, 156 and 154 for years 2010-2015 respectively.

Secondly, the time frame of the PWV data considered to predict a single rainfall event are varied. Previously, a constant time frame of 6 hours was used, now the time frame varies from 10 minutes to upto 6 hours. For each time frame, the maximum increment of PWV value (mm/hr) and the maximum PWV value (mm) are recorded. These values are then compared to the respective threshold values derived in the earlier section (ref. Table 5.5) and a rainfall is predicted if the conditions are satisfied. The predicted rainfall events are then compared to the ground truth and a true detection rate is calculated. The TD rates are then plotted against the length of data considered as shown in Fig. 5.10 (a). The minimum length that can be considered is 10 minutes as the resolution of GPS-PWV is 5 minutes and to calculate the increment rate at least two data points are needed.

In the figure, the first point corresponds to a time frame of 10 minutes, which includes two GPS-PWV data points before the start of the rain event, the second point corresponds to a time frame of 15 minutes, which includes three GPS-PWV

Figure 5.10: TD rates for NTUS for (a) different length of data used (b) for different lead time values. [Best Viewed in Color]

data points before the start of the rain event. Similarly, time frame corresponding to 1 hour indicates the use of twelve GPS-PWV data points before the start of the rain event. It can be observed from the figure, that the TD rate corresponding to the 10 minutes length of data (first point) is less than 40 %. The TD rate gradually increases as the length increases from 10 minutes to 1.5 to 2 hours after which, the TD rates saturate. This analysis shows that a time frame of 2 hours is an optimum value and adding in more PWV values do not contribute much to improve the detection rate.

Thirdly, the lead time values are varied. Lead time is defined as the interval before the start of the rainfall event whose corresponding PWV values are not used in the analysis. A lead time value of 10 minutes, indicates that the PWV values corresponding to the 10 minutes interval before the start of the rainfall event are not used for the analysis; the PWV values before this lead time value are used for the prediction. For each lead time value the length of the data used for prediction is held fixed at 2 hours. The TD rates are then calculated based on the two criteria of maximum rate and maximum PWV values. The TD rates are plotted in Fig. 5.10 (b). From Fig. 5.11, it can be clearly observed that in average over different years, more than 80 % of the total rainfall events can be successfully predicted with a lead time value of up to 45 minutes to 1 hour. As the lead time values increase beyond

an hour, the true detection rates decrease.

Figure 5.11: Zoomed in version of Fig. 5.10, with two hours of time frame for both the figures (a) different length of data and (b) different lead time values. [Best Viewed in Color]

Figure 5.12: TD rates for SALU for (a) different length of data used (b) for different lead time values. [Best Viewed in Color]

Similarly, the experiment was done on the two year data from the SALU station. For SALU station, the total number of rainfall events with 6 hour as the separation

time is 150 and 122 for years 2016 and 2017 respectively. The TD rates for different length of data and different lead time values are shown in Fig. 5.12 (a) and (b) respectively. Similar conclusions can be drawn for SALU station as well. More than 85 % of the total rainfall events can be successfully predicted with a lead time value of up to 45 minutes to 1 hour.

5.5 Summary

A detailed study of the PWV values w.r.t rainfall was performed. The time series plots of PWV and rainfall suggested that the PWV values increase before a rainfall event. A good separation was noted between the PWV values during rain and no-rain cases. Different existing algorithms for rainfall prediction, which use the GPS-PWV values from the temperate and the sub-tropical regions were reviewed and were implemented using data from the tropical region. For the tropical region, the threshold of maximum PWV values was found to play an important role in good detection of rainfall events, in contrast to the existing algorithm. The new criteria were applied on the data from the tropical region and the results were compared with the results from the existing literature. A significant improvement was noted. The FA rates decreased by up to 16.9 % with negligible effect on the TD rates. Moreover, the TD rates were studied by varying the length of the data used for prediction and the lead time values. It was found that using 2 hours of historical data rainfall events can be predicted up to 45 min - 1 hour ahead of the time. In the future, the rainfall prediction will be extensively studied using other ground-based parameters.

Chapter 6

GPS-Derived PWV Values for Rainfall Nowcasting

6.1 Introduction

In Chapter 5 different long-term rainfall prediction algorithms [3, 4] were reviewed and were applied on the data from the tropical region. The algorithm uses the data from a 6 hour window to predict the rain in the next 6 hour's time. The algorithm was only applied for long-term rain prediction scenario and was not tested for rainfall nowcasting scenarios. Particularly, for tropical regions the study of rainfall nowcasting is important because tropical regions mostly experience the convective rain events which are frequent in number and last only for short durations. Thus, in real case scenarios, a time window of 6 hours might hold more than one rain event for tropical regions. In this chapter, we perform rainfall nowcasting (a short-term rainfall prediction) by using GPS-derived PWV in a tropical climate.

6.1.1 Outline of the Contribution

It is well-noted that the rainfall cases in the temperate region are mainly stratiform rain and are commonly associated with low rainfall rates, whereas the rainfall events in the tropical region are mainly convective rain and are commonly associated with

high rainfall rates [127]. Studies show that more than 80 % of the rain events in tropical region like Singapore are convective rain events [127]. In general, these convective rainfall events are short lived with a lifespan of less than 30 minutes [131]. Therefore in this chapter,

- We extensively study the PWV values from the tropical region and propose an algorithm to nowcast a rain event in next 5 minutes given the PWV data of past 30 minutes.
- The proposed algorithm addresses the seasonal characteristics of GPS-PWV values to nowcast rainfall in a tropical climate, which has not been reported previously.

Database:

In this chapter, the GPS-PWV values from IGS-GPS station NTUS and the rain data from the co-located weather station are used in the proposal of the algorithm. The GPS-PWV values from IGS-GPS station SALU and SiReNT GPS station SNUS and the rain data from the respective co-located weather stations are used in the evaluation of the proposed algorithm. The details about the IGS-GPS stations, SiReNT GPS station and the weather stations are given in Table A.1, Table A.2 and Table A.3 respectively of Appendix A.

6.2 PWV and Rainfall

In Chapter 5, we studied the time series observation of GPS-PWV values and rainfall. The PWV values were studied by separating into the rain and the no-rain classes. But the seasonal characteristics of the PWV values were not studied. It is important to consider the seasonal characteristics for rainfall prediction as rainfall initiation are dependent on different seasons. We explained that the rainfall event is dependent on moisture content and saturation values. And we know that the saturation values are dependent on temperature. For different seasons the temperature values are

different. Generally speaking, winter and summer seasons are because of the different temperature values. Therefore, the condition for rainfall must change for different seasons. In this section, we study the seasonal dependency of PWV values and rainfall.

6.2.1 Representation of PWV and Rainfall Data

The IGS-GPS station NTUS, which is used in proposal of the algorithm, is located at Singapore where there are four different seasons, namely, North-East (NE) monsoon generally during Nov.- Dec. and Jan.- Mar., first-inter (FI) monsoon generally during Apr.- May, South-West (SW) monsoon generally during Jun.- Oct., and second-inter (SI) monsoon generally Oct.- Nov. The period of these seasons varies a little from year to year which is updated in yearly weather report provided by the National Environment Agency, Singapore [142].

To study the diurnal and seasonal variations in PWV, 4 years' (2010 – 2013) of PWV data from the NTUS GPS station are firstly divided according to the different seasons; NE, SW monsoons and inter-monsoons (FI and SI). From each of these seasons, PWV values are separated based on rainy and non-rainy days. In this classification, any day with at least one rain event is defined as a rainy day. In four years, there is a total of 250 rainy and 275 non-rainy days in NE, 222 rainy and 357 non-rainy days in SW and 191 rainy and 166 non-rainy days in inter-monsoon seasons.

The PWV data of all the rainy days are composited w.r.t the time of the day. All the PWV values at each time interval is then represented by a boxplot. PWV values are available every 5 min but to make the plot readable, the boxplots in Fig. 6.1 (a-c) are shown at every hour of the day. The red line in the boxplot represents the median, the upper bounding box represents the 75th percentile and the lower bounding box represents the 25th percentile of all the PWV values at the given hour of the day. The mean of all the rainy days' PWV values at the given time is also shown in cyan color with circle marker in Fig. 6.1 (a-c). The PWV values of all the non-rainy days of the respective seasons are averaged at each time stamp and are

shown in Fig. 6.1 (a-c) by a solid curve in green. The occurrences of a rain event at different time of the day for NE, SW and inter-monsoon seasons are plotted in Fig. 6.1 (d-f) respectively. There is a total of 399 rain events in NE, 280 rain events in SW and 234 rain events in inter-monsoon seasons.

Figure 6.1: Diurnal and seasonal variation in PWV and rainfall data for NTUS GPS station (2010 - 2013). The boxplots of PWV values with average of rainy (cyan line with circle markers) and non- rainy days (solid light green line) (a) NE, (b) SW and (c) inter-monsoons. Dashed black lines represent a threshold value of (a) 55 mm for NE, (b) 54 mm for SW and (c) 56.3 mm for inter-monsoons. The total number of rain events occurring at specific hours of the day (d) NE, (e) SW and (f) inter-monsoon seasons. The time for both GPS and weather station for NTUS station is recorded in local time format (UTC+8).

6.2.2 Seasonal and Diurnal Variations in PWV

From Fig. 6.1 (a-c), it can clearly be observed that the average PWV values for rainy days are higher than that of the non-rainy days for all the monsoon and inter-

monsoon seasons. It is interesting to note that although there is a clear shift in the amplitude of average PWV values, the average PWV values tend to follow similar trend based on the time of the day for both rainy and non-rainy cases for the respective seasons, which is mainly affected and dominated by the diurnal variations of ZTD values [143].

In case of NE and SW monsoon seasons, it can be observed that the average PWV for both seasons starts with higher PWV values, which drops in the early morning and then starts to rise again with the temperature as the sun shines. For the SW monsoon the peak of PWV is at around mid-afternoon, 12 : 00 hour, after which the PWV values falls until late evening. Whereas for the NE monsoon, the peak of PWV value is during the afternoon period of around 15 : 00 hour. A possible reasoning for such behavior in PWV is because, for the NE monsoon season, late afternoon rainfall is often experienced, and thus the PWV values tend to increase around this time of the day. For the SW monsoon season, morning downpours are often experienced and thus the PWV values are higher in the morning. This is also illustrated in Fig. 6.1 (d-e), which show the total number of rain events of a specific season as a function of time of the day; Fig. 6.1 (d) represents the NE monsoon season and Fig. 6.1 (e) represents the SW monsoon season. Most of the rain events for the NE monsoon season occur during late afternoon whereas most of the rain events for the SW monsoon season occur during morning and early afternoons

As shown by Fig. 6.1 (c), PWV values during the inter-monsoon season are relatively higher compared to the monsoon seasons. There are several possible reasons as investigated. Firstly, in general, the inter-monsoon season has a higher temperature compared to that of the monsoon seasons. Inter-monsoon season can experience hot afternoons with maximum temperature going beyond 32°C [142] . As the temperature is higher, it can relatively hold more water vapor increasing the overall PWV values. Secondly, it is observed that the rate of occurrence of rain events in inter-monsoon season is significantly higher compared to that in the monsoon seasons. Lastly, it needs to be considered that the NE monsoon season consists of both dry and wet phases. The driest month, generally February, falls under the NE monsoon season, which slightly lowers the overall PWV values of the NE monsoon compared

to that of inter-monsoon season.

6.3 A Season Dependent PWV Threshold Value

From Chapter 5, it is noted that a proper PWV threshold value plays an important role in indicating a rainy condition. In [48], a constant PWV threshold value of 50 mm is defined to indicate a rainy situation, and in [4] a range of PWV values from 23 – 59 mm is used as threshold values for different months. In Chapter 5, a PWV threshold value of 55 mm was found suitable for the station from Singapore. It is clear that these threshold values vary according to seasons and locations. Thus, in this section it is of great interest for us to propose a method to define a season dependent PWV threshold value using the data from a tropical climate.

From Fig. 6.1 (a-c), a clear separation between average PWV values of rainy and non-rainy days can be observed. A constant PWV threshold value to indicate a rainy situation exists. Clearly, the PWV threshold value must change w.r.t the seasons as the inter-monsoon seasons have generally higher PWV values compared to the monsoon seasons. And amongst the monsoon seasons, in general, the NE monsoon experiences higher PWV values as compared to the SW monsoon. The average PWV values of rainy and non-rainy days also shows variations w.r.t the time of the day.

It is interesting to note from Fig.6.1 (a-c) that the maximum inter-quartile range for NE monsoon is 6.72 mm, for SW monsoon is 6.78 mm and for the inter-monsoon is 5.8 mm only. This indicates that the 50 % of the total number of PWV values at any instant of time is distributed within a small range of 6.4 mm in average. Since the 50 % of total number of PWV values are distributed in a small interquartile range, the rainfall prediction results can be very sensitive to the chosen PWV threshold value. A slightly higher PWV threshold value leads to a case where rain events can be undetected and therefore affect the true detection rate. Similarly, a slightly lower PWV threshold value can give rise to many false alarms. Therefore, a season dependent constant PWV threshold value is proposed to ensure high detection rate with a low false alarm rate, which satisfies the following two criteria; The criteria for choosing the threshold values,

- the threshold should be higher than the average PWV values of non-rainy days to make the false alarm as low as possible, and
- the threshold should be lower than the 25th percentile of rainy days' PWV values at the specific hours (during which the probability of rain occurrence is higher) of respective seasons to maximize the true detection rate.

According to the second criteria, a PWV threshold value is chosen in such a way that it gives priority to a specific time-of-day when the probability of occurrence of a rain event is higher. Such specific hours differ according to the seasons. From Fig. 6.1 (d-f) and also from yearly weather reports for Singapore [142], it is clearly noted that, in NE monsoon, afternoon and early evening rainfalls are very frequent. In SW monsoon, Sumatra squalls are experienced during pre-dawn to midday, and short-lived rain falls are often experienced in the afternoon. During inter-monsoon seasons, afternoon to early evening rain events are common.

It should be noted that the threshold value varies from one location to another, depending on the local seasonal and climatic conditions [2]. Therefore, it is important to carefully study the distribution of PWV at a given location to choose a proper threshold value. In the next subsection we will carefully analyze the characteristics of PWV values with respect to (w.r.t) the rainfall initiation time and apply the findings in the rain prediction algorithm.

6.4 Composite Analysis of PWV Derivatives with respect to Rainfall Initiation

In this section we consider all the rain events for the IGS-GPS station NTUS and study the PWV values w.r.t rainfall start time. To do so, first and second derivative of PWV values will be analyzed using the composite analysis method in [144].

(a)

(b)

Figure 6.2: (a) Composite analysis of first derivative of PWV (mm/5min) and (b) second derivative of PWV (mm/(5min)²). Red diamonds represent mean values of PWV derivatives for 208 rain events of year 2010, blue stars for 234 rain events of 2011, green circles for 219 rain events of 2012 and black squares for 252 rain events of 2013.

