Implementation of a regional climate model to improve downscaling and modeling of precipitation over western maritime continent

Singh, Saurabh Kumar

2018

Singh, S. K. (2018). Implementation of a regional climate model to improve downscaling and modeling of precipitation over western maritime continent. Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/73371>

<https://doi.org/10.32657/10356/73371>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

Implementation of a Regional Climate Model to Improve
Downscaling and Modeling of Precipitation over
Western Maritime Continent

Saurabh Kumar Singh

School of Civil and Environmental Engineering

August 2017

Implementation of a Regional Climate Model to Improve
Downscaling and Modeling of Precipitation over
Western Maritime Continent

Saurabh Kumar Singh

School of Civil and Environmental Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirement for the degree of

Doctor of Philosophy

August 2017

ACKNOWLEDGEMENTS

I would like to thank my Ph.D. supervisor, Assoc Prof Lo Yat-Man, Edmond for all the support he gave me in the completion of my thesis. He always encouraged me to do better at every stage, and I am indeed fortunate to get an opportunity to work under him. I have always admired his ability to grasp new knowledge with ease and carry multitudes of research with great efficiency. I consider him as an "ocean of knowledge" and would always look up to him as a source of wisdom.

I would like to thank my Co-supervisor Assoc Prof Qin Xiaosheng for helping me with ideas, resources, and guidance throughout my Ph.D. I would like to thank School of Civil and Environmental Engineering, Nanyang Technological University and the Singapore Government for giving me this wonderful opportunity and funding my research.

I would like to thank, Institute of Catastrophe Risk Management, Nanyang technological university, along with Prof Pan Tso-Chien and other members of the Institute for their invaluable support throughout my Ph.D. research and awesome symposiums along with after parties.

I would like to thank my friends Abhijeet, Sheetal, Rajagopalan, Nikitha, Kaushik, Ashish, Abhishek, Nishanth, Anupam, Shankar, Daxia, Pradeep, Antreep, Manisha and the entire lunch dinner group and TEDxNTU club for their continuous love and support.

The Journey would have never been achievable without the blessing and support of my family. I am will always be grateful to Akansha, my parents Arun Kumar and Nutan Singh and my siblings Yuvraj, Shashank and Shakshi and most importantly Niraj mama.

ABSTRACT

The Western Maritime Continent (WMC) is recognized to have major weather and climate importance on both global and local scales. However, because of excess convection and complex topography, the region faces severe climate and weather modeling challenges. Further, the region has scarce station data, and Global Climate Models (GCMs) such as the Hadley Centre Global Atmospheric Model (HadGAM1) have been systematically found to incorrectly model precipitation. Thus, it is essential that WMC is properly represented by Regional Climate Models (RCMs) to capture and reproduce regional precipitation correctly. The discrepancy in GCM to accurately modeling rainfall over the WMC region results in errors that propagate to both the tropics and extra-tropics precipitation output as well.

The RCMs have been observed to provide “added information” over the GCM data as they consider local topography and climate. The “added information” by RCM was used in three studies conducted in this thesis. In the first study, an integrated downscaling approach, i.e., combination of statistical and dynamical downscaling was applied over the WMC. RCM downscaled predictors having better spatial resolution capture the local information more appropriately and were used as input to a statistical downscaling model. The RCM used was the Weather Research and Forecasting (WRF) model. Integrated downscaling approach would thus fill data gap over data sparse regions of the WMC with a good quality downscaled dataset, thus, enabling a better analysis of precipitation for impact analysis. The most noticeable improvements are for Singapore location where using WRF predictors resulted in an RMSE value of 1.37 (0.81) for mean monthly rainfall for validation (calibration) period while R2 predictors had a much larger RMSE of 3.41 (3.86).

In the second study, one of the outputs of CMIP5 GCM model output data, i.e., CESM1 bias corrected using ECMWF by NCAR was provided to WRF to downscale precipitation over the year 2030 to 2060 for RCP 4.5 and RCP 8.5 scenarios. Furthermore, the WRF precipitation was extracted and was used to generate future IDF curves over three locations in the WMC. In the final study, the Wards method was used to find clusters across 43 locations in WMC. The spatial-temporal groups formed were explained using CHIRPS (for historical) and WRF CMIP5 (for future). This is among the few studies done over WMC that utilizes outputs from two different RCMs to model future changes in extreme rainfall intensity. The output from the study would be significant in planning drainages and accessing future flooding considering the climate change.

A region-specific parametrization scheme validated with observation data to capture convective precipitation was found for WRF and used for the three different studies. An integrated approach capable of producing better mean and extreme rainfall statistics than via conventional statistical downscaling from GCM was verified in the first study over all the four locations.

In the second study, bias-corrected IDF curves were constructed for the two RCPs 4.5 and 8.5 scenarios. Significant increments in precipitation were observed over the three locations in WMC. The WRF-IDF was further compared to CORDEX-IDF, and it was noted that WRF predicted a greater increment in Intensity as compared to CORDEX. Both WRF and CORDEX outputs are considered as different realizations of the two-member ensemble. Hence these two realizations can be used to access the probable uncertainty in future IDF curves. The predicted increment in extreme rainfall intensity by RCP 8.5 was larger compared to RCP 4.5 and historical for almost all the locations and return periods. For Singapore (100 year RP and RCP 8.5) average intensity increase was 36.1% for WRF-IDF at RCP 8.5 while it was 16.8% for

CORDEX-IDF. Over Jakarta (100 year RP) the RCP 8.5 WRF-IDF showed 52.5% average increment while the WRF-IDF showed 74% increase in average intensity. Lastly, the cluster analysis revealed some interesting new grouping in both spatial and temporal dimensions from the third study. This is the first study over WMC that examined the changes in spatial and temporal monthly precipitation using future RCP downscaled CMIP5 GCM data. It was observed that Spatial Groups (SG) formed for WRF-CMIP5 (RCP 4.5 and RCP 8.5) were very different from historical. The difference between SGs would signify changes in precipitation over the next 25 years. While comparing the SGs for RCP 4.5 and RCP 8.5, it was observed that of all the landmasses, Borneo would see the most significant redistribution and would become more homogenous in its rainfall pattern.

Table of Contents

ACKNOWLEDGEMENTS	I
ABSTRACT	II
TABLE OF CONTENTS	V
LIST OF FIGURES	IX
LIST OF TABLES	XIII
ACRONYMS AND ABBREVIATION	XV
1 INTRODUCTION	1
1.1 BACKGROUND	1
<i>1.1.1 Implementation of Dynamical downscaling to obtain High-Resolution Output</i>	<i>3</i>
<i>1.1.2 Implementation of Statistical downscaling to obtain High-Resolution Output</i>	<i>4</i>
1.2 CLIMATE CHANGE AND VULNERABILITY OF WESTERN MARITIME	
CONTINENT	6
1.3 LIMITED OBSERVED DATA AVAILABLE IN WESTERN MARITIME CONTINENT	9
1.4 CLIMATE CHANGE IMPACT ON HYDROLOGY	11
1.5 OBJECTIVE OF THE PH.D. WORK	12
<i>1.5.1 An Integrated Approach to Downscale Precipitation over Western Maritime Continent</i>	<i>12</i>
<i>1.5.2 Intensity-Duration-Frequency Curves Under Changing Climate for Year Range 2030-2060:</i>	<i>14</i>
<i>1.5.3 Cluster Analysis of Spatial and Temporal Monthly Precipitation over Western Maritime Continent</i>	<i>15</i>
1.6 OUTLINE OF THE THESIS	16
2 LITERATURE REVIEW	18
2.1 INTRODUCTION	18
2.2 MARITIME CONTINENT	18
2.3 GENERAL CLIMATE OVER MARITIME CONTINENT	21