6.4.1 First Derivative of PWV Values:

The first derivative (FD) is calculated from the PWV values before and after the start time of each rain event. It is then composited in time, i.e., the mean of the first derivative values at every time stamp before and after the start of the rain event corresponding to all the rain events are calculated. The mean of the first derivative values is plotted in Fig. 6.2(a). The x-axis of the figure represents the time in minutes, where $x = 0$ is the start of the rain, $x < 0$ indicates minutes before the start of the rain and $x > 0$ indicates minutes during and after the start of the rain. Here, a 2-hour (120 minutes) time frame before the start of the rain and a 1-hour (60 minutes) time frame after the start of the rain are presented. The red diamonds represent the mean of the derivatives of the PWV values for all 208 rain events in the year 2010, blue stars represent the mean of the derivatives of the PWV values for all 234 rain events in the year 2011, green circles represent the mean of the derivatives of the PWV values for all 219 rain events of in the year 2012, and the black squares represents the mean of the derivatives of the PWV values for all 252 rain events in the year 2013.

From Fig. 6.2(a), it can be observed that the first derivative of PWV is positive until about 15 minutes before the start of the rain. This indicates that the PWV is accumulated in the atmosphere before the rain event. A significant increase in rate can be noticed after $x = -120$, which indicates in general for tropical regions, the rate of PWV accumulation increases around 2 hours before a rain event. It can be observed that the zero crossing of the first derivative occurs within 30 minutes before the start of the rain, after which, the PWV values start to decrease and derivative of the PWV values becomes negative. Here, the position of zero crossing of the derivative of the PWV values corresponds to the peak PWV values. Statistically, it is seen that for most of the rain events, the peak PWV values occur at around 15 minutes before the start of the rain. Here, the decrease in PWV values just before rain initiates and during the rainfall is because the water vapor condenses into liquid to form the rain clouds and rain drops. It is noted that the rate of change of first derivative of PWV values is sharp within 30 mins before the start of the rain. Thus,

rate of change of first derivative (i.e., second derivative) of PWV before the start of the rain can be an important feature for short term rain prediction and will be the focus of our study in the next section.

6.4.2 Second Derivative of PWV Values:

Like that of first derivative, the second derivative (SD) of the PWV values are calculated from the PWV values before and after the start of a rain event. The composite of the second derivatives are then plotted w.r.t time. Fig. 6.2(b) shows the composite analysis of the second derivative of PWV values. In tropical climate, PWV values are generally higher and show very little day-to-day variations. That is the reason the fluctuations induced in PWV values before a rain event starts are prominent and are easy to be distinguished. This is also reflected in Fig. 6.2(a) and 6.2(b). In general, the rate of change of PWV (first derivative) does not vary much which results in a small or negligible amount of variation in the second derivative of PWV values. Fig. 6.2(b) shows that generally the second derivative of PWV values fluctuate around 0. A prominent feature in Fig. 6.2(b) is noticed when the rain starts, the second derivative values of PWV reach a minimum value. There are significant fluctuations in the second derivative values 30 minutes before and 30 minutes after the start time of the rain event. During this one-hour period, the second derivative values are always less than 0. This feature of second derivative value attaining a minimum value within 30 minutes before the rain start time will be used as a criterion for rain prediction.

6.5 Rainfall Nowcasting

In this section, we propose a simple algorithm to perform nowcasting of the start of a rain event. PWV data from year 2010 to year 2013 for GPS station NTUS are used in the development of the empirical algorithm.

6.5.1 Rainfall Nowcasting Algorithm

Two different criteria will be used to perform nowcasting of the start of a rain event.

- PWV value criteria: Season-dependent PWV threshold values are proposed by using the two criteria described in Section 6.3. Following the two criteria, for NTUS GPS station, for the NE monsoon season, a threshold value of 55 mm is chosen as the constant PWV threshold as shown by the dashed black line in Fig. 6.1 (a), and for the SW monsoon season, a threshold value of 54 mm is chosen as shown by the dashed black line in Fig. 6.1 (b). Similarly, for the inter-monsoon season, a threshold value of 56.3 mm is chosen as shown by the dashed black line in Fig. 6.1 (c) so that both the criteria are fulfilled.
- Second Derivative (SD) value criteria: In previous section, it was observed that a minimum SD value is reached within 30 minutes before a rain starts. We therefore use this minimum SD criteria to predict the start of a rain. For NTUS GPS station it has been noted that the second derivative values of PWV falls to a minimum of about $-0.015 \text{ mm}/(5\text{min})^2$ (ref. Fig 6.2 (b)).

With both the PWV value and SD criteria, the rain prediction, P_{NTUS} for IGS-GPS station NTUS becomes true upon satisfying the given criteria as,

$$P_{NTUS} = \begin{cases} 1, & \begin{array}{c} SD \leq -0.015 \ \& \ PWV \geq 55 \ (NE) \\ \text{or} \\ SD \leq -0.015 \ \& \ PWV \geq 54 \ (SW) \\ \text{or} \\ SD \leq -0.015 \ \& \ PWV \geq 56.3 \ (FI\&SI) \end{array} \\ 0, & \text{Otherwise} \end{cases} \quad (6.1)$$

6.5.2 Evaluation Metrics

True detection rate and the false alarm rate are the two-evaluation metrics that is used to evaluate the proposed nowcasting algorithm.

1) True detection rate:

Six PWV data points before the start of each rain event (corresponding to 30 minutes of PWV values) are considered. An illustration of this prediction process is given in Fig. 6.3. Now for NTUS GPS station, eq. (6.1) is applied on these 6 PWV data points. Firstly, the maximum of these PWV values is recorded. Then as per eq. (6.1), we check if this maximum PWV value exceeds the threshold of 55 mm for NE or 54 mm for SW or if this maximum PWV value exceeds threshold of 56.3 mm for the inter-monsoons. If the condition is satisfied, second derivative values are calculated using the same 6 PWV data points and the minimum second derivative value is recorded. Now if the second derivative value is less than or equal to -0.015 then the start of a rainfall event is treated as predicted. This will then be classified as a true detection. If the maximum PWV value and the minimum second derivative criteria are not satisfied together, the start of the rain event will be classified as no detection.

Figure 6.3: Illustration of data points considered for rainy and non-rainy cases. A window size of 6 data points (30 min) is used for both the cases.

Figure 6.4: Flowchart to explain the rainfall nowcasting algorithm.

2) False alarm rate:

For the non-rain cases, the PWV data points of non-rain days, PWV data points 30 minutes before the start of the rain event and PWV data points after the end of the rain events are considered. All these non-rainy data points are grouped such that each group has 6 PWV points which corresponds to a time frame of 30 minutes. This can also be seen from the illustration diagram in Fig. 6.3. For each window, the maximum PWV value and the minimum second derivative values are calculated, i.e. eq. 6.1 is applied. If the criteria are satisfied, it is considered as a false alarm.

The algorithm is also explained with the help of the flowchart as shown by Fig.

6.4. The flowchart is in accordance to the illustration diagram shown in Fig. 6.3. The definition of different variables that result from the combinations of the actual and predicted cases are same as described in Chapter 5.

6.5.3 Results And Discussion

Using the given procedure, the true detection rate and the false alarm rate have been calculated for all cases from the year 2010 to the year 2013 for NTUS GPS station. The results are summarized in Table 6.1. It is observed that for different years, the true detection rate is very high, around 88.9 %, 85.9 %, 88.5 % and 84.9 % for 2010, 2011, 2012 and 2013 respectively. The corresponding false alarm rate is only 49.1 %, 37.2 %, 36.4 % and 43.7 % for the same years respectively.

Year	No. of Rain Events (NE + SW + Inter-Monsoon Rain Events)	True Detection (%)	False Alarm (%)
2010	208(76 + 74 + 58)	88.9	49.1
2011	234(126 + 55 + 53)	85.9	37.2
2012	219(89 + 72 + 58)	88.5	36.4
2013	252(108 + 79 + 65)	84.9	43.7
2014	231(65 + 102 + 64)	89.1	34.6
2015	222(92 + 63 + 67)	89.1	31.0
Avg		87.7	38.6

Table 6.1: True detection and false alarm rate when using the proposed algorithm for rain prediction for NTUS station

Here, it is important to note that the certain percentage of the total false alarm rate might be the result of comparing the point rainfall rate recorded at NTU with the PWV values which is an average over a cone of 5 to 6 satellite paths. Therefore, some false alarms might not be false alarm but the detection of rainfall that might be occurring at a location some distance away from NTU along one or more of the satellite paths within the cone causing an increase in the average zenith PWV value. Thus, in such cases, the average zenith PWV values might satisfy the given criteria, indicating a buildup of rain clouds but the rain is not at the point location of NTU, resulting in a false alarm.

6.5.4 Performance Validation Using Data from Independent Years at the Same Station

For validation purposes, the proposed algorithm is applied on a set of independent PWV data for NTUS station. Data from years 2014 and 2015 which were not used in previous threshold derivation are used for performance validation here. It is found that the proposed prediction algorithm works well with a true detection rate of 89.1 % for both 2014 and 2015 and with a false alarm rate of only 34.6 % and 31.0 % for 2014 and 2015 respectively.

The six years result show that the proposed algorithm has an average true detection rate of 87.7 % and false alarm rate of 38.6 % only. The true detection rate of 87.7 % suggests that the proposed algorithm can detect most of the rain events without significant misses. The reported true detection rate and the false alarm rate show a significant improvement over the true detection and false alarm rates of 75 % and 60 – 70 % respectively as reported in [3] and the true detection and false alarm rates of 80 % and 66 % respectively as reported in [4]. The values are also tabulated as shown in Table. 6.2.

	From Literature [3]	From Literature [4]	Proposed Method
TD (%)	75.0	80.0	87.7
FA (%)	60 – 70	66	38.6

Table 6.2: Comparison of TD and FA rates from literature and from proposed method.

Thus, the comparison results show that the proposed prediction method is reliable with better performance. It is also simple to implement and easy to use and is applicable for the nowcasting of the start of a rain event.

6.6 Evaluation of the Proposed Algorithm

For the evaluation purpose the proposed algorithm is implemented using data from another GPS station from within Singapore and from Brazil. The results are reported

in the subsequent sections.

6.6.1 Performance Validation Using Data from Another Station in Singapore

In this subsection, one-year (2016) data from SNUS GPS station is processed for performance validation of the proposed method. Since both the stations, NTUS and SNUS are geographically close to each other (16 km apart), SNUS station shall experience similar weather variations as that of NTUS station. Therefore, we apply the thresholds in eq. (6.1) on the SNUS data where a total of 195 rain events were recorded in the year 2016, a true detection rate of 90.7 % and a false alarm rate of 50.2 % are observed. This is a meaningful improvement over recent works by Zhao et al. [15], where their proposed method with site-dependent thresholds in a similar scale give a true detection rate of at most 87.1 % and a false rate of at least 60.5 %. Therefore, the proposed threshold values of eq. (6.1) looks widely applicable (site-independent) within a small-scale region (validated with 16-km apart NTUS and SNUS stations).

Discussion

Although, better results are obtained as compared to those in the literature, a slightly higher false alarm is observed for SNUS station using the threshold values of eq. (6.1) from NTUS station. From our analysis, it is found that the PWV values, which is an average over a cone is slightly higher at the SNUS station as it is located much closer to the coastal area.

To illustrate this point, a probability distribution function of PWV values for four SiReNT GPS stations distributed in Singapore are plotted in Fig. 6.5. The plot in magenta represent the PWV values for SiReNT station SSTS, plot in blue represent the SiReNT station SNUS and the plot in green and red, which are overlapping represent the SiReNT stations SRPT and SNYU respectively. Here, it is important to note that the stations SSTS and SNUS are towards the sea whereas the stations

SNYU and SRPT are mostly surrounded by land area. The PWV values for SSTS and SNYU are respectively higher than the PWV values for SNYU and SRPT.

Therefore, when applying the threshold values of eq. (6.1) from NTUS station (with slightly lower PWV values since the NTUS station is mostly surrounded by the land area), false alarm of rainfall nowcasting becomes slightly higher.

Figure 6.5: Probability distribution of PWV values for SiReNT GPS stations SRPT, SNYU, SNUS and SSTS. [Best viewed in color]

6.6.2 Performance Validation Of The Proposed Methodology For A Station From Brazil

In this section, the proposed rainfall nowcasting algorithm is further validated using two-year data (2016 – 2017) at a GPS station from Brazil. This station, SALU, is located at São Luis of Brazil (another tropical region) with very different climatic conditions compared to that of Singapore. Therefore, the thresholds in eq. (6.1) which were derived from NTUS data in Singapore is not applicable here. Thus in the following, we will first present the analysis of the properties of PWV and rain in São Luis, and then derive the new threshold values.

From the data analysis, the weather conditions in São Luis are broadly divided into dry seasons and wet seasons. The dry season lasts for a very short duration

which is during September to November, and the rest of the months generally falls under the wet season. April is the wettest month and October is the driest month.

Fig. 6.6 (a-b) show the distribution of two years (2016 – 2017) of PWV values for both rainy and non-rainy days of wet and dry seasons respectively. The wet season has higher PWV values compared to the dry season. For wet seasons, the mean rainy PWV values are consistently higher, and there is an increment in the mean rainy PWV values of wet season in the evening hours where there is higher probability of rain occurrence as shown by Fig. 6.6 (c). The dry season, which lasts only for 3 months in a year has very less number of rain events. The PWV values are consistently lower for the dry season.

Figure 6.6: Diurnal and seasonal variation in PWV and rainfall data for SALU GPS station (2016 - 2017). The boxplots of PWV values with average rainy (cyan line with circle markers) and non-rainy days (solid light green line) (a) Wet, (b) Dry Seasons. Dashed black lines represent a threshold value of (a) 54 mm for dry and (b) 45 mm for wet season. (c) total number of rain events occurring at specific hours of the day for wet and dry seasons (d) First derivative values of 2 years. (e) Second derivative values of 2 years. The time for both GPS and weather station for SALU station is recorded in UTC format.

Fig. 6.6 (d-e) show the mean first and second derivative values of PWV before and after a rain starts for all the rain events of 2016 and 2017. Both Fig. 6.6 (d) and (e) have similar trends as those observed in Fig. 6.1 (a) and (b) respectively. The minimum value of second derivative occurs within 30 minutes before the start of a rain event. The minimum value is reported to be $-0.01 \text{ mm}/(5\text{min})^2$. After applying the proposed rainfall nowcasting algorithm, a PWV threshold value of 54 mm (ref. to dashed line in Fig. 6.6(a)) and 45 mm (ref. to dashed line in Fig. 6.6(b)) are selected for the wet and dry seasons respectively, and the SD threshold value of $-0.01 \text{ mm}/(5\text{min})^2$ is chosen. It is interesting to note that the PWV threshold value for the rainy season in Brazil is similar to that for the monsoon season in Singapore, and similar observation also holds for SD threshold value. Therefore, with both the PWV value and SD conditions, the rain prediction, P_{SALU} for SALU GPS station becomes true upon satisfying the given criteria as,

$$P_{SALU} = \begin{cases} 1, & SD \leq -0.01 \text{ \& } PWV \geq 54 \text{ (Wet)} \\ & \text{or} \\ & SD \leq -0.01 \text{ \& } PWV \geq 45 \text{ (Dry)} \\ 0, & \text{Otherwise} \end{cases} \quad (6.2)$$

Year	No. of Rain Events (NE + SW + Inter-Monsoon Rain Events)	True Detection (%)	False Alarm (%)
Station from Singapore : SNUS			
2016	91(NE) + 51(SW) + 53(Inter)	90.7	50.2
Station from Brazil: SALU			
2016	232(Wet) + 16(Dry)	82.6	38.5
2017	309(Wet) + 28(Dry)	86.9	35.5
Avg		84.7	37

Table 6.3: Validation Of proposed algorithm for a station from Singapore and Brazil

The eq. (6.2) is applied on PWV and rain data from SALU (2016 and 2017) and the results are tabulated in Table 6.3. From Table 6.3, it can be observed that, the rain events are predicted with a true detection rate of 82.64 % and a false alarm rate

of 38.5 % for 2016 and a true detection rate of 86.9 % and a false alarm rate of 35.5 % for 2017.

For SALU GPS station, the wet season has a total of 541 rain events; 232 in 2016 and 309 in 2017, whereas the dry season has only 44 rain events; 16 in 2016 and 28 in 2017. This is different from Singapore, where all the monsoon seasons and inter-monsoon seasons experience equitable number of rainfall events. Therefore, it is also interesting to access the performance based on different seasons separately.

When the algorithm is applied onto only wet seasons of 2016 and 2017, a true detection rate of 84.4 % and a false alarm rate of 26.8 % were determined. Similarly, when it was applied only onto the dry seasons from 2016 and 2017, a true detection rate of 89.7 % and a false alarm rate of 42.7 % were observed. The proposed algorithm shows better performance for the wet seasons with constantly lower false alarm rate as compared to that of dry season. From rainfall prediction point of view, wet seasons are more important as most of the rain events fall under wet seasons.

The results show that the proposed algorithm works with good accuracy for a tropical station from Brazil as well, and thus it is widely useful and applicable for tropical stations. Here it is important to point out that the PWV threshold values clearly vary from seasons to seasons and the values will change for different locations (large scale, ref. to eq. (6.1) and eq. (6.2). The proposed rainfall nowcasting algorithm can help to derive a priori value of threshold in a location for a particular season, which can be used to select the threshold for other events, which makes the process to forecast easier [4].

6.7 Summary

A detailed study of the PWV values w.r.t rainfall was performed and a simple, yet effective algorithm was proposed for the prediction of the start of a rain event in the tropical region. The proposed prediction algorithm implemented a season dependent PWV threshold value, which was optimized by considering the time-of-day when most of the rain events occur for different seasons and a SD threshold to predict the start of the rain within the next 5 min using data from the past 30 min in a

tropical region. The algorithm was derived based on the data from Singapore, NTUS GPS station. The results showed that the proposed algorithm works well with a true detection of 87.7 % and a false alarm of 38.6 % in average for NTUS. The proposed algorithm was validated using the data from two more tropical stations SNUS and SALU. The proposed algorithm showed good accuracy for both the independent stations as well.

Chapter 7

A Data-driven Approach to Improve the Rainfall Prediction Results

7.1 Introduction

From the previous chapters it is clear that the GPS-PWV values have very good potential in rainfall prediction and with the growing GPS stations, the use of GPS-PWV in rainfall prediction becomes more important. Overall, the true detection rates for rainfall prediction using GPS-PWV values is very good in all regions. The false alarm rates for the tropical region are lesser compared to the false alarm rates reported for the temperate and the sub-tropical regions in the literature. Therefore, in this chapter, along with the GPS-PWV values, we study the usefulness of other meteorological parameters like temperature, relative humidity, solar radiation, dew point temperature, seasonal and diurnal factors in rainfall prediction to lower the false alarm rates maintaining a good true detection rate.

7.1.1 Outline of the contribution

This section outlines the contributions of this chapter:

- A detailed study of different weather parameters is presented using the feature

correlation matrix. Along with the different weather parameters, the seasonal and the diurnal factors are also considered, which are generally neglected in most of the related studies.

- These weather parameters and the seasonal factors are individually assessed for rainfall detection and prediction and the parameters that are important for rainfall prediction with varying lead time values are identified.
- A machine learning algorithm is implemented and a systematic study is done which shows an improvement in the rainfall prediction accuracy.

Database:

GPS PWV and Weather Station Data: Data from GPS and weather stations from Singapore are used. Data from years 2012-2015 from IGS GPS station, NTUS and the co-located weather station from NTU are used. Similarly, data from year 2016 from the SiReNT GPS station, SNUS and the co-located weather station from NUS are used. From the weather station at NTU, the weather parameters namely, temperature ($T, ^\circ\text{C}$), relative humidity ($RH, \%$), dew point temperature ($DPT, ^\circ\text{C}$) and solar radiation ($SR, \text{W/m}^2$) are used. From the weather station at NUS, except DPT all other parameters are available and are used for this Chapter. Throughout the Chapter, the data from NTUS station is used in doing the experiments and discussing the results. The data from SNUS station are used in validating the results. More detailed information on the data is given in Appendix A.

Wide Angle High Resolution Sky Imaging System: Along with the PWV and weather station data, we also use the sky images in this Chapter. Wide Angle High Resolution Sky Imaging System (WAHRISIS) is a system that is developed to take pictures of sky which can be useful in analyzing the sky conditions like: cloudy, partly cloudy, rain cloud, etc and forecasting cloud movements [145]. The WAHRISIS is installed in NTU. It takes the pictures and uploads in the server at an interval of 2 minutes. Fig. 7.1 (a) show WAHRISIS and (b) shows the sky image captured by it.

Figure 7.1: (a) Sky Imager (WAHRIS) (b) Sky image taken by WAHRIS
We use these images in this Chapter for illustration purposes.

7.2 Approach & Tools

In this section, we describe the different tools and techniques that have been implemented to use the various weather features for prediction of a rainfall event.

7.2.1 Supervised Machine Learning Technique

For the purpose of rainfall prediction, the output is always labeled. The label is either 1 or -1 for cases with rain and no-rain respectively. Therefore, supervised machine learning techniques can be implemented for rainfall prediction. Artificial Neural Network (ANN) and Support Vector Machine (SVM) are the most commonly used supervised machine learning techniques for dealing with such geoscience-based problems [123]. We use SVM as it is computationally efficient and works sufficiently for our application.

SVM is a parametric method that generates a hyperplane or a set of hyperplanes in the vector space by maximizing the margin between classifiers to the nearest neighbor data [124, 125]. In this paper, we use SVM to classify rain and no-rain cases using different weather parameters as features. Consider a feature matrix X of dimension $m \times n$, where n is the number of features (weather variables like T , DPT , RH , e.t.c) and m is the number of samples. Consider a output matrix Y ,

a column matrix of dimension $m \times 1$, where values can be either 1 or -1 . 1 and -1 indicate the two classes; rain and no-rain respectively. SVM is trained with a data set of i points represented by $(\vec{x}_1, y_1), (\vec{x}_2, y_2), \dots, (\vec{x}_i, y_i)$, where $\vec{x}_i$ is the i^{th} row of feature matrix X and y_i is the i^{th} row of the output matrix Y . Here, length of i depends on the training size. Based on the training samples, SVM generates a maximum-margin hyperplane that separates the group of points $\vec{x}_i$ for which $y_i=1$ (i.e. rain) from the group of points for which $y_i=-1$ (i.e. no-rain). This is achieved such that the distance between the hyperplane and the nearest point $\vec{x}_i$ from either group is maximized. After the SVM is trained, the remaining samples ($m-i$) are used in testing the model. The test samples are used and the output is predicted. The predicted output is then compared to the real observation data and evaluation metrics are calculated.

7.2.2 Evaluation Metrics

In the study of rainfall prediction, it is important to see how well the rainfall is predicted and how often the methodology makes false predictions. Therefore, the results are generally expressed in terms of true detection and false alarm rates [3, 4]. Likewise, we report the true detection rate (TD), false alarm rate (FA) and the overall accuracy (A) of the algorithm.

Table 7.1 shows the confusion matrix. The matrix shows all the possible cases when the predicted output is compared to the real observation data. Using confusion matrix, TP , TN , FP and FN samples are calculated. The true detection rate is defined as $TD = TP/(TP + FN)$, the false alarm (FA) rate is defined as $FA = FP/(TN + FP)$ and the accuracy is defined as $A = (TP + TN)/(TP + FN + TN + FP)$. We also report the final mis-detection rate (MD) of the algorithm; $MD = 1 - TD$.

7.2.3 Downsampling Technique

In the following, SVM is trained and tested using the weather database of four years (2012 – 2015). The weather data have a temporal resolution of 1 minute,

	Predicted (No)	Predicted (Yes)
Actual (No)	True Negative (TN)	False Positive (FP)
Actual (Yes)	False Negative (FN)	True Positive (TP)

Table 7.1: Confusion Matrix

therefore one year database generally includes a total of $365 * 1440$ data points. Out of these 525,600 data points, there are far fewer data points with rain than without, because rain is relatively a rare event. Therefore, it imposes a problem of imbalanced dataset, which might bias the classification results of the SVM. Therefore, we employ a downsampling technique that balances the number of positive- and negative- labels.

Out of 525120 valid data points from the year 2015, the data points with rain referred to as *minority* cases was only 5017 in number, whereas the the data points without rain referred to as *majority* cases was 520103. The *minority* to *majority* ratio here is nearly 1 : 104, which indicates that the database is highly imbalanced with respect to rain events. Consequently, the traditional way of separating a database viz. 70% of training size and 30% of test size might result in a biased model, which is dominated by the characteristics of the *majority* database. Therefore, we balance the training dataset using a random downsampling technique. In this method, we consider all the cases from the *minority* scenario and the cases from *majority* scenario are randomly chosen such that the *minority* to *majority* ratio is balanced. There is a general practice to make the ratio 1 : 1, but other ratios can also be considered [126]. We implement 1 : 1 random downsampling technique, which is less time consuming and works efficiently for our purpose of rainfall prediction.

7.3 Feature Identification & Correlation

7.3.1 Features

Weather Features

We identified five important weather features; T , RH , SR , DPT and PWV . These weather features are recorded by the weather sensors installed in the weather station.

Weather features T , RH are measured by the temperature/humidity sensor, SR is measured by the solar sensor and PWV is processed from GPS data. DPT is calculated using eq. (7.1). Eq. (7.1) is the simplified approximation to calculate DPT based on RH and T . Here, T is temperature in $^{\circ}C$ and RH in %.

$$DPT = T - \frac{100 - RH}{5} \quad (7.1)$$

These weather features have inherent diurnal and seasonal properties, which can be helpful in rainfall prediction. Therefore, in the next section, we will identify the seasonal and diurnal features for our proposed task.

Seasonal and Diurnal Features

As we discussed in the previous Chapter 6.2.1, in tropical island of Singapore, we experience four main seasons – North-East Monsoon (NE) from Nov-Mar, First-Inter Monsoon (FI) from Apr-May, South-West (SW) Monsoon from Jun-Oct, and Second-Inter Monsoon (SI) from Oct-Nov. In Singapore, the rainfall pattern shows some correlation with the seasons. We often experience late afternoon showers during the NE Monsoon. Sumatra squalls are experienced during pre-dawn to midday, and short-lived rain falls often take place in the afternoon during the SW Monsoon. During inter-monsoon seasons, afternoon to early evening rain events are common [142]. This was shown in results of Chapter 6. Therefore, we consider day-of-year (DoY) as a feature that takes the seasonal effect into consideration.

The diurnal characteristics of all the features can be clearly observed in the time series observation from Fig. 7.4, which will be discussed in detail in Section 7.3.3. We consider hour-of-day (HoD) as a feature that takes the diurnal effect into consideration.

For HoD feature, the hour values repeat after every 24 hours and for the DoY feature the number of days repeat after every 365 days for non-leap years and after 366 days for the leap years. Thus, the DoY and the HoD features are both cyclic in nature, as the same values repeat after a specific period of time. Therefore, each of these features DoY and HoD are expressed into its *sine* and the *cosine* components

Figure 7.2: (a) X and Y components of HoD and (b) X and Y components of DoY [Best viewed in color]

so that their cyclic properties are properly captured. eq. (7.2) and eq. (7.3) are used for expressing the feature HoD into its *sine* and *cosine* components; HoD_x and HoD_y respectively and are plotted as shown in Fig. 7.2(a). The DoY feature can be similarly expressed into its *sine* and *cosine* forms; DoY_x and DoY_y using eq. (7.4) and eq. (7.5) respectively. Fig. 7.2(b) shows the DoY_x and DoY_y values.

$$HoD_x = \cos \frac{2\pi \cdot HoD}{24} \quad (7.2)$$

$$HoD_y = \sin \frac{2\pi \cdot HoD}{24} \quad (7.3)$$

$$DoY_x = \cos \frac{2\pi \cdot DoY}{365} \quad (7.4)$$

$$DoY_y = \sin \frac{2\pi \cdot DoY}{365} \quad (7.5)$$

In the rest of the thesis, if any method includes the HoD feature, both HoD_x and HoD_y are used (same for the DoY feature). Eq. (7.2-7.5), show an example for the non-leap year. For leap years, the total number of days is 366.

7.3.2 Feature Correlation Matrix

It is important to analyze the correlation coefficients between different features, because if two features are perfectly correlated then the second feature will not provide any additional information, as it is already determined by the other.

The degree of correlation amongst all the features is shown in Fig. 7.3. Various observations can be made from the off-diagonal elements. T has a good correlation with features RH , SR , and DPT . A high negative correlation coefficient of -0.9 is observed between T and RH , which is expected because when the air is warm, it can hold more water vapor thus the saturation point increases and the relative humidity values become lower. T and SR , and T and DPT have positive correlation coefficients of 0.7 . During the day time, as the Sun rises, SR and T values both increase, while the RH values decrease. The features RH and SR are negatively correlated with a correlation coefficient of -0.6 . This can be explained as SR values are lowest in the night, whereas the RH values at night in a tropical island like Singapore is generally very high. Therefore, a negative correlation is observed between the two variables. These observations are supported by the time series plot in Fig. 7.4. It is interesting to note that the feature PWV does not show strong correlation with any of the other features. However, it has a small positive correlation with RH . For temperate regions, a higher degree of correlation is observed between RH and PWV [146] as temperate region has a much wider temperature range over different seasons and locations. This has a direct impact on the behavior and correlation of these variables.