2.4	NUMERICAL WEATHER PREDICTION PRINCIPLES	24
2.4.1	<i>Global Climate Models (GCMs)</i>	27
2.4.2	<i>RCM Models</i>	27
2.5	“ADDITION INFORMATION” ADDED BY RCM	29
2.6	CLIMATE CHANGE IMPACTS	32
2.7	CLIMATE DOWNSCALING	36
2.7.1	<i>Dynamic Downscaling</i>	38
2.7.2	<i>Statistical Downscaling Method (SDSM) and Automated Downscaling Method (ADM)</i>	
	40	
3	AN INTEGRATED DYNAMICAL AND STATISTICAL APPROACH TO DOWNSCALE PRECIPITATION OVER WESTERN MARITIME CONTINENT	44
3.1	INTRODUCTION	44
3.2	MODEL DESCRIPTION AND IMPLEMENTATION	46
3.2.1	<i>WRF Model</i>	47
3.2.2	<i>ASD Model</i>	49
3.2.3	<i>Data Used</i>	54
3.3	RESULTS AND DISCUSSION	55
3.3.1	<i>Verification of dynamically downscaled output from WRF</i>	55
3.3.2	<i>Comparison of WRF and CFSR variables with TRMM</i>	58
3.3.3	<i>Calibration and Validation of ASD model using Integrated Approach</i>	60
3.4	CONCLUSIONS OF INTEGRATED DOWNSCALING APPROACH	69
4	CONSTRUCTION OF FUTURE IDF CURVES USING RCM DATA OVER WESTERN MARITIME CONTINENT	71
4.1	INTRODUCTION	71
4.2	DATA AND EXTREME VALUE DISTRIBUTION USED TO CONSTRUCT IDF CURVES	74
4.2.1	<i>NCAR CESM Global Bias-Corrected CMIP5</i>	74

4.2.2	<i>CORDEX dataset over East Asia</i>	75
4.2.3	<i>Extreme Value Distribution</i>	76
4.3	CONSTRUCTION AND BIAS CORRECTION OF HISTORICAL IDF CURVES	77
	IDF CURVE BIAS CORRECTION	88
4.4	COMPARISON OF PROJECTED FUTURE IDF CURVE FOR SINGAPORE, KUALA LUMPUR AND JAKARTA USING BIAS-CORRECTED WRF AND CORDEX DATA	91
4.4.1	<i>WRF derived IDF curve and historical IDF curves at station level</i>	95
4.4.2	<i>Comparison between CORDEX derived IDF curve and historical IDF curves at station level</i>	105
4.4.3	<i>Comparison between WRF and CORDEX derived IDF curves</i>	110
4.5	CONCLUDING REMARKS	110
5	CLUSTER ANALYSIS OF SPATIAL AND TEMPORAL MONTHLY PRECIPITATION OVER WESTERN MARITIME CONTINENT	112
5.1	INTRODUCTION	112
5.2	STUDY AREA AND DATA	114
5.3	METHODS	118
5.3.1	<i>Cluster Analysis (Ward's Method)</i>	118
5.4	RESULTS AND FINDINGS FOR 5 KM CHIRPS DATASET	120
5.4.1	<i>Grouping in Space</i>	122
5.4.2	<i>Monthly Rainfall Over Spatial Clusters</i>	125
5.4.3	<i>Temporal Clusters</i>	128
5.5	RESULTS FOR CLUSTER ANALYSIS ON 25 KM CHIRPS AND FUTURE WRF-CMIP5 DATASET TO INVESTIGATE REDISTRIBUTION OF RAINFALL	130
5.5.1	<i>Spatial Grouping of 25 km Historical CHIRPS data</i>	132
5.5.2	<i>Spatial Grouping for RCP 4.5 WRF-CMIP5 data (2030 – 2060)</i>	134
5.5.3	<i>Spatial Grouping for RCP 8.5 WRF-CMIP5 data (2030 – 2060)</i>	136
5.5.4	<i>Temporal Groupings of 25 km Historical CHIRPS data</i>	138

5.5.5	<i>Temporal Grouping of RCP 4.5 WRF-CMIP5 data (2030 – 2060)</i>	139
5.5.6	<i>Temporal Grouping of RCP 8.5 WRF-CMIP5 data (2030 – 2060)</i>	140
5.6	SUMMARY	142
5.6.1	<i>Distribution of rainfall using Cluster Analysis of 5 Km CHIRPS dataset over WMC</i>	143
5.6.2	<i>Redistribution of Rainfall over WMC</i>	144
	<i>Comparison of RCP clusters over Borneo</i>	145
6	SUMMARY AND CONCLUSION	147
6.1	SUMMARY	147
6.2	INTEGRATED DYNAMICAL-STATISTICAL APPROACH TO DOWNSCALE PRECIPITATION OVER WMC	148
6.2.1	<i>Results from Downscaling of daily precipitation</i>	148
6.2.2	<i>Recommendations for Future Studies</i>	149
6.3	FUTURE IDF CURVES USING RCM DATA OVER WMC	150
6.3.1	<i>Calculating bias and downscaling factors using historical IDF Curves</i>	150
6.3.2	<i>Bias-corrected future IDF curves</i>	151
6.3.3	<i>Recommendation for Future Work</i>	151
6.4	CLUSTER ANALYSIS OF SPATIAL AND TEMPORAL MONTHLY PRECIPITATION OVER WMC	152
6.4.1	<i>Cluster analysis using 5 km CHIRPS data from Jan 1980 to Dec 2016</i>	152
6.4.2	<i>Cluster analysis using 25 km WRF-CMIP5 data along with 25 km averaged CHIRPS data</i>	153
6.4.3	<i>Recommendations for Future Studies</i>	153

LIST OF FIGURES

Figure 1.1 A map for multiple climate hazards over Southeast Asia. [Reproduced from:(Yusuf et al., 2009)]. Legends indicate 0 (lowest vulnerability) and 1(highest vulnerability)	8
Figure 1.2 Annual Flood frequency over Southeast Asia [Reproduced from:(Yusuf et al., 2009)]. The legends show 0.762 (highest) and 0(lowest) flood frequency events per year from 1980 to 2001.	10
Figure 2.1 Plot showing the MC: 10°S–20°N and 90°–150°E.	19
Figure 2.2 Mean seasonal daily precipitation (mm/day) plotted from Climate Prediction Centre (CPC) Merged Analysis of Precipitation (CMAP) climatology (a) DJF (b) MAM (c) JJA (d) SON [Reproduced from (Strachan, 2007)]	22
Figure 2.3 Seasonal wind m/sec (at 925 hPa) derived from ECMWF data. (a) DJF (b) MAM (c) JJA (d) SON. The color shows the magnitude of the winds while arrow shows the direction of the zonal winds [Reproduced from (Strachan, 2007)]	23
Figure 2.4 Monthly mean rainfall from TRMM with sea surface temperature (°C) superimposed onto Quick Scatterometer (QuickSCAT) wind (mm/sec) for August(top) and February (bottom). Data average from 1998 to 2003 and 2000 to 2004 respectively [Reproduced from (Qu et al., 1994)	24
Figure 2.5 A framework representation of NWP [Reproduced from http://public.homepages.uconn.edu/~ria08001/ceml/ceml-both.html]	26
Figure 2.6 A flow chart summarizing components of WRF model [Reproduced from http://www2.mmm.ucar.edu/wrf/users/docs/user_guide/users_guide_chap1.htm l]	28
Figure 2.7 A figure showing topographic details as resolved by RCM-50km(left) and GCM-300km(right) and [Reproduced from Christensen O.B., 2001]	30
Figure 2.8 Simulated precipitation by GCM and RCM over Britain with observation [Reproduced from Jones et al., 2004]	31
Figure 2.9 (a) GCM (PCM) and(b)) RCM (MM5) simulations over the Rocky Mountains for temperature change over the Rocky Mountains [Reproduced from Leung et al., 2004]	32
Figure 2.10 Global Average Surface change under different scenarios. Reproduced from: (Alexander et al., 2013).....	36
Figure 3.1 Flow chart of integrated approach (left) and conventional SD method (right)	47
Figure 3.2 WRF parametrization scheme in this study and their interdependence (adapted from Skamarock et al., 2008)	49