Here the *sine* and *cosine* components of the DoY and the HoD features clearly show correlation with features like T , RH , DPT and SR . The use of the cyclic

Figure 7.3: Degree of correlation between various features [Best viewed in color]. properties of HoD and DoY have helped to clearly show the existing correlation between the weather variables and the seasonal and the diurnal factors; this property was underestimated previously when the HoD and the DoY values were directly used instead of their *sine* & *cosine* components.

In summary, different weather features along with the seasonal and the diurnal features were identified. The correlation coefficients between these features were studied. Different features showed correlation between each other which can be explained with reference to a particular situation. However, only RH and T features have a strong correlation coefficient. This indicates that these weather variables can individually contribute towards a particular weather phenomenon. Therefore, in next section all these weather variables are studied with respect to a rainfall event.

7.3.3 Time Series Observations

In this section, a time series plot of different weather features are shown for the GPS station NTUS. We have performed a detailed study to identify the features that are important for rainfall prediction.

In window W1 of Fig. 7.4, rain events can be observed at 14:00 hours and at 18:00 hours. With respect to these rain events, different properties of the weather parameters can be analyzed. For both the rain events, the PWV values significantly increase within 3 hours before the rainfall starts. For the rain event at 14:00, the PWV values start to increase at around 11:00. For the rain event at 18:00, the PWV values start to increase at around 15:00. Such observation was also reported in [5], which statistically shows that the PWV values in the tropical region increases within 2 to 3 hours before a rainfall starts. In section W1, it can be noticed that the temperature drops during the rain as the surface cools down. It decreases and reaches values which are very similar to the dew point values. Whereas, a sharp increment in the RH values can be noticed during the rain. The RH values reach up to 90% and 100% for the rain events at 14:00 and at 18:00 respectively. The solar radiation values on the other hand decrease before and during the rain. The drop in the solar radiation values before the initiation of a rainfall event might be because of the presence of the rain clouds. The WAHRSIS images taken during and before the rain events at 14:05 and at 17:30 respectively show the presence of thick dark clouds. The presence of the clouds blocks the solar radiation and hence the values drop significantly.

Similar observations can be made from window W2, where a significant increment in PWV values can be noticed before the start of the rain event. The PWV values start to increase at around 12:00 for a rain event that occurs at 15:30. The temperature values drop to the dew point values and the RH values increase up to 100% during the rain event. Similarly, the solar radiation values decrease before and during the rain event. The WAHRSIS image taken at 15:15 shows the presence of clouds. Here, a clear sky image taken at 12:00 is also shown for reference (black dot in the image marks the position of the Sun). In Singapore for a clear sky day, the solar radiation values can reach up to 1000 W/m^2 [147] and the values can fall very low before and during a rain event.

It is important to note that the weather parameters like T , RH , DPT and SR show a very distinct diurnal variations. These variables show fluctuations in the day time which could be correlated to rainfall but in the night time they generally

Figure 7.4: Time series of different weather variables. (a) PWV (blue solid line) in mm and Rainfall Rate (stem plot in green) in mm/hr (b) Surface Temperature in $^{\circ}\text{C}$ (solid magenta line), Dew point Temperature in $^{\circ}\text{C}$ (dashed cyan line) and Relative humidity in % (Red dots) (c) Solar Radiation in watt/ m^2 . The X-axis for all the subplots represent the time of the day for four consecutive days : 28th, 29th, 30th Nov and 1st Dec, 2015. The sky images captured by WAHRSIS are also shown for specific time of the day. [Best viewed in color]

follow a similar pattern with minimum variations. As can be seen from Fig. 7.4(b), relative humidity is always high, nearly 100% during the night and the early morning hours. Similarly, the temperature and the dew point readings are almost the same during the night and the early morning hours. Naturally, the solar radiation values are always zero for the night hours and early morning hours. PWV values show a distinct diurnal pattern [2, 143], but unlike T , DPT , RH and SR values, PWV values show fluctuations during the night time in response to rain events.

Section W3 shows a midnight rain event. A distinct increment in the PWV values can be observed 3 hours before the start of the rain event at 00:30. This observation

is correlated with the observations made for the day time rain events from sections W1 and W2. As expected, the T , RH , DPT and the SR values do not show any significant changes corresponding to this midnight rain event as shown in W3.

In summary, we analyzed the different weather parameters which are important for rainfall prediction. We observed that T , RH and DPT show sudden changes during the rain but not before, whereas PWV and SR values show relatively distinct changes before a rain event in day time, and only PWV values show distinct changes before a night time rain event.

7.4 Rainfall Prediction

In Chapter 6, we used the PWV values and its second derivative to develop a model for tropical region to predict a rain event after 5 minutes given a data of 30 minutes duration. The model was sub-divided into 3 sections based on the seasons (North East, South West and Inter monsoons). For this paper, we use the same rain prediction scenario where by, 1) we divide our whole feature database into segments consisting of 30 minutes of data, 2) for each 30 min segment, it is checked whether a rain event occurs after a lead time of 5 minutes or not and 3) all rainfall cases which are separated by a duration of 30 minutes or less are considered as a single rain event [127, 131]. Different than [5], here we study the combined effect of using different meteorological parameters along with PWV values in rainfall prediction and instead of using separate seasonal models, we use the seasonal and the diurnal features to propose one combined model.

The methodologies described in Section 7.2 are implemented to develop this rainfall prediction model. The evaluation metrics are reported after the model is trained and tested using data from years 2012-2015 for NTUS station and data from year 2016 for SNUS station.

Figure 7.5: Performance of an individual feature for (a) Rainfall classification (b) Rainfall Prediction. The TD values are represented by black and FA is represented by blue. D and N indicate results for Day and Night respectively. Each reading shows 95% confidence intervals. The left y axis shows values for TD and the right y axis shows values for FA . (*Best Viewed in Color*)

7.4.1 Assessment of Individual Features for Rainfall Prediction

From the discussion of the time series observation (cf. Fig. 7.4), it was observed that a few weather parameters show changes during the rain, which can be useful for rainfall classification, whereas a few show changes before a rain event which can be useful for rainfall prediction. The effect of the diurnal behavior of the weather features were also highlighted. Thus, in this section, results are presented to show the performance of an individual feature for rainfall classification and prediction. To show the effect of the time-of-day, the results are segregated into day and night time. The 95% confidence intervals are indicated for all the results.

Fig. 7.5(a) shows the rainfall classification results with TD and FA values for day and night in black color and blue color respectively. Similarly Fig. 7.5(b) shows the results for rainfall prediction. From Fig. 7.5(a), it can be observed that all the features can clearly differentiate rain and no-rain conditions in the day time. Most of these features have good performance for rainfall classification in night time as well

except the solar radiation. Solar radiation values are zero in the night time, thus has no effect on the rain classification/prediction. Therefore, we report only day time results for *SR* feature.

From Fig. 7.5(b) it can be observed that unlike in rainfall classification scenario, all the features do not have good true detection rate and false alarm rate for rainfall prediction. Similar to the rainfall classification results, *SR* gives the highest value of *TD* in the day time, whilst it is not useful in the night cases. Whereas features like *T*, *RH* and *DPT* showed a good capability in rainfall classification but their performances are not very useful for rainfall prediction as these parameters changed only during the rain but not before the rain. *PWV* is the only feature, which shows a good separation between *TD* and *FA* for rainfall prediction in both day and night cases. These results are consistent with our time series observation in Fig. 7.4.

Therefore, it is clear that not all the features might be useful for rainfall prediction and the inclusion of the diurnal and the seasonal features together with the weather features will improve the accuracy.

7.4.2 Selection of Optimum Features

Adding Features

In Chapter 6, the rainfall prediction is done based on the *PWV* values and its seasonal behavior only. And in this section, we want to analyze how the evaluation metrics (*TD*, *FA* and *A*) improve by adding in other features. Therefore, we start by taking the combination of *PWV*, *HoD* and *DoY* features as suggested above. Then the other features are added one by one in steps. The first reading in Fig. 7.6 is therefore obtained by using the combination of the *PWV*, the seasonal and the diurnal (*DoY* and *HoD*) features. The subsequent readings are obtained when each feature is added to the previous pool of features as shown by the labels. We do the experiment over a number of iterations (50), with 45% of the total data as the train set and remaining as the test set for each iteration. The reported evaluation metrics are an average over the 50 iterations. Therefore, along with the mean evaluation metrics we also report their 95% confidence intervals.

Figure 7.6: Test results (TD , FA and A) when different features are added one at a time. The TD values are represented by solid black line, A is represented by dashed black line and FA is represented by blue dots. Each reading shows 95% confidence intervals. The left y axis shows values for TD and A and the right y axis shows values for FA . (*Best Viewed in Color*)

The first reading of Fig. 7.6 shows a TD value of 81.9%, FA value of 33.6% and an overall accuracy of 66.6% for rainfall prediction with lead time of 5 minutes. When SR is added in the second step, it can be observed that the FA rate decreases significantly from 33.6% to 22.0%, the TD rate improves and reaches 82.3%. Therefore, the overall accuracy increases to 78.0%. From Fig. 7.5 it was observed that the features T , RH and DPT are not significant in rainfall prediction. Therefore, when these features are subsequently added to the pool of features, the TD values increase but the FA values also increase. When all the features are involved, the TD , FA and A values are 86.8%, 26.1% and 74% respectively. This experiment shows that the highest accuracy is achieved when the feature combination of PWV , SR , DoY and HoD is used. To emphasize on this finding, in the next section the features are systematically eliminated one at a time and the evaluation metrics are evaluated.

Eliminating Features

Fig. 7.7 shows the evaluation metrics when different features are eliminated one at a time from a pool of all the features. In Fig. 7.7, the first reading is obtained by using all the 7 features for the rainfall prediction. The subsequent readings are when the respective feature as shown by the x-axis of the plot are removed from the pool of all the 7 features leaving behind always 6 features. When only RH is removed from the pool, it is labeled as "All-RH". For our analysis, the readings corresponding to the elimination of an individual feature are compared to the first reading when all the features are used.

Figure 7.7: Test results (TD , FA and A) when different features are removed one at a time. The TD values are represented by solid black line, A is represented by dashed black line and FA is represented by blue dots. Each reading shows 95% confidence intervals. The left y axis shows values for TD and A and the right y axis shows values for FA . (*Best Viewed in Color*)

As mentioned earlier, when all the features are used, a TD value of 86.8%, a FA value of 26.1% and an overall accuracy of 74.0% are obtained. When either of the

features, DPT , T or RH is removed from the pool of all the features, TD values decrease slightly and FA and A values remain same as that of the first reading. This indicates that the presence of either of these features do not contribute much for rainfall prediction. Now, when the PWV feature is eliminated, a significant drop in the TD values can be noticed as compared to the first reading. This indicates that the PWV values play an important role in maintaining a high detection rate and high accuracy. When the seasonal and the diurnal features (DoY and HoD) are removed from the pool of all the features, TD remains almost the same whereas, FA increases by almost 8.5 % and the accuracy decreases compared to the first reading. Similarly, when the SR feature is eliminated, the FA values increase from 26.1 % to 38.8 %. These results show that the features SR , DoY and HoD are important in helping us to reduce the false alarm rate and increase the accuracy.

Therefore, the features DPT , RH and T do not contribute in improving the TD rates and removal of these features one at a time actually improves the accuracy by slightly lowering the false alarms. Whereas, HoD , DoY and SR are important to control the false alarms and PWV is important to maintain a high TD rate. This is in line with our previous work [5], where PWV is shown to provide a good rainfall prediction ability. Similarly, in [5], different models were proposed for different seasons for lowering the FA rates. This is in line with the DoY and HoD features and their importance for FA reduction.

In summary, from these experiments it is concluded that the combination of the features PWV , SR , DoY and HoD gives the best results. The TD rate for this combination is 82.3%. The miss detection rate of the algorithm is around 17.7%. The FA rate for this combination is 22% and the overall accuracy for this combination is 78%.

7.4.3 Benchmarking

In this section, the results obtained by using PWV , SR , HoD and DoY features for rainfall prediction are compared to the literature. The true detection rate and the false alarm rate reported in this paper show a significant improvement over the true

detection and false alarm rates of 75% and 60%–70%, respectively, as reported in [3]; the true detection and false alarm rates of 80% and 66%, respectively, as reported in [4] and the true detection and false alarm rates of 85.18% and 66%, respectively, as reported in [1]. A good improvement is observed in the false alarm rates when the proposed method is applied.

Station	Years	Results from Chapter 6 (<i>PWV</i>)		Applied Algorithm (<i>PWV</i> , <i>SR</i> , <i>DoY</i> & <i>HoD</i>)	
		<i>TD</i> (%)	<i>FA</i> (%)	<i>TD</i> (%)	<i>FA</i> (%)
NTUS	2012	88.5	36.4	80.5	20.7
	2013	84.9	43.7	80.9	24.8
	2014	89.1	34.6	83.8	20.3
	2015	89.1	31.0	89.0	21.5
	Avg	87.9	36.4	83.5	21.8
SNUS	2016	90.7	50.2	81.1	19.2

Table 7.2: Benchmarking

Similarly, a detailed comparison is done between the proposed method and the results reported in Chapter 6. Table 7.2 shows the data for comparison. It can be clearly observed that for all the years (2012 - 2015), by incorporating the effect of *SR*, *HoD* and *DoY* with the *PWV* feature the false alarm rate can be significantly lowered. On average, the *FA* rate is reduced by 14.6% whereas the *TD* decreases by only 4.4%. Similarly, the algorithm was implemented using the *PWV* data from the SNUS GPS station and the ground based data from the co-located weather station. When the features *PWV*, *SR*, *DoY* and *HoD* are used for the rainfall prediction, the *FA* rates are reduced by more than a half as compared to the literature. Table 7.2 reports the results for SNUS station, which shows that there is a significant reduction in the false alarms as compared to the true detection rates. Therefore, the proposed method shows a better performance than the literature and a single model is capable of replacing the season specific models.

7.4.4 Different Lead Time Values

The results discussed so far were for a lead time of 5 minutes. In this section, firstly, characteristics of an individual feature in higher lead-time values are shown. Secondly, characteristics of the combination of the optimum features in higher lead-time values are discussed.

Characteristics of An Individual Feature

Here each feature (PWV , T , RH , DPT and SR) is combined with the seasonal and the diurnal features (DoY and HoD) and are used one at a time for rainfall prediction with varying lead time values. Fig. 7.5(b) showed the TD and FA values for a lead-time of 5 minutes. Here we study the characteristics of each feature combined with the seasonal and the diurnal features in higher lead-time values. The results corresponding to the higher lead-time values are shown in Fig. 7.8.

In Fig. 7.8, for all the plots, the black curve with square markers show the TD values and the blue curve with circle marker show the FA rates. The left y-axis of all the plots show the TD readings and the right y-axis show the FA readings.

From these plots, it can be observed that if only single feature is considered at a time, PWV is the only feature which can be useful in rainfall prediction with a higher lead-time value, greater than 20 minutes; maintaining a good detection rate with a low false alarm rate. The SR feature shows good performance till 15 – 20 minutes of lead-time value, after which there is a sharp increment in the FA values which reach upto 50 %. As discussed previously, the plots for T and RH clearly show that the features T and RH are useful in detection but are not useful in prediction. From Fig. 7.8(d), it can be seen that at 0 min lead-time, the TD is nearly 80 % and FA is below 40 %. But as the lead-time increases there is a sharp drop in TD values and a sharp increment in the FA values. The DPT feature shows the least usefulness in the prediction of a rainfall event. The plots shown here re-emphasizes the argument that the features PWV and SR along with the HoD and DoY features are the optimum set of features for short-term rainfall forecasting.

Figure 7.8: TD and FA rates for different features combined with seasonal and diurnal features at higher lead time values (a) PWV , (b) SR , (c) T , (d) RH and (e) DPT .

Figure 7.9: Accuracy at different lead time values

Characteristics of Combination of Optimum Features

This section reports the accuracy of the algorithm for a range of lead-time values when the optimum set of features (*PWV*, *SR*, *DoY* and *HoD*) are used. Fig. 7.9 shows the performance of the algorithm at different lead times, based on data from the previous 30 minutes. As expected, the accuracy decreases with longer lead times. Up to 20 minutes lead time, the accuracy remains above 60%. In tropical location like Singapore, we generally experience short-lived convective rainfall with smaller gaps between two rainfalls [127, 131]; as a result, higher lead times increase the false alarm rates, thus the accuracy of rainfall prediction decreases.

The point of operation depends largely upon the requirements and applications. For higher accuracy requirements, a shorter lead time can be chosen. To obtain better accuracy at longer lead times, the amount of historical data that has been considered can vary.

7.5 Summary

In this chapter, different ground based weather features which are important from the view point of rain were identified. A detailed analysis was done to study the interdependency of these variables. The seasonal and the diurnal factors were incorporated with the weather variables. It was found that all the features play a significant role in the rainfall classification while features like *PWV*, *SR*, *DoY* and *HoD* have potential in rainfall prediction as well. The *PWV* feature contributes the most to maintain a high detection rate, and the features *SR*, *DoY* and *HoD* contribute for lowering the false alarm rates. When compared to the results from the literature, *FA* rates can be significantly reduced by incorporating *SR*, *DoY* and *HoD* features along with the *PWV* feature.

Chapter 8

Conclusion & Future Work

8.1 Conclusion

In this thesis, we have investigated the retrieval of PWV values from GPS in detail and its applications in the area of remote sensing.

The PWV values calculated from the radiosondes, microwave radiometers and satellite images have lower spatio-temporal resolution and these instruments have restriction of being useful under good weather conditions only. GPS derived PWV values have an advantage of having better spatio-temporal resolution and GPS instruments work under all weather conditions.

In this thesis, GIPSY-OASIS v6.4 software was used in deriving the ZWD values. For the calculation of the ZWD values different parameters need to be considered. There are different mapping functions which can be used for mapping the slant delay into the zenith delay. The results presented in the thesis concluded that the different mapping functions do not significantly affect the final PWV values. Similarly, there are different hydrostatic delay functions. It was found that the VMF1 model performs the best for the temperate region and for the sub-tropical region it has the similar performance as that of the GPT and the GPT2 models. The RMSE values improve for the tropical region when the VMF1 models are used but for some stations the performance of the VMF1 model comes closer to the static

model. A comprehensive comparison was done between the GPS-PWV and the radiosonde-PWV values for different stations from the tropical, the sub-tropical and the temperate regions. It was found that the GPS-PWV values have very good correlation with the radiosonde-PWV values for all the regions. The GPS-PWV values were also calculated for different elevation cut-off angles and were compared against radiosonde-PWV values for GPS and radiosonde stations from Singapore. It was concluded that the correlation coefficients/RMS errors decrease/increase as the distance between the radiosonde station and the GPS stations increases. It was found that the effect was more prominent in the PM timings compared to the AM timings in Singapore as in the evening (PM), the climate in Singapore is localized with frequent local convective rain events. The GPS-PWV values were also compared to the MODIS-PWV values. The comparison was done under different cloud mask values. It was concluded that the GPS-PWV and the MODIS-PWV values correlate better during the cloud-free conditions with the RMSE values of 7.8 mm, 5.4 mm and 4.1 mm for the tropical, the sub-tropical and the temperate stations respectively.

In this thesis, a simplified, latitude and day-of-year based PI value model was proposed for the retrieval of PWV values from GPS signals. The proposed equation overcomes the site-specific disadvantages and poor temporal and spatial resolution issues of the traditional techniques of using $(T_m - T_s)$ equations. PI value residuals were studied with respect to station altitude and was concluded that the altitude of a station plays a significant role when the stations are from higher altitude compared to lower altitude. For simplified model it is recommended that the altitude factor should only be used when the altitude of a given station is higher than 1000 m. Thus, by examining the relationship of PI with respect to its geographic coordinates and considering the seasonal effects, a latitude and DoY based PI model was proposed.

A detailed study of the PWV values w.r.t rainfall was performed. Different existing algorithms for rainfall prediction, which use the GPS-PWV values from the temperate and the sub-tropical regions were reviewed and were implemented using data from the tropical region. It was found that different from the existing algorithms, for the tropical region, the threshold of maximum PWV values plays an important role in good detection of rainfall event. The new criteria were applied

on the data from the tropical region and the results were compared with the results from the existing literature. A significant improvement was achieved. The FA rates were decreased by up to 16.9 % with negligible effect on the TD rates. Moreover, the TD rates were studied by varying the length of the data used for prediction and the lead time values. It was found that using 2 hours of historical data rainfall events can be predicted up to 45 min - 1 hour ahead of the time.

A simple algorithm was proposed for the prediction of the start of a rain event in the tropical region. The proposed prediction algorithm implemented a season dependent PWV threshold value, which was optimized by considering the time-of-day when most of the rain events occur for different seasons and a second derivative threshold to predict the start of the rain within the next 5 min using data from the past 30 min in a tropical region. The algorithm was derived based on the data from Singapore, NTUS GPS station. The results showed that the proposed algorithm works well with a true detection of 87.7 % and a false alarm of 38.6 % on average for NTUS station. The proposed algorithm was validated using the data from two more tropical stations SNUS and SALU. The proposed algorithm showed good accuracy for both the independent stations as well.

Finally, a machine learning based algorithm was also applied to improve the rainfall prediction accuracy. For this purpose, different ground based weather parameters, which are important from the view point of rain were identified and were used along with the GPS- PWV values. A detailed analysis was done to study the inter-dependency of these parameters. The seasonal and the diurnal factors were incorporated with the weather parameters. It was found that all the parameters play a significant role in rainfall classification while parameters like PWV , SR , DoY and HoD have potential in rainfall prediction as well. The PWV values contribute the most to maintain a high detection rate, and the parameters like SR , DoY and HoD contribute for lowering the false alarm rates. When compared to the results from the literature, FA rates can be significantly reduced by 12.5 % by incorporating SR , DoY and HoD along with the PWV values.

In a summary, in this thesis, accurate retrieval of PWV values from the GPS signal delays has been studied. We used the GPS- PWV values for rainfall predic-

tion applications. We also worked on data-driven approach to further improve the prediction accuracies.

8.2 Future Work

This thesis explored various aspects of GPS-PWV values and its applications. While this thesis addresses most of the important aspects related to GPS-PWV, it also helped us in generating interesting future research proposals. This section discusses some of the future directions in brief.

- GPS Slant Water Vapor: In this thesis, the zenith GPS-PWV values were used. As a future work, slant PWV values can be used in studying the overall 3D distribution of the water vapor. The study of distribution of the water vapor can be useful in the study of clouds and cloud movements. This can also assist in the prediction of rainfall direction.
- Rainfall prediction algorithm: In this thesis, the proposed rainfall nowcasting algorithm is tested for few locations from the tropical region only. The weather station data is not easily available because of which it is difficult to get the rain data. In future, we will use radar data to indicate rain and no-rain conditions at different IGS GPS station location as radar has higher spatial coverage. Also the algorithm can be applied to the data from the GPS stations from temperate and sub-tropical stations and the evaluation parameters can be analyzed. It will be an interesting work as the rainfall pattern in the sub-tropical and the temperate regions can vary a lot from the tropical region.
- GPS-PWV values for water vapor attenuation: With the advancement in technology, the satellite communication systems are gradually shifting towards higher frequency end (like Q and V bands in the near future). As the frequency becomes higher, atmospheric impairments like water vapor attenuation become prominent. The attenuation caused by water vapor of the atmosphere typically provide a lower contribution to the total attenuation if compared to

clouds and rainfall except for water vapor absorption frequency of 22 GHz. But for some applications which involve lower cut-off angles, the water vapor attenuation can be up to several decibels. In literature, the water vapor attenuation is calculated using the ITU-R models which use the integrated water vapor content calculated from the radiosonde observations [148].

As a future work, the GPS-PWV values can be used instead of the radiosonde-PWV values and the accuracy of the water vapor attenuation can be assessed. Since, radiosonde observations are generally not available for extreme weather conditions and GPS is applicable under all weather conditions, use of the GPS-PWV values can give overall understanding of the water vapor attenuation at a particular frequency.

- Radio Occultation (RO): We know that the GPS-PWV values have lower vertical resolution. Unlike, radiosonde observations, GPS does not give information on the distribution of the water vapor at different heights, it only gives a measure of the water vapor content in a column of the atmosphere. Recently, there is a growing interest in the remote sensing society to use the radio occultation to study the distribution of different parameters w.r.t the altitude [9, 149]. RO uses the radio signals from the Low Earth Orbit (LEO) satellites and the GPS satellites. As a future work, the PWV values can be calculated from the radio occultation technique to improve its vertical resolution.

Author's Publications

Journals

- J1 S. Manandhar, Y. H. Lee, Y. S. Meng and J. T. Ong, "GPS-Derived PWV Values for Long-Term Rainfall Prediction" *Remote Sens.*, Mar 2019. [Under Review].
- J2 S. Manandhar, Y. H. Lee and Y. S. Meng, "Effect of Different Mapping Functions and *ZHD* Models on GPS-Derived Precipitable Water Vapor," *IEEE Trans. Geosci. Remote Sens.*, Nov 2018. [Under Review].
- J3 S. Manandhar, S. Dev, Y. H. Lee, Y. S. Meng and S. winkler, "A Data-Driven Approach to Improve the Rainfall Prediction for Tropical Region," *IEEE Trans. Geosci. Remote Sens.*, 2018. [Major Revision, Decision received on 25th Sep 2018].
- J4 F. Yuan, Y. H. Lee, Y. Meng, S. Manandhar and J. T. Ong, "High Resolution ITU-R Cloud Attenuation Model for Satellite Communications in Tropical Region," *IEEE Trans. Ant. Prop.*, 2018. [Minor Revision, Decision received on 5th Jan 2019].
- J5 S. Manandhar, Y. H. Lee, Y. S. Meng, F. Yuan and J. T. Ong, "GPS Derived PWV for Rainfall Nowcasting in Tropical Region," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4835-4844, Aug, 2018.
- J6 S. Manandhar, Y. H. Lee, Y. S. Meng and J. T. Ong, "A Simplified Model for the Retrieval of Precipitable Water Vapor From GPS Signal," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6245-6253, Nov, 2017.

Conferences

- C1 S. Manandhar, S. Dev, Y. H. Lee and Y. S. Meng, "Analyzing Solar Irradiance Variation From GPS and Cameras," in *Proc. Antennas and Propagation and USNC-URSI Radio Science Meeting (APS/URSI)*, 2018, USA.
- C2 S. Manandhar, S. Dev, Y. H. Lee and Y. S. Meng, "On the Importance of PWV in Detecting Precipitation," in *Proc. Antennas and Propagation and USNC-URSI Radio Science Meeting (APS/URSI)*, 2018, USA.
- C3 S. Manandhar, S. Dev, Y. H. Lee, S. Winkler and Y. S. Meng, "Systematic study of weather variables for rainfall detection," in *Proc. International Geoscience and Remote Sensing Symposium (IGARSS)*, 2018, Spain.
- C4 S. Manandhar, S. Dev, Y. H. Lee, Y. S. Meng and S. Winkler, "A data-driven approach to detecting precipitation from meteorological sensor data," in *Proc. International Geoscience and Remote Sensing Symposium (IGARSS)*, 2018, Spain.
- C5 S. Manandhar, Y. H. Lee, Y. S. Meng, F. Yuan and S. Dev, "A Potential Low Cost remote sensing using GPS derived PWV," in *Proc. International Geoscience and Remote Sensing Symposium (IGARSS)*, 2018, Spain.
- C6 S. Manandhar, Y. H. Lee, Y. S. Meng and K. S. Kumar, "Accuracy Assessment of MODIS Derived Precipitable Water Vapor," in *Proc. Asia-Pacific Conference on Antennas and Propagation (APCAP)*, 2018, New Zealand.
- C7 S. Dev, S. Manandhar, Y. H. Lee and S. Winkler, "Study of Clear Sky Models for Singapore," in *Proc Progress in Electromagnetics Research Symposium (PIERS)*, Singapore, Nov. 2017.
- C8 L. S. Kumar , S. Manandhar, Y. H. Lee and Y. S. Meng, "GPS derived PWV for monitoring Cloud Evolution," in *Proc Progress in Electromagnetics Research Symposium (PIERS)*, Singapore, Nov. 2017.

- C9 S. Dev, S. Manandhar, Y. H. Lee, and Y. S. Meng., "Analyzing Cloud Optical Properties Using Sky Cameras," in *Proc Progress in Electromagnetics Research Symposium (PIERS)*, Singapore, Nov. 2017.
- C10 S. Manandhar, S. Dev, Y. H. Lee, and Y. S. Meng, "Correlating Satellite Cloud Cover with Sky Cameras," in *Proc Progress in Electromagnetics Research Symposium (PIERS)*, Singapore, Nov. 2017.
- C11 S. Manandhar, F. Yuan, Y. H. Lee and Y. S. Meng, "A constant conversion factor for retrieval of PWV from GPS signals in tropical region," in *Proc IEEE APS Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting (APS/URSI)*, San Diego, USA, Jul. 2017.
- C12 S. Dev, S. Manandhar, F. Yuan, Y. H. Lee and S. Winkler, "Cloud radiative effect study using sky camera," in *Proc IEEE AP-S Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting*, San Diego, USA, Jul. 2017
- C13 F. Yuan, S. Manandhar, Y. H. Lee, Y. S. Meng, "Investigation of gaseous attenuation on a Ka-band propagation link in tropical region," in *Proc IEEE AP-S Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting*, San Diego, USA, Jul. 2017
- C14 S. Manandhar, F. Yuan, S. Dev, Y. H. Lee, Y. S. Meng, "Weather Radar to Detect Cloud Occurrence Level," in *Proc IEEE Region 10 Conference (TENCON)*, Singapore, Nov. 2016.
- C15 F. Yuan, S. Manandhar, Y. H. Lee, Y. S. Meng, "Model Comparison for Estimating Cloud Liquid Water Content and Attenuation in Tropical Region," in *IEEE Region 10 Conference (TENCON)*, Singapore, Nov. 2016.
- C16 S. Dev, S. Manandhar, Y. H. Lee, S. Winkler, "Detecting Rainfall Onset Using Sky Images", in *Proc IEEE Region 10 Conference (TENCON)*, Singapore, Nov. 2016.