Figure 3.3 Correlations of the top 5 predictors extracted from WRF (left) and R2 (right) with rainfall for Singapore station (Significant predictors with p-value less than 0.05)	52
Figure 3.4 Decision tree corresponding to the two grids WRF (top right) and R2 (bottom right). The first parent is the top-most node in the decision tree shown. Time range covers 1st Jan 1980 – 1st Jan 2005 or 9133 daily records.....	53
Figure 3.5 Simulated WRF (50km x 50km) (left) and observed TRMM (25km x 25km) (right) from Jan 2000 – Dec 2008, 9-year seasonal precipitation (mm/day) comparison.	57
Figure 3.6 Correlation between predictors extracted from WRF (left) and CFSR (right) with observed TRMM data over the WMC (time range from Jan 2000 to Dec 2008).....	59
Figure 3.7 Simulated 100 realization of calibrated precipitation monthly mean (time range Jan 1980 - Dec 1998) ensemble for Singapore region using (a)WRF predictors (top) & (b)R2 (bottom)	60
Figure 3.8 25th and 75th quantile of the calibration results for the Singapore location using the WRF predictors (left) and R2 predictors (right) along with observed median in dotted line. The statistics plotted in the figures are (a) mean, (b) standard deviation (c) wet days (d) excess kurtosis	62
Figure 3.9 25th and 75th quantile of the validation results for Singapore location using the WRF predictors (left) and R2 predictors (right) along with observed median in dotted line. The statistics plotted in the figures are (a) mean, (b) standard deviation (c) wet days (d) excess kurtosis	62
Figure 3.10 Absolute error between simulated monthly mean (left) and number of wet days (right) over the calibration period at four locations (1) Singapore (2) Hat Yai (3) Labuan and (4) Jakarta, using (a) R2 predictors (dark gray) and (b) WRF predictors	64
Figure 3.11 Absolute error between simulated monthly mean (left) and number of wet days (right) over the validation period at four location (1) Singapore (2) Hat Yai (3) Labuan and (4) Jakarta, using (a) R2 predictors (dark gray) and (b) WRF predictors	65
Figure 3.12 Comparison of seasonal daily precipitation rate mean over the four locations for the validation dataset.	69
Figure 4.1 A flowchart of the methodology for constructing bias corrected future IDF curve.	74
Figure 4.2 CORDEX East Asia Model domain and orography (m). [Reproduced from https://cordex-ea.climate.go.kr/main/modelsPage.do]	76
Figure 4.3 WRF physics parameterization for running CMIP5 model.....	80
Figure 4.4 Comparison plots for seasonal rainfall rate (mm/day) between CHIRPS (left) and WRF (right) for the baseline period (1980-2005).	81

Figure 4.5 Future mean precipitation as predicted by RCP 4.5 (left) and RCP 8.5 (right) using WRF model for the year 2030-2060.	82
Figure 4.6 percentage precipitation changes (future modeled precipitation – observed historical precipitation)/observed precipitation) for WRF model RCP 4.5 (left) and RCP 8.5 (right) for the year 2030-2060	83
Figure 4.7 Gumbel fit with observed annual extreme values for Singapore rain gauge from the year 1980 to 2005.	85
Figure 4.8 Constructed IDF curve for this study using Changi (Singapore) station data (left) and the IDF curve published by Public Utilities Board (PUB) Singapore for station averaged over Singapore (right)	86
Figure 4.9 Present station IDF curves (a) Singapore (b) Jakarta (c) Kuala Lumpur	87
Figure 4.10 Singapore WRF-IDF curves along with CFSR IDF for time range 1980-2005, RP=100(left) and RP=50(right).	89
Figure 4.11 Singapore CFSR-IDF curves along with station IDF for time range 1980-2005, RP=100(left) and RP=50(right).	91
Figure 4.12 Future bias corrected IDF curves generated from WRF dataset for RCP 4.5 (left) and RCP 8.5 (right) over (a) Singapore (b) Kuala Lumpur and (c) Jakarta	93
Figure 4.13 Future bias corrected IDF curves generated from CORDEX dataset for RCP 4.5 (left) and RCP 8.5 (right) over (a) Singapore (b) Kuala Lumpur and (c) Jakarta.....	94
Figure 4.14 Comparison plot for Singapore (RP=100 year) showing IDF points corresponding to historical, RCP 4.5 and RCP 8.5.....	95
Figure 4.15 Comparison plot for Jakarta (RP=100 year) showing IDF points corresponding to historical, RCP 4.5 and RCP 8.5	98
Figure 4.16 Comparison plot for Kuala Lumpur (RP=100 year) showing IDF points corresponding to historical, RCP 4.5 and RCP 8.5	99
Figure 4.17 A Comparison plot for Singapore RP=50 year (left) and RP=25 (right) showing IDF points corresponding to historical, RCP 4.5 and RCP 8.5	100
Figure 4.18 Comparison plot for Jakarta RP=50 year (left) and RP=25 (right) showing IDF points corresponding to historical, RCP 4.5 and RCP 8.5	102
Figure 4.19 Comparison plot for Kuala Lumpur RP=50 year (left) and RP=25 (right) showing IDF points corresponding to historical, RCP 4.5 and RCP 8.5	102
Figure 5.1. A flowchart showing clustering framework in this study	114
Figure 5.2 Spatial map showing a) topography and b) land use category, for the study region Western Maritime Continent.....	115
Figure 5.3 The steps involved in wards clustering (reproduced from cs-people.bu.edu/evimaria/cs565-12/lect5.pptx)	120
Figure 5.4 Spatial distribution of CHIRPS accumulated seasonal rainfall (mm) over WMC averaged from the year 1980 to 2016,	122

Figure 5.5 Cluster Analysis groups (Group1 – Group 5) (a) dendrograms based on their spatial similarity and (b) spatial representation of the spatial clusters...	124
Figure 5.6 The box-plots denote the 25% and 75% of the monthly rainfall(mm/day) for the Spatial Groups 1 to 5. (a to e).....	127
Figure 5.7 Cluster analysis applied to monthly rainfall for SG 1 to 5 based on temporal grouping for each SG.....	129
Figure 5.8 Cluster Analysis groups (SG1 – SG3) for CHIRPS historical (25 km) (a) dendrograms based on their spatial similarity and (b) Spatial representation of the clusters on a map.....	133
Figure 5.9 Cluster Analysis groups (SG1 – SG 3) for RCP 4.5 (a) dendrograms based on their spatial similarity and (b) SG representation on a map	135
Figure 5.10 Cluster Analysis groups (SG11 – SG3 3) for RCP 8.5 (a) dendrograms based on their spatial similarity and (b) spatial representation of the clusters on a map.....	137
Figure 5.11 The box plots (left) and Dendrograms of TGs (right) for different 3 SGs discovered in the CHIRPS historical data (25 km)	139
Figure 5.12 The box plots (left) and Dendrograms of TGs (right) for different 3 SGs discovered in the WRF-CMIP5 (RCP 4.5)	140
Figure 5.13 The box plots (left) and Dendrograms of TGs (right) for different 3 SGs discovered in the WRF-CMIP5 (RCP 8.5)	142
Figure a.1 Percentage error in between WRF and CFSR IDFs (left) and CFSR and Observed station ODFs (right) for RP=100 years (top) and RP= 50 years (bottom) over Kuala Lumpur	155
Figure a.2 Percentage error in between WRF and CFSR IDFs (left) and CFSR and Observed station ODFs (right) for RP=100 years (top) and RP= 50 years (bottom) over Jakarta	156

LIST OF TABLES

Table 2.1 Table summarizing different statistical downscaling methods and tools used in downscaling literature (Wilby R., Charles, et al., 2004; Wilby R., Charles, S., Zorita, E., Timbal, B., Whetton, P., Mearns, L., 2004).	41
Table 2.2 Pros and cons of statistical and dynamical downscaling (Wilby R. L. et al., 2002).	43
Table 3.1 Station location and data interval used for calibration and validation	46
Table 3.2 List of predictors extracted from WRF output and R2 data in statistical downscaling	51
Table 3.3 Land/grid percentage for Reanalysis-II grid and WRF grid.	64
Table 3.4 Error parameters for the statistical parameters of the ensemble w.r.t the observed station values in calibration. The RMSE value has units mm for the mean, standard deviation and excess kurtosis, and day for the number of wet days.	67
Table 3.5 Error parameters for the statistical parameters of the ensemble w.r.t the observed station values in validation. The RMSE value has units mm for the mean, standard deviation and excess kurtosis, and day for the number of wet days.	68
Table 4.1 The locations over which IDF curves are constructed	73
Table 4.2 Table showing IDF values corresponding to which historical-station level IDF curves were constructed.	88
Table 4.3 100-year return period IDF curve and the relative change intensity percentage corresponding to RCP 4.5 and RCP 8.5 at Singapore, Jakarta, and Kuala Lumpur	96
Table 4.4 50-year return period WRF-IDF curve and the relative change intensity percentage corresponding to RCP 4.5 and RCP 8.5 at Singapore, Jakarta and Kuala Lumpur	103
Table 4.5 25-year return period WRF-IDF curve and the relative change intensity percentage corresponding to RCP 4.5 and RCP 8.5 at Singapore, Jakarta and Kuala Lumpur	104
Table 4.6 100-year return period CORDEX- IDF curve and the relative change intensity percentage corresponding to RCP 4.5 and RCP 8.5 at Singapore, Jakarta, and Kuala Lumpur	106
Table 4.7 50-year return period CORDEX-IDF curve and the relative change intensity percentage corresponding to RCP 4.5 and RCP 8.5 at Singapore, Jakarta and Kuala Lumpur	108

Table 4.8 25-year return period CORDEX-IDF curve and the relative change intensity percentage corresponding to RCP 4.5 and RCP 8.5 at Singapore, Jakarta and Kuala Lumpur	109
Table 5.1 Stations, coordinates (latitude and longitude), and altitude of the locations used in this study. Time range of the monthly data is from Jan 1981 to Dec 2016.....	117
Table 5.2 The distribution of SGs over four landmasses for Historical and future time period (RCP 4.5 and RCP 8.5).....	145
Table 5.3 Distribution of locations among the clusters over the five different landmasses in WMC for RCP 4.5and RCP 8.5 monthly rainfall data	146

ACRONYMS AND ABBREVIATION

AR4	Fourth Assessment Report
AR5	Fifth Assessment Report
ASD	Automated Statistical Downscaling
BMJ	Betts-Miller-Janjic scheme
CA	Cluster Analysis
CESM	The Community Earth System Model
CFSR	Climate Forecast System Reanalysis
CHIRPS	Climate Hazards Group InfraRed Precipitation with Station data
CMAP	CPC Merged Analysis of Precipitation
CMIP5	Coupled Model Intercomparison Project
CPC	Climate Prediction Centre
CORDEX	Coordinated Regional Climate Downscaling Experiment
DD	Dynamic Downscaling
DJF	December-January-February
ECMWF	European Centre for Medium-Range Weather Forecasts
EEPSEA	Economy Environmental Program for Southeast Asia
ERA40	ECMWF Reanalysis
GCM	Global Climate Model
GEV	Gumbel Extreme Value
GHGs	Green House Gases
HadCM3	Hadley Centre Climate Model, Version 3
HadGEM2-AO	Hadley Centre Global Environment Model,

	Version 2
HadRM3P	Hadley Centre Regional Climate Model, version 3
IDF	Intensity Duration Frequency
IPCC	Intergovernmental Panel on Climate Change
IQR	Interquartile Range
JJA	June-July-August
LARS-WG	Long Ashton Research Station Weather Generator
LW	Longwave
MAM	March-April-May
MM5	Mesoscale Model Version 5
MC	Maritime Continent
MJO	Madden-Julian Oscillation
NCAR	National Center for Atmospheric Research
NSE	Nash–Sutcliffe efficiency
NWP	Numerical Weather Prediction
PBL	Planetary Boundary Layer
PCM	Parallel Climate Model
PPC	Precipitating Convective Clouds
PRUDENCE	Prediction of Regional scenarios and Uncertainties for Defining European Climate change risks and Effects
QuickSCAT	Quick Scatterometer
R2	Reanalysis 2
RCM	Regional Climate Model

RCP	Representative Concentration Pathways
RMSE	Root Mean Square Error
RP	Return Period
RRTMG	Rapid Radiative Transfer Model for GCMs
SD	Statistical Downscaling
SG	Spatial Group
SON	September-October-November
SPCZ	South Pacific Convergence Zone
SST	Sea Surface Temperature
SW	Shortwave
TG	Temporal Group
TRMM	Tropical Rainfall measuring Mission
TSS	Total Sum of Squares
YSU	Yonsei University
WGEN	Weather Generation Model
WMC	Western Maritime Continent

1 Introduction

1.1 BACKGROUND

Climate change forced by either anthropogenic or natural cause can alter the likelihood as well as the strength of extreme weather and climate events. Inter-Governmental Panel on Climate Change (IPCC) fifth assessment report (AR5) (IPCC, 2013), reported the effect of climate change on temperature and water cycle. Alterations in temperature and water cycle were cited as the primary reasons behind enhanced extreme atmospheric events. IPCC observed that General Circulation Models (GCM) projected change in global surface temperature for the 21st century is likely to increase by 1.5°C to 2°C or higher (depending on the climate change scenarios) over the year 2060-2080, relative to the period 1850 to 1900. The increment in temperature accelerates evaporation and thus increases subsequent water vapor in the atmosphere (Trenberth, 2011). This change in the water cycle will result in an increment in the contrast between wet and dry seasons in different regions. It is projected that the extreme precipitation events over tropical and mid-latitude land surfaces will most likely be more frequent and intense by the year 2100.

Studying the impact of climate change has become important for both scientists, as well as water and catchment related management strategists. Most of the information related to the climate change as applied in hydrological studies are primarily obtained from coarse resolution Global Climate Models (GCMs). GCMs simulate the circulation over the land surface, atmosphere, cryosphere, and oceans, and are primary tools used for climate predictions. Coarse resolution GCMs are, however, an inadequate way of assessing regional climate. The inadequacy is because the resolution of GCMs is too coarse to resolve sub-grid processes. Sub-grid level

hydrologic processes are the most critical aspect not accounted in GCMs. This results in the unreliability of GCM at sub-grid scales (Firth et al., 2012).

Previous studies have reported that, although GCMs can fairly well represent global atmospheric circulations, their ability to simulate regional temperature and precipitation details are poor (Giorgi, 1990; Knutti et al., 2010; Randall et al., 2007). The inability of GCMs to properly simulate precipitation at fine scales is their major drawback making them unsuitable for hydrological impact studies. It is thus important not to just rely on GCMs for fine-scale precipitation studies as precipitation is highly variable in time and space and GCMs would then impart substantial uncertainty for hydrological impact studies.

Higher resolution data that captures local or sub-grid scale feature with greater precision is thus required for quality impact assessment. Higher precision data necessary for the hydrological impact assessments can be achieved using “downscaling.” Downscaling is a method of bringing low resolution to higher resolution, thus reducing the discrepancy between GCM scenarios and impact assessment. The downscaling process assumes that large-scale weather and climate have a strong influence on local weather and climate while disregarding the impact of local scales back on the global scale.

There are mainly two downscaling approaches: a) Dynamical downscaling and b) Statistical downscaling. In dynamical downscaling, a Regional Climate Model (RCM) is nested within a GCM or it can be run in stand-alone with GCM outputs to include higher grid resolution atmospheric physics for the area of interest. The statistical downscaling on the other hand tries to establish a statistical linkage between large-scale GCM data and observed local scale data.

1.1.1 Implementation of Dynamical downscaling to obtain High-Resolution Output

RCMs have a similar representation of atmospheric dynamics and physical processes as GCMs. However, RCMs have a higher resolution (typically 50 km or less) and cover a part of the earth (e.g., Asia, Southeast Asia, Europe, America). The higher resolution of RCMs also demands that the model should be driven at higher temporal resolution (~3 minutes or less) compared to GCMs (~30 minutes or less) to maintain numerical stability. Dynamical downscaling is conditioned to the state of the large-scale atmosphere. However, there is always an “additional value” expected when this approach is used to downscale large-scale information to a local level (McCarthy et al., 2001) <https://paperpile.com/c/JtUXiX/ohvt>. The dynamical downscaling is a useful technique to transform coarse resolution data to data of higher spatial resolution and information in studies such as a) study of the spatiotemporal precipitation structure over a complex topography, b) distribution of atmospheric variables over different land use and c) understanding local circulations within the atmosphere as dependent on sub-grid scale convection, land-sea breeze and tropical storms (Giorgi 1990). The high temporal resolution of dynamic downscaling allows the study of processes that have a high temporal frequency as well as only captured at high temporal resolution. Information such as precipitation frequencies, the variability of surface winds, occurrence and withdrawal of extreme events can only be captured by higher resolution outputs generated from dynamical downscaling.

Dynamical downscaling is achieved using RCM. It is essential that RCMs also represent scales below which they cannot explicitly resolve for atmospheric processes such as convection, radiation, and land surface processes. RCMs resolve these atmospheric processes using parameterization schemes. The parameterization within the RCMs is a simple representation of real-world meteorological processes. For example, the precipitation is simulated using both large and sub-grid scale schemes.

The larger-scale grid scheme accounts for clouds and precipitation from large-scale atmospheric processes that are resolved by the RCMs. Whereas, the convection scheme resolves the cloud and precipitation resulting from sub-grid processes.