- C17 S. Manandhar, F. Yuan, Y. H. Lee, Y S Meng, "Weather radar to detect and differentiate clouds from rain events," in *Proc IEEE AP-S International Symposium on Antennas and Propagation and USNC/URSI National Radio Science Meeting (APS/URSI)*, Puerto Rico, USA, Jul. 2016.
- C18 S. Manandhar, Y. H. Lee, and S. Dev, "GPS derived PWV for rainfall monitoring," in *Proc IEEE International Geoscience and Remote Sensing Society (IGARSS)*, Beijing, China, 10 - 15 Jul, 2016.
- C19 S Manandhar, Y. H. Lee and S. Winkler, "Mean Weighted Temperature for PWV Estimation," in *IEEE Asia-Pacific Conference on Antennas and Propagation (AP-CAP)*, Bali, July 2015.

References

- [1] Q. Zhao, Y. Yao, W. Yao, and Z. Li, “Real-time precise point positioning-based zenith tropospheric delay for precipitation forecasting,” *Sci. Rep.*, vol. 8, no. 7939, 2018, doi:10.1038/s41598-018-26299-3.
- [2] S. G. Jin and O. F. Luo, “Variability and climatology of pwv from global 13-year gps observations,” *IEEE Trans. on Geosci. Remote Sens.*, vol. 47, no. 7, pp. 1918–1924, Jul 2009.
- [3] P. Benevides, J. Catalao, and P. M. A. Miranda, “On the inclusion of GPS precipitable water vapour in the nowcasting of rainfall,” *Nat. Hazards Easrth Syst. Sci.*, vol. 15, pp. 2605–2616, 2015.
- [4] Y. Yao, L. Shan, and Q. Zhao, “Establishing a method of short term rainfall forecasting based on gnss-derived pwv and its application,” *Sci. Rep.*, vol. 7, no. 1, pp. 3452–3459, 2017.
- [5] S. Manandhar, Y. H. Lee, Y. S. Meng, F. Yuan, and J. T. Ong, “GPS derived PWV for rainfall nowcasting in tropical region,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4835–4844, Aug 2018.
- [6] P. Jiang, S. Ye, D. Chen, Y. Liu, and P. Xia, “Retrieving precipitable water vapor data using GPS zenith delays and global reanalysis data in china,” *Remote Sensing*, vol. 8, no. 5, pp. 2072–4292, 2016.
- [7] S. G. Jin, J.U. Park, J.H. Cho, and P. H. Park, “Seasonal variability of GPS-derived zenith tropospheric delay (1994-2006) and climate implications,” *J. Geophys. Res.*, vol. 112, no. D9, Jun 2007.
- [8] G. G. Amenu and P. Kumar, “Nvap and reanalysis-2 global precipitable water products: intercomparison and variability studies,” *Bull. Amer. Meteor. Soc.*, vol. 86, no. 2, pp. 245–256, 2005.
- [9] S. P. Ho, L. Peng, C. Mears, and R. A. Anthes, “Comparison of global observations and trends of total precipitable water derived from microwave radiometers and COS-

- MIC radio occultation from 2006 to 2013,” *Atmos. Chem. Phys.*, vol. 18, pp. 259–274, Jan 2018.
- [10] T. Ning, “GPS meteorology: With focus on climate applications,” *PhD Dissertation*, Department of Earth and Space Sciences, Chalmers University of Technology, Gothenburg, Sweden, 2012.
 - [11] X. Wang, K. Zhang, S. Wu, Z. Li, Y. Cheng, L. Li, and H. Yuan, “The correlation between GNSS-derived precipitable water vapor and sea surface temperature and its responses to El niño-southern oscillation,” *Remote Sensing of Environment*, vol. 216, pp. 1–12, 2018.
 - [12] Y. Shoji, K. Sato, M. Yabuki, and T. Tsuda, “Comparison of shipborne GNSS-derived precipitable water vapor with radiosonde in the western north pacific and in the seas adjacent to japan,” *Earth, Planets and Space*, vol. 69, no. 153, 2017.
 - [13] P. Barman, S. Jade, A. Kumar, and W. Jamir, “Inter annual, spatial, seasonal, and diurnal variability of precipitable water vapour over northeast india using GPS time series,” *International Journal of Remote Sensing*, vol. 38, no. 2, pp. 391–411, Dec 2017.
 - [14] K. Ansari, O. Corumluoglu, S. K. Panda, and P. Verma, “Spatiotemporal variability of water vapor over turkey from GNSS observations during 2009-2017 and predictability of ERA-interim and ARMA model,” *The Journal of Global Positioning Systems*, vol. 16, no. 8, 2018.
 - [15] Q. Zhao, Y. Yao, and W. Yao, “GPS-based PWV for precipitation forecasting and its application to a typhoon event,” *J. Atmospheric Sol.-Terr. Phys.*, vol. 167, pp. 124–133, Jan 2018.
 - [16] P. Hu, G. Huang, Q. Zhang, X. Wang, and M. Mao, “Algorithm and performance of precipitable water vapor retrieval using multiple GNSS precise point positioning technology,” in *Proc. China Satellite Navigation Conference (CSNC) 2018*, 2018, vol. 497, pp. 139–151, doi:https://doi.org/10.1007/978-981-13-0005-9_12.
 - [17] Y. Liu and Y. Chen, “Precision of precipitable water vapour from radiosonde data for GPS solutions,” *Geomatica*, vol. 54, no. 2, pp. 171–175, 2000.
 - [18] “Wyoming university, department of atmospheric sciences,” <http://weather.uwyo.edu/upperair/sounding.html>, [Online; latest accessed 17-Mar-2018].
 - [19] I. Durre, R. S. Vose, and D. B. Wuertz, “Overview of the integrated global radiosonde archive,” *J. Climate*, vol. 19, pp. 53–68, 2006.

- [20] Z. Wang, X. Zhou, Y. Liu, D. Zhou, H. Zhang, and W. Sun, "Precipitable water vapor characterization in the coastal regions of china based on ground-based GPS," *Advances in Space Research*, vol. 60, no. 11, pp. 2368–2378, Sep 2017.
- [21] K. Gui, H. Che, Q. Chen, and H. Liu et. al, "Evaluation of radiosonde, MODIS-NIR-clear, and AERONET precipitable water vapor using IGS ground-based GPS measurements over china," *Atmospheric Research*, vol. 197, pp. 461–473, Nov 2017.
- [22] G. Gurbuz and S. G. Jin, "Long-time variations of precipitable water vapour estimated from GPS, MODIS and radiosonde observations in turkey," *Int. J. Climatol.*, vol. 37, pp. 5170–5180, Jun 2017.
- [23] J. V. Martinez, M. Anton, and et.al J. P. O. de Galisteo, "Validation of MODIS integrated water vapor product against reference GPS data at the iberian peninsula," *International Journal of Applied Earth Observation and Geoinformation*, vol. 63, pp. 214–221, Mar 2017.
- [24] R. Ohtani and I. Naito, "Comparisons of GPS-derived precipitable water vapors with radiosonde observations in japan," *J. Geophy. Res.*, vol. 105, no. D22, pp. 26,917–26,929, Nov 2000.
- [25] D. S. Song, H. S. Yun, and D. H. Lee, "Verification of accuracy of precipitable water vapour from GPS during typhoon rusa," *Survey Review*, vol. 40, no. 307, pp. 19–28, 2013.
- [26] V. V. Kalinnikov and O. G. Khutorova, "Diurnal variations in integrated water vapor derived from a GPS ground network in the volumega-ural region of russia," *Ann. Geophys.*, vol. 35, pp. 453–464, Mar 2017.
- [27] Y. Liou, Y. Teng, T. V. Hove, and J. C. Liljegren, "Comparison of precipitable water observations in the near tropics by GPS, microwave radiometer, and radiosondes," *Journal of Applied Meteorology*, vol. 40, no. 1, pp. 5–15, 2001.
- [28] E. Pottiaux and R. Warnant, "First comparisons of precipitable water vapor estimation using GPS and water vapor radiometers at the royal observatory of belgium," *R. GPS Solutions*, vol. 6, pp. 11–17, Nov 2002.
- [29] J. V. Baelen, J. Aubagnac, and A. Dabas, "Comparison of near-real time estimates of integrated water vapor derived with GPS, radiosondes, and microwave radiometer," *J. Atmos. Oceanic Technol.*, vol. 22, pp. 201–210, Nov 2005.
- [30] K. Saito, M. Kunii, and K. Araki, "Cloud resolving simulation of a local heavy rainfall event on 26 august 2011 observed in tomacs," *J. Met. Soc. of Japan*, vol. 96A, pp. 175–199, Jan 2018.

- [31] T. Yeh, H. Shih, C. Wang, S. Choy, C. Chen, and J. Hong, "Determining the precipitable water vapor thresholds under different rainfall strengths in taiwan," *Advances in Space Research*, vol. 61, no. 3, pp. 941–950, 2018.
- [32] "MODIS data archive website," <https://ladsweb.modaps.eosdis.nasa.gov/archive/allData/6/>, [Online; latest accessed 30-June-2018].
- [33] M. D. King, Y. J. Kaufman, W. P. Menzel, and D. Tanre, "Remote sensing of cloud, aerosol, and water vapor properties from the moderate resolution imaging spectrometer (MODIS)," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 1, pp. 2–27, Jan 1992.
- [34] J. Chang, S. G. Jin, and X. He, "Assessment of inSAR atmospheric correction using both MODIS near-infrared and infrared water vapor products," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5726–5735, Sep 2014.
- [35] B. C. Gao and Y. J. Kaufman, "Water vapor retrievals using moderate resolution imaging (MODIS) near-infrared channels," *J. Geophys. Res.*, vol. 108, no. D13, pp. 4389–4398, Jul 2003.
- [36] S. W. Seemann, J. Li, W. P. Menzel, and L. E. Gumley, "Operational retrieval of atmospheric temperature, moisture, and ozone from MODIS infrared radiances," *J. Appl. Meteorol.*, vol. 42, no. 8, pp. 1072–1091, Aug 2003.
- [37] S. S. Ningombam, S. Jade, T.S. Shrungeshwara, and H.-J. Song, "Validation of water vapor retrieval from moderate resolution imaging spectro-radiometer (MODIS) in near infrared channels using GPS data over IAO-hanle, in the trans-himalayan region," *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 137, pp. 76–85, Jan 2016.
- [38] H. Liu, S. Tang, S. Zhang, and J. Hu, "Evaluation of MODIS water vapour products over china using radiosonde data," *International Journal of Remote Sensing*, vol. 36, no. 2, pp. 680–690, 2015.
- [39] N. Lu, J. Qin, K. Yang, Y. Gao, X. Xu, and T. Koike, "On the use of GPS measurements for moderate resolution imaging spectrometer precipitable water vapor evaluation over southern tibet," *J. Geophys. Res.*, vol. 116, no. D23117, pp. 680–690, 2011.
- [40] B. Bramanto, I. Gumilar, T. P. Sidiq, W. Kuntjoro, and D. A. Tampubolon, "Sensing of the atmospheric variation using low cost GNSS receiver," in *Proc. IOP Conference Series: Earth and Environmental Science*, 2018, vol. 149, p. 012073, doi:<http://stacks.iop.org/1755-1315/149/i=1/a=012073>.

- [41] W. Suparta, R. Rahman, M. Singh, J. Singh, and M. T. Latif, "Investigation of flash flood over the west peninsular malaysia by global positing system network," *Advanced Science Letters*, vol. 21, no. 2, pp. 153–157, 2015.
- [42] S. Choy, C. S. Wang, T. K. Yeh, J. Dawson, M. Jia, and Y. Kuleshov, "Precipitable water vapor estimates in the australian region from ground-based GPS observations," *Advances in Meteorology*, vol. 2015, no. 956481, 2015, <https://doi.org/10.1155/2015/956481>.
- [43] "Igs gps station network," <http://www.igs.org/network>, [Online; latest accessed 30-June-2018].
- [44] "Crustal dynamics data information system: NASA's archive of space geodesy data," <ftp://cddis.gsfc.nasa.gov/pub/gps/data/>, [Online; latest accessed 30-June-2018].
- [45] C. Lu, X. Li, and T. Nilsson, "Real-time retrieval of precipitable water vapor from GPS and BeiDou observations," *J. Geod.*, vol. 89, no. 9, pp. 843–856, Sep 2015.
- [46] X. Li, X. Zhang, X. Ren, M. Fritsche, J. Wickert, and H. Schuh, "Precise positioning with current multi-constellation global navigation satellite systems: GPS, GLONASS, galileo and BeiDou," *Sci. Rep.*, vol. 5, no. 8328, Feb 2015.
- [47] K. Zhang, T. Manning, S. Wu, W. Rohm, D. Silcock, and S. Choy, "Capturing the signature of severe weather events in australia using GPS measurements," *IEEE Trans. Geosci. Remote Sens.*, vol. 8, no. 4, pp. 1839–1847, Apr 2015.
- [48] J. Shi, C. Xu, J. Guo, and Y. Gao, "Real-time GPS precise point positioning-based precipitable water vapor estimation for rainfall monitoring and forecasting," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3452–3459, June 2015.
- [49] C. J. Hegarty and E. Chatre, "Evolution of the global navigation satellite system (gnss)," *Proceedings of the IEEE*, vol. 96, no. 12, pp. 1902–1917, Dec 2008, doi:10.1109/JPROC.2008.2006090.
- [50] M. Bevis, S. Businger, and S. Chiswell, "GPS meteorology: Mapping zenith wet delays onto precipitable water," *J. Appl. Meteor.*, vol. 33, no. 3, pp. 379–386, Mar 1994.
- [51] N. Jiang and et. al T. Xu, "GPS/BDS short-term isb modelling and prediction," *GPS Solut.*, vol. 21, no. 1, pp. 163–175, 2017.
- [52] Z. Liu, Y. Li, F. Li, and J. Guo, "Estimation and evaluation of the precipitable water vapor from GNSS PPP in asia region.," in *Proc. China Satellite Navigation Conference (CSNC)*, May 2017, vol. I, pp. 85–95.