Various RCMs are presently available for many regions; they are developed and maintained by different modeling centers around the world. Most of these RCMs have a different model formulation and internal variability and thus output different output scenarios. A significant effort has been conducted to quantify uncertainty within multiple RCMs. Some of these projects are PRUDENCE (Prediction of Regional Scenarios and Uncertainties for Defining European Climate change risks and Effects) (Christensen, 2001), ENSEMBLES that aimed to deliver probabilistic estimates of climate risk by integrating ensemble achieved from Earth system models (Hewitt, 2004) and NARCCAP (North American Regional Climate Change Assessment Program) (Kang et al., 2012). One another such program which is also used in this Ph.D. study is the CORDEX initiative (Strandberg et al., 2014).

RCMs are typically run at a horizontal resolution between 25km to 50km for climate modeling experiments (IPCC, 2007). The present study used an RCM, the Weather Research and Forecasting (WRF) Model at various resolutions. The resolution was decided on the resolution requirement of the study. For example, historical runs for downscaling from Reanalysis dataset were done at 15km, 25km, and 50km. However, when WRF was run for GCMs for historical and future scenarios, the horizontal resolution of the model was 25 km throughout. A detailed overview of RCM WRF is discussed in Section 3.

1.1.2 Implementation of Statistical downscaling to obtain High-Resolution Output

Many approaches have been developed to bridge the gap between the GCM and RCM outputs to the station scale weather data as required to improve the accuracy of impact

estimation. Most of the approaches are statistical and are commonly referred as statistical downscaling. Statistical downscaling in general tries to create a mapping $E(Y|X)$ in between large-scale (GCM) predictors X and observed station data or a local predictand Y , expressed as

$$E(Y|X) = f(X, Y) \quad (1.1)$$

Equation (1.1) is a mapping vector that should be estimated to achieve statistical downscaling where E is the projected or mapped value, and $f(X, Y)$ is the mapping function that requires station data and GCM predictors as input. Advanced statistical downscaling also models the variability not completely explained by the dependence of Y on X by using a random variable commonly represented as noise. Statistical downscaling is achieved using regression. The term X for predictors is also termed as covariates. In statistical downscaling, there is always a large-scale covariate corresponding to every observed (station) data. Therefore, the calibration is done using observed, station data, i.e., a relationship between predictors and predictand is established to relate a time series rather than just relating the statistical distribution between predictors and predictand.

The statistical downscaling approach is divided into two steps; the first step involves selection of predictors and the second is developing the model and parameter estimation. The first step often requires a transformation of the predictors before feeding it into the model which is step 2. Some of the transformations include: predictor selection, filling missing data, normalization and checking for covariance between the predictors. An improvement in statistical downscaling approach can be achieved either by improving step 1 and step 2. In step 1, an improved predictor dataset can be used or a better method methodology for filling missing data and scaling. However, improving the results of statistical downscaling by altering step 2 can only be achieved by using an advanced methodology and minimizing the uncertainty.

In this Ph.D. study, the accuracy of statistical downscaling was improved by improving the step 1. Higher resolution predictors extracted from WRF output was used in this study. The hypothesis behind feeding higher resolution predictors was that higher resolution predictors have more local scale information needed for rainfall prediction. A detailed discussion of this approach can be found in Chapter 3.

1.2 CLIMATE CHANGE AND VULNERABILITY OF WESTERN MARITIME CONTINENT

The tropical weather and climate are controlled by convection and latent heat released from convection. The three core regions having dominant convection in the tropics are Southeast Asia (SEA) a part of WMC, the equatorial African landmass, and the Amazon (Schumacher, 2013; Tan, 2014). However, the monsoonal region of the Maritime Continent (MC) is recognized as the most dominant source of convection (Meehl, 1987). The changes in the convection over the Maritime region has been observed to affect the circulation around the world (Meehl, 1987; Sardeshmukh et al., 1988)

The mean climate over the WMC (Figure 2.1) is a result of various combinations of a weather system that occurs over various ranges of spatial and temporal scales. The climatic events over WMC range from a single cumulus cloud to cloud superclusters and tropical storms, and further up to global scales, an inter-annual variation of events resulting in the El Nino Southern Oscillations. The multi-modal variability of convection over the spatiotemporal scale makes the WMC a challenging domain of study.

The climate of WMC is of great importance and has a significant influence on inhabitants of this densely populated region. Agriculture within the region that serves as a livelihood is also largely dependent on the climate. For example, agriculture employs nearly 50% of the Indonesian population and contributes to

around 16% of the Gross National Product (data.worldbank.org/country/Indonesia). Most farmers within the WMC rely entirely on rainfall for irrigation, and any variabilities in rainfall can impact on cultivation and the national economy. For example, an El Nino-induced drought over Indonesia could result in an arid condition which is coupled with late and weak monsoonal rains (Glover, 2006). The El-Nino conditions also could lead to forest and crop fires resulting in health hazard of millions of people in WMC. Beside El-Nino several other weather and climate phenomenon determine the climate over WMC. Some of these climate anomalies are the Indian Ocean Dipole (IOD) (NH et al., 1999), Madden Julian Oscillation (Zhang, 2005) and La-Nina (Holton et al., 1990).

Land change associated with industries and agriculture, which involves logging and clearing of crops also alters the climate within the region. The land-atmosphere interaction is altered, changing the radiation balance and exchange of sensible heat between the atmosphere and land surface. The intensive agriculture also sometimes results in land degradation which would lead to soil erosion after heavy, intense rainfall. On the other hand, the WMC also consists of urban areas such as Singapore, Jakarta, Malaysia, e.t.c. The increasing trend of intense rainfall over these urban areas affects and challenges the drainage system. A panel tasked to monitor drainage design for Singapore (Chua, 2002) observed that the rainfall intensity had been continuously increasing since last 30 year.

A 2009 study done over Southeast Asia by the Economy Environmental Program for Southeast Asia (EEPSEA) mentioned that this region is most vulnerable to climate change risk (see Figure 1.1) (Yusuf et al., 2009). The vulnerability in the region was defined by the scale 0 (less vulnerable) to 1 (highly susceptible). The parameters used to determine vulnerability were: exposure, sensitivity and adaptive capacity. This vulnerability could further be related to exposure, sensitivity and adaptive capacity of a region as shown schematically by Equation 1.2.

Figure 1.1 A map for multiple climate hazards over Southeast Asia. [Reproduced from:(Yusuf et al., 2009)]. Legends indicate 0 (lowest vulnerability) and 1(highest vulnerability)

$$\text{Vulnerability} = f(\text{exposure, sensitivity, adaptive capacity}) \quad (1.2)$$

According to the IPCC AR4 (2007), **exposure** can be defined as “the degree and nature to which an area is exposed to a significant climatic variations”, **sensitivity** is “the extent to which a system is affected, either adversely or beneficially, by climate-related stimuli” and **adaptive capacity** is defined as “the ability of a system to adjust to climate change (including climate variability and extremes) (Yusuf et al., 2009). The climate change would result in intensified and frequent extreme weather events over Southeast Asia. The extreme events would lead to loss of rich forest, lesser crop yield, coastal resource damage, disease outbreak and many other related economic losses and human sufferings. It was found that Thailand, Indonesia, Philippines, and Vietnam would suffer the most severe effects of climate change compared to other regions over Southeast Asia. These studies suggested that Southeast Asia has a significant stake in taking a proper action for mitigating climate change risks.

1.3 Limited Observed Data available in Western Maritime Continent

Simulation and modeling methods using historical data are widely applied to prepare a response to weather and climate impacts. However, these approaches require long historical records to provide reliable information for past and future (Measey, 2010). Furthermore, in most of the regions around the world and especially in developing nations, the collection of climate and weather data are intermittent. The poor maintenance of data and lack of historical records could lead to inadequate mitigation plans (Boer et al., 2004).

Precipitation is one of the most important variables that are required by hydrological models. The success of the model can be more dependent on the quality of precipitation data than the complexity of the model (Gan et al., 1997). However, most rainfall-runoff simulation outputs in sparse data region are inadequate because the spatial variability of rainfall is not adequately represented. Furthermore, the catch of each rain gauge is small and represents only a small radius around each gauge instrument. Thus, rain gauge networks are typically too sparse, especially so in the WMC to provide an adequate quality of spatial and temporal scale rainfall data. The quality of the available rainfall is therefore inadequate over most of the region to capture a convective storm data correctly.

It has also been found that flood events are more severe and frequent (see Figure 1.2) in the tropical regions and the river basins are also found to have inadequate gauge data (Hossain et al., 2006). The inadequate station data prevents real-time flood monitoring as well as loss calculations. Monitoring floods are gradually becoming more challenging in medium and large river basins because of the gradual decline in in-situ rainfall networks throughout the world. The decrement in these conventional rain data sources has been identified recently as one of the central concerns hindering the advancement of hydrological modeling, especially around ungauged sites

(Shiklomanov et al., 2002; Stokstad, 1999). Ideally, for a detailed monitoring and analysis of these vulnerable sites, a dense rain gauge network is necessary (Collischonn et al., 2008; Tucci et al., 2008).

Figure 1.2 Annual Flood frequency over Southeast Asia [Reproduced from:(Yusuf et al., 2009)]. The legends show 0.762 (highest) and 0(lowest) flood frequency events per year from 1980 to 2001.

Estimation of the catchment-scale flood is one the major parts of applied hydrology in the calculation of risk and the design of hydraulic structures. For this, it is vital to have precipitation intensity, duration, and frequency data. The relation between rainfall Intensity(I), Duration(D), and Frequency(F) is represented by a family of IDF curves. IDF curves are widely used in engineering practice to determine the capacity of storm sewers. The evaluation of extreme events is also studied using IDF curves as they include most of the needed information.

As such, one of the main aims of this thesis is to construct present and future IDF curves for regions in the WMC. This thesis will demonstrate how to derive both current and future, as well as point and gridded IDF curves through results obtained

using dynamical downscaling. The IDF curves thus obtained would be helpful for policymakers in coming up with better climate change adaptation strategies.

1.4 Climate change impact on Hydrology

On considering the importance of water on environment and society, it is imperative to understand the response of the water cycle to climate change. It is observed that climate change affects both variability and availability of regional water resources. Both regional, as well as global climate simulations, under different scenarios are documented in the 2013 IPCC AR5 report. The IPCC panel observed that the climate change is likely to affect most of the world. Hydrological changes will affect both surface and groundwater that are primarily dependent on changes in temperature and precipitation patterns. Variations in these components of hydrology will likely alter the soil moisture, evapotranspiration, and other related hydrological components. A wide number of studies have been done to investigate the impact of climate change on hydrology which primarily comprises of three steps (Wetterhall et al., 2005; Xu et al., 2005):

- i. Use of GCMs to simulate climate response corresponding to the increment in Green House Gases (GHGs)
- ii. Implementation of downscaling (either statistical or downscaling) to bring GCM data to regional scale. The downscaled output has a scale that is compatible with the hydrological model and regional impact studies.
- iii. Development and use of hydrological applications to study the response of hydrological regimes at different scales corresponding to climate change.

One of the significant impacts of climate change over the MC would be its effect on water resources. The result would be thus being flooding as well as water scarcity problems occurring together. Further, extreme rainfall events would lead to an

excessive runoff on the fragile ecosystem of the region leading to floods, soil erosion, landslides.

Climate change is likely to result in a significant increase in intensity and frequency of extreme rainfall events as well as redistribution of rainfall within many regions. Therefore, it is of interest to study the past IDF curve of the region and then to model the future IDF curves using the IPCC scenarios to assess the changes in intensity and distribution. The IDF for past and future climate have been carried out in detail and presented in Chapter 4. Along with the IDF curves, it is also essential to assess the present distribution and grouping of rainfall over a region. Finding the present spatial and temporal clusters help to understand how the climate change is redistributing the rainfall. The cluster analysis for spatial and temporal distribution of rainfall over WMC is done in this Ph.D. study and is presented in Chapter 5.

1.5 Objective of the Ph.D. Work

There are three primary objectives of this Ph.D. work as described from in 1.5.1 to Section 1.5.3.

1.5.1 An Integrated Approach to Downscale Precipitation over Western Maritime Continent

Regional climate downscaling has developed quite extensively and has further scope for future improvement. A vast extent of research is being carried out throughout the research community to improve the quality of downscaled output predictands of interest such as rainfall, temperature, and the wind. The improvement is achieved either by finding a source of uncertainty and mitigating it specifically or by using novel approaches for downscaling or by using novel approaches for downscaling.

An accurate simulation of the precipitation, particularly in the tropical regions of the MC is very complicated. The inaccuracy is primarily because GCMs are run at a very coarse resolution, and hence the local climate variability such as from regional

topography are not adequately captured. Reanalysis data like GCM are also coarse and would ignore sub-grid and mesoscale local features. Thus, using GCMs or Reanalysis outputs directly in tropical regions can result in substantial uncertainty and biases in the downscaled predictands, specifically precipitation for the Ph.D. work here. The problems encountered by Reanalysis coarse grids can be resolved by downscaling the Reanalysis dynamically (physical downscaling) or statistically (empirical downscaling). Reanalysis inputs like the Reanalysis-II have a coarser resolution, and thus predictors extracted from Reanalysis-II cannot be directly used in statistical downscaling. A further downscaling of Reanalysis-II data can better capture mesoscale features when precipitation is to be downscaled.

As the typical GCMs outputs are very coarse when they are regressed along with station data, the downscaled model may fail to capture much of the regional weather. This partial failure to capture regional information makes the derived output unreliable for the impact studies. Therefore, to address this shortcoming, a substitute approach of using the RCM output in a statistical regression model is developed, and verification performed to assess the improvement in the case of precipitation. The study was carried out on four stations over the WMC. Daily precipitation output from statistical downscaling for these four stations was demonstrated and compared with statistically downscaled RCM and with Reanalysis predictors via the biases w.r.t observed station data. A comparison was made to compare mean and extreme statistics of rainfall for both integrated downscaling approach and pure statistical downscaling method. The methodology and results of this study are described in Chapter 3.

1.5.2 Intensity-Duration-Frequency Curves Under Changing Climate for Year Range 2030-2060:

IDF curves are widely used in different applications of hydrology. The IDF curve holds a paramount importance when a drainage system is designed. However, these IDF curves are based on rainfall stationary hypothesis as stated by (Kuo et al., 2015; Mailhot et al., 2007) that “the estimation of IDF curves rely on the assumption of rainfall series stationary, namely that intensities and frequency of extreme events remained unchanged with time.” However, the recent climatic change is significantly driven by global warming, and it is expected to modify extreme rainfall events and correspondingly, the existing IDF curves. A research objective here is to assess IDF under future climate conditions. The Community Earth System Model (CESM1) is one of the models of CMIP5 GCMs in IPCC AR5 (Monaghan et al., 2014) already used in many future downscaling studies (mostly statistical downscaling). Further, prediction of future climate (Miao et al., 2014; Yu et al., 2014) is utilized as an input boundary condition for WRF for future climate downscaling over the year 2030 to 2060. Verification is established by comparing the CESM1 CMIP5-WRF data outputs with CHIRPS dataset for the historical period. After that, the output precipitation from WRF model would be used for updating already existing IDF curves for different regions within the WMC.

This second objective of this Ph.D. work was thus achieved in two steps: firstly, an IDF curve was plotted for four study sites using the historical dynamically downscaled data. The IDF curves from the historical runs were compared against the IDF curves constructed from hourly data as extracted from Climate Forecast System Reanalysis (CFSR) dataset (Saha et al., 2010). CFSR data was chosen instead of other reanalysis datasets such as JRA-55, MERRA2, NCEP-2 and ERA-Interim as CFSR has very good temporal (hourly) and spatial (38 km) resolution. The resolution of the CFSR is thus much closer to the RCMs used in this study. In contrast other reanalysis products

have a spatial resolution in the range of 50 km to 200 km and temporal resolution in the range of 3hr to 6hr. The historical IDF achieved from WRF was bias corrected using the scaling method to match with the IDF curves derived from the CFSR data. In the second step, the IDF curve for the future emission scenarios which are defined by IPCC was developed. The IDF curves for future RCP scenarios (RCP 4.5 and RCP 8.5) were constructed using WRF, driven by CMIP5 data and using the CORDEX data (downloaded from <https://www.cordex.org/output.html>). The details regarding CORDEX data is discussed in Section 4.2.2. The IDF curves for RCP 4.5 and RCP 8.5 was then bias corrected using the historical IDF curves. The methodology and results of this study are described in chapter 4.