- [53] C. Kitpracha, D. Promchot, P. Srestasathiern, and C. Satirapod, "Precise tropospheric delay map of thailand using GNSS precise point positioning technique.," *International Journal of Geoinformatics*, vol. 13, no. 2, pp. 17–21, 2017.
- [54] S. A. E. Younes, "Study the effect of new egypt wet mapping function on space geodetic measurements," *American Journal of Remote Sensing*, vol. 6, no. 1, pp. 29–38, Mar 2018.
- [55] T. A. Herring G. Elgered, J. L. Davis and I. I Shapiro, "Geodesy by radio interferometry: water vapor radiometry for estimation of the wet delay," *J. Geophys. Res.*, vol. 96, no. B4, pp. 6541–6555, Apr 1991.
- [56] F. Zhou, G. Shengfeng, W. Chen, and D. Dong, "Comprehensive assessment of positioning and zenith delay retrieval using GPS+glonass precise point positioning," *Acta Geodynamica et Geomaterialia*, vol. 14, pp. 317–326, Jun 2017.
- [57] J. M. Astudillo, L. Lau, and T. Moore, "Analysing the zenith tropospheric delay estimates in on-line precise point positioning (PPP) services and PPP software packages," *Sensors*, vol. 18, no. 2, Feb 2018.
- [58] P. Tregoning and T. A. Herrin, "Impact of a priori zenith hydrostatic delay errors on GPS estimates of station heights and zenith total delays," *Geo. Phy. Let.*, vol. 33, no. L23303, 2006, doi:10.1029/2006GL027706.
- [59] Y. Shoji, H. Nakamura, and T. Iwabuchi et. al, "Tsukuba gps dense net campaign observation: Improvement in GPS analysis of slant path delay by stacking one-way postfit phase residuals," *Journal of the Meteorological Society of Japan. Ser. II*, vol. 82, no. 1B, pp. 301–314, 2004.
- [60] D. Landskron and J. Bohm, "VMF3/GPT3: refined discrete and empirical troposphere mapping functions," *J Geod*, vol. 92, pp. 349–360, 2018.
- [61] Z. Liu, Y. Dang, Q. Yang Q, and H. Wang, "Applicability analysis of troposphere mapping functions in china," in *Proc.China Satellite Navigation Conference (CSNC) 2017*, May 2017, doi:https://doi.org/10.1007/978-981-10-4594-3_27.
- [62] K. Wilgan, T. Hadas, P. Hordyniec, and J. Bosy, "Real-time precise point positioning augmented with high-resolution numerical weather prediction model," *GPS Solut.*, vol. 21, pp. 1341–1353, Mar 2017.
- [63] A. Klos, A. Hunegnaw, F. N. Teferle, K. B. Abraha, F. Ahmed, and J. Bogusz, "Statistical significance of trends in zenith wet delay from re-processed GPS solutions," *GPS Solut.*, vol. 22, no. 2, Mar 2018.

- [64] L. Urquhart, F. G. Nievinski, and M. C. Santos, "Assessment of troposphere mapping functions using three-dimensional ray-tracing," *GPS Solutions*, vol. 18, no. 3, pp. 345–354, Jul 2014.
- [65] J. W. Marini, "Correction of satellite tracking data for an arbitrary tropospheric profile," *Radio Sci.*, vol. 7, no. 2, pp. 223–231, 1972.
- [66] A. E. Niell, "Improved atmospheric mapping functions for VLBI and GPS," *Earth Planets Space*, vol. 52, no. 10, pp. 699–702, 2000.
- [67] A. E. Niell, "Global mapping functions for the atmosphere delay at radio wavelengths," *J. Geophys. Res.*, vol. 101, no. b2, pp. 3227–3246, 1996.
- [68] "GGOS website," <http://ggosatm.hg.tuwien.ac.at/DELAY/>, [Online; latest accessed 30-June-2018].
- [69] M. A. Abdelfatah, A. E. Mousa, and G. S. El-Fiky, "Assessment of tropospheric delay mapping function models in Egypt: Using PTD database model," *J. Ast. Geophys.*, vol. 7, pp. 47–51, 2018.
- [70] J. Boehm, A. Niell, P. Tregoning, and H. Schuh, "Global mapping function (GMF): A new empirical mapping function based on numerical weather model data," *Geophys. Res. Lett.*, vol. 33, no. L07304, 2006.
- [71] J. Böhm, G. Möller, M. Schindelegger, G. Pain, and R. Weber, "Development of an improved empirical model for slant delays in the troposphere (GPT2w)," *GPS Solut.*, vol. 19, pp. 433–441, 2015.
- [72] M. Bevis, S. Businger, T. A. Herring, C. Rocken, R. A. Anthes, and R. H. Ware, "GPS meteorology: remote sensing of atmospheric water vapor using the global positioning system," *J. Geo. Res.*, vol. 97, no. D14, pp. 15,787–15,801, Oct 1992.
- [73] A. Saeed, A. Madjid, A. Jamal, and G. Abdolreza, "Precipitable water vapour estimation using the permanent single GPS station in zanzan, iran," *Meteorological Applications*, vol. 24, no. 3, pp. 415–422, Jul 2017.
- [74] X. Wang, k. Zhang, S. Wu, C. He, Y. Chen, and X. Li, "Determination of zenith hydrostatic delay and its impact on GNSS-derived integrated water vapor," *Atmospheric Measurement Techniques*, vol. 10, no. 8, pp. 2807–2820, Aug 2017.
- [75] J. Liu, X. Chen, J. Sun, and Q. Liu, "An analysis of GPT2/GPT2w+saastamoinen models for estimating zenith tropospheric delay over asian area," *Advances in Space Research*, vol. 59, no. 3, pp. 824–832, 2017.

- [76] J. Z. Kalita and Z. Rzepecka, “Impact of the initial tropospheric zenith path delay on precise point positioning convergence during active conditions,” *Measurement Science and Technology*, vol. 28, no. 4, Feb 2017.
- [77] A. Leontiev and Y. Reuveni, “Combining meteosat-10 satellite image data with GPS tropospheric path delays to estimate regional integrated water vapor (IWV) distribution,” *Atmospheric Measurement Techniques*, vol. 10, no. 2, pp. 537–548, 2017.
- [78] S. Desai, D. Kuang, and W. Bertiger, “Gipsy/oasis (gipsy) overview and under the hood,” Tech. Rep., Near Earth Tracking Systems and Applications Groups, Jet Propulsion Laboratory, California Institute of Technology, Mar 2014, ftp://ehzftp.wr.usgs.gov/svarc/GIPSY_pdfs/GIPSY_Overview.pdf.
- [79] T. Hadas, F. N. Teferle, K. Kazmierski, P. Hordyniec, and J. Bosy, “Optimum stochastic modeling for GNSS tropospheric delay estimation in real-time,” *GPS Solut.*, vol. 21, pp. 1069–1081, Dec 2017.
- [80] Y. Yuan, K. Zhang, W. Rohm, S. Choy, R. Norman, and C. S. Wang, “Real-time retrieval of precipitable water vapor from gps precise point positioning,” *J. Geophys. Res. Atmos.*, vol. 119, pp. 10,044–10,057, 2014.
- [81] F. Ahmed, P. Vaclavovic, F. N. Teferle, J. Dous, R. Bingley, and D. Laurichesse, “Comparative analysis of real-time precise point positioning zenith total delay estimates,” *GPS Solut.*, vol. 20, pp. 187–199, 2016.
- [82] P. S. de Oliveira Jr., L. Morel, F. Fund, R. Legros, J. F. G. Monico, S. Durand, and F. Durand, “Modeling tropospheric wet delays with dense and sparse network configurations for PPP-RT,” *GPS Solut.*, vol. 21, pp. 237–250, 2017.
- [83] P. Golaszewski, K. Stepniak, and P. Wielgosz, “Zenith tropospheric delay estimates using absolute and relative approaches to GNSS data processing - preliminary results,” in *Proc.2017 Baltic Geodetic Congress (BGC Geomatics)*, Aug 2017, pp. 414–418.
- [84] D. K. Adams, H. M. J. Barbosa, and K. P. G. De Los Rios, “A spatiotemporal water vapor-deep convection correlation metric derived from the amazon dense GNSS meteorological network,” *Monthly Weather Review*, vol. 145, no. 1, pp. 279–288, 2017.
- [85] C. Yu, N. T. Penna, and Z. Li, “Generation of real-time mode high-resolution water vapor fields from GPS observations,” *J. Geophys. Res. Atmos.*, vol. 122, pp. 2008–2025, 2017, doi:10.1002/2016JD025753.
- [86] R. Pacione et al, “EPN-repro2: A reference GNSS tropospheric data set over europe,” *Atmospheric Measurement Techniques*, vol. 10, no. 5, pp. 1689–1705, 2017, <http://dx.doi.org/10.5194/amt-10-1689-2017>.

- [87] F. Alshawaff, T. Fuhrmann, A. Knopfler, X. Luo, M. Mayer, S. Hinz, and B. Heck, "Accurate estimation of atmospheric water vapor using GNSS observations and surface meteorological data," *IEEE Trans. on Geosci. Remote Sens.*, vol. 53, no. 7, Jul 2015.
- [88] S. Amir and T.A.Musa, "GPS meteorology activities in the malaysian peninsula," in *Proc.10th International Symposium And Exhibition on Geoinformation (ISG2011) And ISPRS Commission*, Sept 2011.
- [89] H. Seko, H. Nakamura, Y. Shoji, and T. Iwabuchi, "The meso scale water vapor distribution associated with a thunderstorm calculated from a dense network of GPS receivers," *J. Mete. Soc.*, vol. 82, no. 1B, pp. 569–586, 2004.
- [90] Y. Shoji et. al, "Semi-diurnal and diurnal variation of errors in GPS precipitable water vapor at tsukuba, japan caused by site displacement due to ocean tidal loading," *J. Mete. Soc.*, vol. 52, pp. 685–690, 2000.
- [91] X. Li, L. Xu, Y.Fang, Y. Zhang, J. Ding, H. Liu, and X. Deng, "Estimation of the precipitable water vapor from ground-based GPS with GAMIT/GLOBK," in *Proc. IITA Int. Conf. Geosci. Remote Sens*, Aug 2010, number 3, pp. 210–214.
- [92] W. Suparta, M. Alauddin, M. Ali, B. Yatim, and N. Misran, "Remote sensing of antarctic atmospheric water vapour using ground-based GPS meteorology," in *Proc. 5th Student Conf. Research and Development (SCOReD)*, Dec 2007, pp. 1–8.
- [93] L. S. Kumar, F. Yuan, and Y. H. Lee, "Comparison of precipitable water vapor derived from GPS and radiosonde data for singapore," in *Proc.2014 Int. Symp Antennas Propag*, Dec 2014, pp. 437–438.
- [94] J. Wang, L. Zhang, and A. Dai, "Global estimates of water-vapor-weighted mean temperature of the atmosphere for GPS applications," *J. Geophys. Res*, vol. 110, no. D21, pp. 379–386, Nov 2005.
- [95] A. Susnik, "GPS for weather and spaceweather studies," *Ph. D thesis*, University of Nova Gorica 2013.
- [96] Y. Cao, F. Zheng, Y. Xie, and Y. Bi, "Impact of the weighted mean temperature on the estimation of GPS precipitable water vapor," in *Proc.2008 Int. Conf. Microw. Millimeter Wave Technol.*, Apr 2008, pp. 799–801.
- [97] L. Li, S. Q. Wu, Z. M. Wang, Y. Tian, C. Y. He, and K. F. Zhang, "Modelling of weighted-mean temperature using regional radiosonde observations in hunan china," *Terr. Atmos. Ocean. Sci., volume.*, vol. 29, no. 2, pp. 187–199, Apr 2018.

- [98] “GGOS website for weighted mean temperature,” <http://ggosatm.hg.tuwien.ac.at/DELAY/ETC/TMEAN/>, [Online; latest accessed 30-June-2018].
- [99] L. Li, X. Wang, Y. Tian, C. He, and K. Zhang, “Modelling of weighted-mean temperature using regional radiosonde observations in Hunan, China,” *Terr. Atmos. Ocean. Sci.*, vol. 29, no. 2, pp. 187–199, 2018.
- [100] Y. Yao, S. Zgu, and S. Yue, “A globally applicable, season-specific model for estimating the weighted mean temperature of the atmosphere,” *J. Geod.*, vol. 86, pp. 1125–1135, Apr 2012.
- [101] Y. Yao, B. Zhang, S. Yue, C. Xu, and W. Peng, “Global empirical model for mapping zenith wet delays onto precipitable water,” *J. Geod.*, vol. 87, no. 5, pp. 439–448, May 2013.
- [102] Y. Yao, C. Xu, B. Zhang, and N. Cao, “GTm-III: a new global empirical model for mapping zenith wet delays onto precipitable water vapour,” *Geophysical Journal International*, vol. 197, no. 1, pp. 202–212, 2014.
- [103] P. Chen, W. Yao, and X. Zhu, “Realization of global empirical model for mapping zenith wet delays onto precipitable water using NCEP re-analysis data,” *Geophysical Journal International*, vol. 198, no. 3, pp. 1748–1757, 2014.
- [104] G. Petit and B. Luzum, “IERS conventions 2010,” Tech. Rep. 36, IERS Technical Note, <http://www.iers.org/IERS/EN/Publications/TechnicalNotes/tn36.html>.
- [105] F. Lyard, F. Lefevre, T. Letellier, and O. Francis, “Modelling the global ocean tides: modern insights from FES2004,” *Ocean Dynamics*, vol. 56, pp. 394–415, 2006, doi:10.1007/s10236-006-0086-x.
- [106] “Nasa ftp server, total tropospheric delay,” <ftp://cddis.gsfc.nasa.gov/gps/products/troposphere/new/>, [Online; latest accessed 30-June-2018].
- [107] D. S. McMillan, “Atmospheric gradients from very long baseline interferometry observations,” *Geophy. Res. Lett.*, vol. 22, no. 9, pp. 1041–1044, May 1995.
- [108] M. Fritsche, R. Dietrich, C. Knofel, and A. Rulke et. al, “Impact of higher-order ionospheric terms on gps estimates,” *Geophysical Research Letters*, vol. 32, no. 23, 2005.
- [109] T. Manning, K. Zhang, W. Rohm, S. Choy, and F. Hurter, “Detecting severe weather using GPS tomography: An Australian case study,” *J. of Glob. Pos. Sys.*, vol. 11, no. 1, pp. 58–70, 2012.