1.5.3 Cluster Analysis of Spatial and Temporal Monthly Precipitation over Western Maritime Continent

Precipitation over WMC is highly influenced by prominent convection occurring in the region. This region also has a relatively homogeneous air mass and evenly distributed pressure (Gabler et al., 2008) and sea surface temperature. Thus, the weather and climate in this region are governed heavily by local and mesoscale factors. The objective of this work is to identify the spatial and temporal patterns of monthly precipitation over this region, using Ward's clustering algorithm (Ward Jr, 1963). The objective is to relate the clustering patterns with other meteorological variables as well as topography and land use within the region. 37 year of monthly precipitation data over 43 evenly distributed locations over four major large islands: Java, Sumatra, Borneo, and Malaysia are used. Overall, four major spatial rainfall clusters were identified within the region using the Wards clustering algorithm. The methodology and results of this chapter are discussed in Chapter 5.

1.6 Outline of the Thesis

The focus of the Ph.D. work is to use dynamical downscaling specifically the WRF RCM over WMC to improve statistical downscaling of precipitation, study the impact of climate change on precipitation and lastly understanding different spatiotemporal precipitation clusters or distribution. The outline of the thesis is as follows

Chapter 2 is a literature review on the MC covering dynamical downscaling, statistical downscaling, Numerical Weather Prediction (NWP) and climate scenarios. Overall, the review covers the classification and requirement of downscaling methods, construction of IDF curves for present and future climate RCP scenarios, and the past works related to spatial and temporal clustering of rainfall.

Chapter 3 presents the ability of the integrated downscaling framework using the WRF combined with the Automated Statistical Downscaling Model (ASD) over WMC. Results are presented from both climatological aspects at the regional WRF output level and statistically at the station level after ASD.

Chapter 4 presents the results for present and future for point IDF curves at three stations: Singapore, Jakarta and Kuala Lumpur as a typical location in the WMC. The station IDF approach was developed using CFSR data, and bias-corrected the historical station data via the scaling method. Point IDF curves for the future were constructed for these three locations. Changes in future IDF curves as predicted by WRF is driven by the CMIP5 GCM are presented.

Chapter 5 presents the results of clustering analysis of spatial and temporal monthly precipitation over the WMC. The study relates the four clustering patterns found with other meteorological variables as well as topography and land use within the region. This chapter also presents the change in spatial and temporal groupings of monthly

precipitation induced because of climate change. Chapter 6 summarizes the main finding of this study as well as proposed future work for all the three studies.

2 LITERATURE REVIEW

2.1 Introduction

As mentioned in Chapter 1, this research focuses on the use of dynamical downscaling over the WMC to improve statistical downscaling, study the change and variability of precipitation via IDF curves and analyze spatiotemporal grouping of rainfall via CA. This chapter, therefore, reviews the past related studies that have applied these methods to monitor climate change effects. The drawbacks of coarse resolution GCM as used for regional climate modeling was discussed in the introduction chapter. Therefore, it is of importance to review and understand the “additional information” that an RCM imparts to various variables. Section 2.3 and Section 2.4 discuss the general climate system over MC and NWP principles respectively. The use of RCM in hydrology is discussed in Section 2.5 along with case studies where the same has been successfully done. Section 2.6 discusses the impact of climate change on hydrological cycles and overview of AR5 RCP scenarios. Lastly, Section 2.7 introduces the concept of downscaling and explains the two main categories of downscaling, i.e., dynamical and statistical.

2.2 Maritime Continent

The MC ranges from 10°S–20°N and 90°–150°E (Figure 2.1) as initially defined by Ramage (1968). The MC consist of many islands and mountains with an elevated topography and orography. The MC region is termed the “boiler box” of the tropical region since they hold the warmest part of the ocean (Chang, 2005; Neale et al., 2003; Ramage, 1968). RCMs forced by GCMs or Reanalysis data have been extensively used over the world. However, there are only a few downscaling studies reported over MC (Rahmat et al., 2014) and thus the MC is under-represented in the downscaling literature. Given its importance for the global climate, and that a significant fraction

of the world's population lives in this region, it is of interest to dynamically and statistically downscaled global climate data over the MC.

Figure 2.1 Plot showing the MC: 10°S–20°N and 90°–150°E.

Convection is one of the most dominant meteorological processes in the MC, and most of the precipitation over the MC is from rainfall resulting from convective clouds. The complex convection process forming these convective clouds are associated with the interaction between diurnal heating cycle, the islands topography and predominant large-scale circulation (Keenan et al., 1989).

According to Ramage (1968), convective process over the MC is dominated by interaction amongst the diurnal heating, the local topography of the islands and synoptic circulation. These interactions lead to the formation of island thunderstorms which are rampant in the tropical region and especially over the MC. These convective storms play a crucial role in tropical dynamics and resulting in a large proportion of rainfall (Williams et al., 1987).

Many studies have been conducted to verify the extent of diurnal activity in MC using satellite images (Holland et al., 1980). It is observed that the effect of convection on the diurnal timescales is variable between land and sea over the region. Studies have shown that the complex surface topography throughout MC (Hock et al., 2005; Qian, 2008; Wang et al., 2007; Wu et al., 2009) introduces inhomogeneity which further complicates the diurnal forcing of convection. When it is winter in the Northern hemisphere, the diurnal amplitude of convection related precipitation over land can increase to about three times to that of rainfall in the adjacent oceans (Kraus, 1963; Neale et al., 2003). The increase is because the land has a smaller capacity to hold heat in the northern region leading to low-level instability in the large diurnal variations (Yang et al., 2001). However, there also exist some further changes in the response. The diurnal cycle of offshore propagation of convection over the MC is related to high diurnal heating of the land, linked land-sea breeze circulations and gravity wave propagation. As mentioned by Love *et al.* (2011), there is significant variability over the islands influenced by the gravity waves. For example, spatial variation in heating over Sumatra throughout the daytime leads to an alteration in the makeup of the offshore gravity wave response. The variability in heating propagates to the oceans and consequently results in coherent changes where the convection peak (maximum convection rate within a day) is reached. Most of these variations are related to a standard evolutionary life of a convection system, which forms and evolves over land and moves to the ocean bodies.

The influence of inter-seasonal activities is very commonly seen in the MC. For example, circulations like the Madden –Julian Oscillations (MJO) travel over this part of the globe before reaching the West Pacific in its developed stage. Further, the inhomogeneity of the underlying surface diurnal forcing results in periodic modulations of MJO which then disturbs the diurnal variations in MC. The MC is also found to be responsible for modulating the phase speed and strength of the MJO by

inducing and splitting the active phase before the MJO rejuvenates while moving to the South Pacific (Sui et al., 1992; Zhu et al., 1993). Thus, a definite impact of MCs over global climate is undeniable.

Because of the direct and indirect influence of climate anomalies such as MJO and ENSO over the MC on the world climate, there is a need to study climate and weather with sufficient accuracy over the MC. However, the most readily available data to carry out these studies are from GCMs, which are run at a very coarse resolution and therefore do not adequately capture the local climate variability such as regional topography, and such variability is an integral constituent in determining the weather over the MC. Hence, using these GCMs outputs in the tropical region can result in substantial uncertainty and biases in the downscaled predictands of interest, such as rainfall. The problems encountered by GCMs coarse grids can be resolved by downscaling the GCMs output dynamically using RCMs (physical or dynamical downscaling) or statistically (empirical downscaling).

2.3 General Climate Over Maritime Continent

GCMs and RCMs simulate the global and regional mean climate conditions respectively. It is essential to understand the observed mean climate over the Maritime regions before using these models. In this section, global datasets from past studies as used to describe the seasonal climate over MC is reviewed. Observational studies reveal a variability in seasonal climate over both temporal and spatial dimensions over the geographically complex MC. The variabilities are the result of MC region's local complexities as well as its climate sensitivity to remote effects. These challenges are faced when choosing a correct model that can capture the variability in climate over the region.

The climate of the Maritime region is categorized by its heavy rainfall throughout the year. Further, the regions in the MC also show a strong seasonal variation and cycle

associated with wind regimes. The seasonal daily mean rainfall for DJF, MAM, JJA, SON derived by the Climate Prediction Centre (CPC) (Development; Fan et al., 2004) is shown in Figure 2.2. Further, Figure 2.3 shows the zonal winds at 925 hPa derived from 40 years of ECMWF data (Strachan, 2007). It can be seen in Figure 2.2 that a wet season with rainfall intensity up to 14 mm/day with prevailing westerly winds of 6 m/sec is experienced over most of Indonesia and Southeast Asia over the boreal winters that coincide with the Australia-Asian monsoon. The MC is particularly more conducive to orographic storms or rainfall over the mountains and in the coasts exposed to South China Sea northeast wind flow.