- [110] W. Suparta, S. K. Zulkeple, and W. S. Putro, “Estimation of thunderstorm activity in tawau, sabah using GPS data,” *Advanced Science Letters*, vol. 23, no. 2, pp. 1370–1373, Feb 2017.
- [111] M.J. Singh W. Suparta M.T. Islam F.Tangang M. Muhammad, M. Abdullah, “Characterization of GPS PWV during flooding event over keningau, sabah,” in *Proc.IconSpace*, Jul 2013, pp. 429–433, Malaysia.
- [112] K. Zhang S. Choy, C. Wang and Y. Kuleshov, “GPS sensing of precipitable water vapour during the march 2010 melbourne storm,” *Adv. Space Res.*, vol. 52, no. 9, pp. 1688–1699, Nov 2013.
- [113] W. Rohm, Y. Yuan, B. Biadeglne, K. Zhang, and J. L. Marshall, “Ground-based GNSS ZTD/IWV estimation system for numerical weather prediction in challenging weather conditions,” *Atm. Res.*, vol. 138, no. 0169-8095, pp. 414–426, 2014.
- [114] L. S. Kumar, S. Manandhar, Y. H. Lee, and Y. S. Meng, “GPS derived pwv for monitoring cloud evolution,” in *Proc.2017 Progress in Electromagnetics Research Symposium - Fall (PIERS - FALL)*, Nov 2017, pp. 1421–1423, doi:10.1109/PIERS-FALL.2017.8293353.
- [115] G. Vali P. Wechsler et al. Z. Wang, J. French, “Single aircraft integration of remote sensing and in situ sampling for the study of cloud microphysics and dynamics,” *Bull. Amer. Meteor. Soc.*, vol. 93, no. 5, pp. 653–668, 2012.
- [116] H. Takiguchi, T. Kato, H. Kobayashi, and T. Nakaegawa, “GPS observations in thailand for hydrological applications,” *Earth Planets Space*, vol. 52, no. 2, pp. 913–919, 2000.
- [117] S. Barindelli, E. Realini, G. Venuti, A. Fermi, and A. Gatti, “Detection of water vapor time variations associated with heavy rain in northern italy by geodetic and low-cost GNSS receivers,” *Earth, Planets and Space*, vol. 70, no. 1, Feb 2018.
- [118] P. W. Li, X. Y. Wang, Y. Q. Chen, and S. T. Lai, “Use of GPS signal delay for real-time atmospheric water vapour estimation and rainfall nowcast in hongkong,” in *Proc.First Int. Symp Cloud-prone & Rainy Area Rem. Sens.*, Oct 2005, HongKong.
- [119] M. A. Sharifi, A. S. Khaniani, and M. Joghataei, “Comparison of GPS precipitable water vapor and meteorological parameters during rainfall in tehran,” *Meteorol. Atmos. Phys.*, vol. 127, pp. 701–710, 2015.
- [120] K.C. Luk, J.E. Ball, and A. Sharma, “A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting,” *J. of Hydrology*, vol. 227, pp. 56–65, 2000.

- [121] G. J. Sawale and S. R. Gupta, "Use of artificial neural network in data mining for weather forecasting," *International J. Of Computer Sci. And App.*, vol. 6, no. 2, Apr 2013.
- [122] E. Hernandez, V. Sanchez-Anguix, V. Julian, J. Palanca, and N. Duque, "Rainfall prediction: A deep learning approach," in *Proc. International Conference on Hybrid Artificial Intelligence Systems*, Apr 2016.
- [123] D. J. Lary, A. H. Alavi, A. H. Gandomi, and A. L. Walker, "Machine learning in geosciences and remote sensing," *Geoscience Frontiers*, vol. 7, 2016.
- [124] S. Dev, B. Wen, Y. H. Lee, and S. Winkler, "Machine learning techniques and applications for ground-based image analysis," *IEEE Geoscience and Remote Sensing Magazine, Special Issue on Advances in Machine Learning for Remote Sensing and Geosciences*, vol. 4, no. 2, pp. 79–93, Jun 2016.
- [125] N. Cristianini and J. Shawe-Taylor, "An introduction to support vector machines and other kernel-based learning methods," *Cambridge University Press*, 2000.
- [126] M. M. Rahman and D. N. Davis, "Cluster based under-sampling for unbalanced cardiovascular data," in *Proc. World Congress on Engineering*, 2013.
- [127] J. X. Yeo, Y. H. Lee, and J. T. Ong, "Performance of site diversity investigated through radar derived results," *IEEE Trans. Antennas Prop.*, vol. 59, no. 10, pp. 3890 – 3898, 2000.
- [128] J. Braun, C. Rocken, and R. Ware, "Validation of line-of-sight water vapor measurements with GPS," *Radio Science*, vol. 36, no. 3, pp. 459–472, 2001.
- [129] J. Braun, C. Rocken, and J. Liljegren, "Comparisons of line-of-sight water vapor observations using the global positioning system and a pointing microwave radiometer," *Journal of Atmospheric and Oceanic Technology*, vol. 20, no. 5, pp. 606–612, 2003.
- [130] T. Ning and G. Elgered, "Trends in the atmospheric water vapor content from ground-based GPS: The impact of the elevation cutoff angle," *IEEE J. Sel. Topics in App. Earth Obser. and Remote Sens.*, vol. 5, no. 3, pp. 744–751, 2012.
- [131] J. X. Yeo, Y. H. Lee, and J. T. Ong, "Use of weather radar rain cell motion forecasting for site diversity system," in *Proc. MILCOM*, Nov 2016.
- [132] "Wyoming university, department of atmospheric sciences," <http://weather.uwyo.edu/surface/meteorogram/>, [Online; latest accessed 17-Mar-2018].
- [133] S. Manandhar, Y.H. Lee, and S. Winkler, "Mean weighted temperature for PWV estimation," in *Proc. 2015 IEEE 4th Asia-Pacific Conf. Antennas Propag*, Jul 2015, pp. 437–438.

- [134] Z. Chen, J. Li, L. Liu, G. Luo, and X. Mo, "Model establishment and analysis of weighted mean temperature in the region of guilin," in *Proc.China Satellite Navigation Conference (CSNC) 2017 Proceedings*, 2017, vol. 437, doi:https://doi-org.ezlibproxy1.ntu.edu.sg/10.1007/978-981-10-4588-2_1.
- [135] S. Manandhar, F. Yuan, Y. H. Lee, and Y. S. Meng, "A constant conversion factor for retrieval of PWV from GPS signals in tropical region," in *Proc. IEEE APS Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting (APS/URSI)*, Jul 2017.
- [136] X. Wang, K. Zhang, S. wu, S. Fan, and Y. Chen, "Water vapor weighted mean temperature and its impact on the determination of precipitable water vapor and its linear trend," *J. Geophy. Res.*, vol. 121, no. 2, pp. 833–852, Jan 2016.
- [137] F. Alshawaff, T. Fuhrmann, A. Knopfler, M. Mayer X. Luo, S. Hinz, and B. Heck, "Accurate estimation of atmospheric water vapor using GNSS observations and surface meteorological data," *IEEE Trans. on Geosci. Remote Sens.*, vol. 53, no. 7, July 2015.
- [138] S. G. Jin, O.F. Luo, , and J. Cho, "Systematic errors between VLBI and GPS precipitable water vapor estimations from 5 year co-located measurements," *J. Atm. Sol. Ter. Phy.*, vol. 71, no. 2, pp. 264–272, 2009.
- [139] "VLBI website," <ftp://cddis.gsfc.nasa.gov/vlbi/ivsproducts/trop/>, [Online; latest accessed 17-Mar-2018].
- [140] "National university of singapore, singapore. geography weather station," <https://inetapps.nus.edu.sg/fas/geog/>, [Online; latest accessed 30-June-2018].
- [141] "Cemaden's observational network for natural disaster risk monitoring," <http://www.cemaden.gov.br/mapainterativo/>, [Online; latest accessed 30-June-2018].
- [142] National Environment Agency, "Annual weather review," <https://www.nea.gov.sg/corporate-functions/resources/publications/annual-reports>, Oct. 2017.
- [143] S. G. Jin, O. F. Luo, and S. Gleason, "Characterization of diurnal cycles in ZTD from a decade of global GPS observations," *J. Geod.*, vol. 83, no. 6, pp. 537–545, June 2009.
- [144] G. Li, F. Kimura, T. Sato, and D. Huang, "A composite analysis of diurnal cycle of gps precipitable water vapor in central japan during calm summer days," *Theor. Appl. Climatol.*, vol. 92, pp. 15–29, 2008.
- [145] S. Dev, F. M. Savoy, Y. H. Lee, and S. Winkler, "WAHRIS: A low-cost, high-resolution whole sky imager with near-infrared capabilities," in *Proc.IS&T/SPIE Infrared Imaging Systems: Design, Analysis, Modeling, and Testing*, May 2014.

- [146] W. P. Elliott and J. K. Angell, “Variations of cloudiness, precipitable water, and relative humidity over the United States 1973-1993,” *Geophys. Res. Letters*, vol. 24, no. 1, pp. 41–44, Jan 1997.
- [147] S. Dev, S. Manandhar, Y. H. Lee, and S. Winkler, “Study of clear sky models for Singapore,” in *Proc. PIERS*, 2017.
- [148] L. Luini and C. G. Riva, “Improving the accuracy in predicting water-vapor attenuation at millimeter-wave for earth-space applications,” *IEEE Trans. Ant. Prop.*, vol. 64, no. 6, pp. 2487–2493, Jun 2016.
- [149] E. Fu, K. Zhang, F. Wu, X. Xu, and K. Marion, “An evaluation of gnss radio occultation technology for Australian meteorology,” *J. Glob. Pos. Sys*, vol. 6, no. 1, pp. 74–79, 2007.

Appendix A

Details on location of GPS, Radiosonde and Weather Stations

The table below provides the details on the IGS GPS and radiosonde station locations and distance between the two co-located stations.

S.N	IGS GPS Station Id (Lat, Long)	Radiosonde Station Id (Lat, Long)	Station Height (m)	Distance (km)	Years
Temperate Stations					
1	ALRT (82.49 N, 062.34 W)	71082 82.50 N, 062.33 W)	76.0	1	2010-2015
2	NYAL (78.93 N, 011.87 E)	01004 78.93 N, 011.87 E)	8.0	3	2010-2015
3	CHUR (58.75 N, 094.08 W)	71913 58.73 N, 094.08 W)	29.0	2	2010-2015
4	FALK (51.69 S, 057.87 W)	88889 51.81 S, 058.45 W)	73.0	42	2010-2015
5	HERS (50.86 N, 000.33 E)	03882 50.90 N, 000.32 E)	52.0	4	2010-2015
6	GLSV (50.36 N, 030.49 E)	33345 50.40 N, 030.56 E)	167.0	7	2010-2015
7	BRST (48.38 N, 004.49 W)	07110 48.45 N, 004.42 W)	98.0	9	2010-2015
8	STJO (47.59 N, 052.67 W)	71802 47.51 N, 052.78 W)	113.0	12	2010-2015
9	HOB2 (42.80 S, 147.43 E)	94975 42.83 S, 147.50 E)	27.0	7	2010-2015

Appendix A. *Details on location of GPS, Radiosonde and Weather Stations*

10	ANKR (39.88 N, 032.75 E)	17130 39.95 N, 032.88 E)	891.0	14	2010-2015
11	VNDP (34.55 N, 120.61 W)	72393 34.75 N, 120.56 W)	121.0	23	2010-2015
Sub-Tropical Stations					
12	TSKB (36.10 N, 140.08 E)	47646 (36.05 N, 140.13 E)	31.0	7	2010-2015
13	BRMU (32.37 N, 064.69 W)	78016 (32.37 N, 064.68 W)	37.0	1	2010-2015
14	AIRA (31.82 N, 130.59 E)	47827 (31.55 N, 130.55 E)	31.0	30	2010-2015
15	LHAZ (29.65 N, 091.10 E)	55591 (29.66 N, 091.13 E)	3650.0	3	2010-2015
16	MAST (27.76 N, 015.63 W)	60018 (27.67 N, 015.58 W)	105.0	11	2010-2015
17	UFPR (25.44 S, 049.23 W)	83840 (25.51 S, 049.16 W)	908.0	10	2010-2015
18	KUNM (25.02 N, 102.79 E)	56778 (25.01 N, 102.68 E)	1892.0	11	2010-2015
19	ALIC (23.67 S, 133.88 E)	94326 (23.80 S, 133.88 E)	546.0	14	2010-2015
20	HKSL (22.37 N, 113.92 E)	45004 (22.31 N, 114.17 E)	66.0	14	2010-2015
Tropical Stations					
21	NRMD (22.22 S, 166.48 E)	91592 (22.26 S, 166.45 E)	72.0	5	2010-2015
22	TOW2 (19.26 S, 147.05 E)	94294 (19.25 S, 146.76 E)	9.0	30	2010-2015
23	LAUT (17.60 S, 177.44 E)	91680 (17.75 S, 177.44 E)	18.0	17	2010-2015
24	HYDE (17.41 N, 078.55 E)	43128 (17.45 N, 078.46 E)	545.0	11	2010-2015
25	ABMF (16.26 N, 061.52 W)	78897 (16.26 N, 061.51 W)	11.0	2	2010-2015
26	PIMO (14.63 N, 121.07 E)	98433 (14.56 N, 121.36 E)	614.0	30	2010-2015
27	IISC (13.02 N, 077.57 E)	43295 (12.96 N, 077.58 E)	921.0	7	2010-2015
28	SALU (02.59 S, 044.21 W)	82281 (02.60 S, 044.23 W)	53.0	2	2010-2015
29	NTUS (01.34 N, 103.67 E)	48698 (01.33 N, 103.88 E)	16.0	28	2010-2015

Table A.1: IGS GPS stations and radiosonde stations

Appendix A. Details on location of GPS, Radiosonde and Weather Stations

The table below provides the details on the SiReNT GPS station locations and data availability.

Station Id	Station Coordinates (Lat, Long)	Years
SRPT	1.44N, 103.78E	2015
SNPT	1.37N, 103.84E	2015
SLYG	1.37N, 103.97E	2015
SNYU	1.34N, 103.67E	2015
SNUS	1.29N, 103.77E	Aug, 2015 to Aug, 2016
SSTS	1.23N, 103.64E	Aug, 2015 to Aug, 2016

Table A.2: SiReNT station location and data availability

The table below provides the details on the Weather station locations and data availability.

Country	Place (Lat, Long)	Weather Parameters	Resolution	Years
Singapore	NTU (01.34 N, 103.67 E)	T , RH , SR DPT and RR	1 min	2010-2015
Singapore	NUS (01.29 N, 103.77 E)	RR	5 min	2016
Brazil	São Luis (02.59 S, 044.23 W)	RR	10 min (Rain) 1 hr (No-Rain)	2016-2017
Brazil	Recife (08.05 S, 034.92 W)	RR	10 min (Rain) 1 hr (No-Rain)	2017
India	Bangalore (13.03 N, 077.58 E)	RR	1 hr	2010

Table A.3: Weather stations

The table below summarizes different databases and their data availability.

Instrument	Number of Stations	Temporal Resolution	Number of Years
Radiosonde	174	2 times a day	4 Years (2012-2015)
METAR	174	30 mins - 1 hr	4 Years (2012-2015)
GPS (IGS)	29	5 mins	6 Years (2010-2015)
GPS (SiReNT)	6	5 mins	1 Year (2015)
MODIS	9	2 times a day	1 Year (2015)
Weather Station	5	1 min - 1 hr	6 Years (2010-2015)

Table A.4: Summary of all instruments and related data availability