Figure 2.2 Mean seasonal daily precipitation (mm/day) plotted from Climate Prediction Centre (CPC) Merged Analysis of Precipitation (CMAP) climatology (a) DJF (b) MAM (c) JJA (d) SON [Reproduced from (Strachan, 2007)]

Figure 2.3 shows strong easterlies were persistent for boreal winters (DJF) over the western Pacific. A characteristic westerly wind is observed over the eastern Indian Ocean, across the MC and into the South Pacific Convergence Zone (SPCZ), this region is observed to have maximum convection in Maritime region. Although, a

persistent westerly wind is seen for boreal autumn, spring and winter; a considerable variability was observed in the sign of 925 hPa zonal winds across the region. The variability of the wind was evident at smaller islands.

Figure 2.3 Seasonal wind m/sec (at 925 hPa) derived from ECMWF data. (a) DJF (b) MAM (c) JJA (d) SON. The color shows the magnitude of the winds while arrow shows the direction of the zonal winds [Reproduced from (Strachan, 2007)]

A comparatively dry season with prevailing easterly wind is experienced by the Maritime continent during boreal summers. A peak in the dry season was observed during August. Over the region, the transition period of monsoon was observed in SON and MAM (see Figure 2.4). However, a mean rainfall of 10 mm/day was observed even during SON and MAM. A wind reversal in the transition period is seen over the region which occurs rapidly in MAM

compared to SON. The anomaly in the reversal of inter-sessional winds can be (see Figure 2.4) that also shows a high SST over Maritime continent from August to February.

Figure 2.4 Monthly mean rainfall from TRMM with sea surface temperature (°C) superimposed onto Quick Scatterometer (QuickSCAT) wind (mm/sec) for August(top) and February (bottom). Data average from 1998 to 2003 and 2000 to 2004 respectively [Reproduced from (Qu et al., 1994)]

2.4 Numerical Weather Prediction Principles

NWP is the process wherein fundamental equations are solved to obtain the solution for weather. Both GCM and RCM models are primarily implemented via Computational Fluid Dynamics (CFD) solution techniques. The framework of NWP models is represented in Figure 2.5. The four primitive equations used in NWP models are (Stocker, 2014):

1. The geostrophic momentum equations

$$\frac{Du}{Dt} - fv = -\frac{\partial \phi}{\partial x} \quad (2.1)$$

$$\frac{Dv}{Dt} + fu = -\frac{\partial \phi}{\partial y} \quad (2.2)$$

2. The hydrostatic equation

$$0 = \frac{\partial \phi}{\partial p} - \frac{RT}{p} \quad (2.3)$$

3. The continuity equation

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial p} = 0 \quad (2.4)$$

4. The thermodynamic energy equation

$$u \frac{\partial T}{\partial x} + v \frac{\partial T}{\partial y} + w \left(\frac{\partial T}{\partial p} + \frac{RT}{p c_p} \right) \frac{\partial w}{\partial p} = \frac{J}{c_p} \quad (2.5)$$

In these equations, the symbols denote the following:

- **u** is the zonal velocity (velocity in the east/west direction tangent to the sphere).
- **v** is the meridional velocity (velocity in the north/south direction tangent to the sphere).
- **w** is the vertical velocity
- **T** is the temperature
- **φ** is the geopotential
- **f** is the term corresponding to the Coriolis force and is equal to $2\Omega \sin(\phi)$, where Ω is the angular rotation rate of the Earth ($2\pi / 24$ radians/hour), and ϕ is the latitude.
- **R** is the gas constant
- **p** is the pressure
- **c_p** is the specific heat
- **J** is the heat flow per unit time per unit mass

Figure 2.5 A framework representation of NWP [Reproduced from <http://public.homepages.uconn.edu/~ria08001/ceml/ceml-both.html>]

Almost all the RCMs use some numerical scheme to solve the above equations. These models treat water differently than in hydrological models. As mentioned by Te Chow (1988) the property of water in the hydrological model does not change with a change in temperature (typically ambient range). Water is described using mass and momentum conservation laws. However, in the case of the atmospheric gases, compressibility is considered as the gas compressibility increases significantly with temperature. Therefore, along with mass and momentum equations to describe the motion, an additional conservation equation for thermal energy (as following the first law of thermodynamics) and the perfect gas law equation are required. The ideal gas law is needed to set up a link in between pressure, temperature, and volume (Williams et al., 1987).

2.4.1 Global Climate Models (GCMs)

GCMs are the numerical models that represent the physical processes in the ocean, atmosphere and cryosphere, and on the land surface. Numerical models such as GCMs are the most progressive tools presently available for modeling the response of the global climate system to various inputs such as an increment in greenhouse gas concentration. GCMs, sometimes integrated with RCMs, can provide consistent estimates geographically and physically for regional climate change required (Change, 2014; Hänsel et al., 2013; Moriondo et al., 2006).

Due to the scaling issue and limited knowledge of the atmospheric process and circulation, GCMs outputs are subject to significant uncertainties. The main uncertainties involved in long-term climate projections are model structure uncertainty, emission scenario uncertainty as well as uncertainty as a result of internal variability (Yip et al., 2011). There are considerable uncertainties in GCMs arising from the climate forcing scenarios. Another source of uncertainty comes from the initial condition used by the model. GCMs usually start from an arbitrary time in the pre-industrial era (approximately around the year 1850). It is noted that even simulation carried out using the same GCMs but with different initial times can sometimes result in drastically different in outputs for the future (<http://www.aip.org/history/climate/GCM.htm>).

2.4.2 RCM Models

RCMs can significantly improve on the GCM outputs. The improvement is seen especially regarding convective and non-convective precipitation phenomenon. The primary driving forces for precipitation events in the atmosphere are microscale or mesoscale whereas GCMs running at a coarser resolution fails to capture such significant local information. RCMs captures the microscale and mesoscale features

better than GCMs as they are comparatively operated at a much finer resolution and thus give better rainfall output (Skamarock et al., 2005).

The non-hydrostatic WRF Model which is used here utilizes the advanced research WRF dynamical core and is a fully compressible, primitive equation model. The model was built and then was continuously improved and developed by the National Centre for Atmospheric Research (NCAR), National Oceanic and Atmospheric Administration (NOAA) and many leading educational institutions and government organizations. Figure 2.6 shows the inner working with WRF and its utilities via a flowchart.

Figure 2.6 A flow chart summarizing components of WRF model [Reproduced from http://www2.mmm.ucar.edu/wrf/users/docs/user_guide/users_guide_chap1.html]

The WRF model allows the user to mix the dynamic cores along with different physics parameterization and hence improves the forecasts and downscaling both regarding

accuracy of output and optimization in time. WRF model is commonly used for atmospheric research, operational forecasting and dynamic downscaling of GCMs data (Shan, 2011).

2.5 “Addition information” Added by RCM

This section reviews several past examples works related to regional climate modeling that illustrate the RCM provided “additional value” to the input data. As discussed in Chapter 0, RCM uses fine-scale topographic, discretization and advanced atmospheric equations and parameterization to output a finer-scale climate output. Figure 2.7 shows that an RCM of resolution 50 km can capture the topography better than a GCM of resolution 300km. As reproduced from a study done for Prediction of Regional scenarios and Uncertainty for Defining European Climate change risk and Effects (PRUDENCE) project (Christensen, 2001), Figure 2.7 makes it evident that GCMs fail to adequately capture the local topography details that have a prominent effect on the local climate. However, a 50 km RCM does a reasonably good job in capturing local topographic details, though there might be a need to go to a higher resolution to resolve steep topography. The PRUDENCE is one of the first experiments done where the RCMs added information contribution was realized.

Figure 2.7 A figure showing topographic details as resolved by RCM-50km(left) and GCM-300km(right) and [Reproduced from Christensen O.B., 2001]

In another study done by Jones et al., (2004) over Britain, winter precipitation comparisons were made amongst RCM, GCM and observed data. The simulated precipitation along with observations is shown in Figure 2.8. The study implemented the GCM HadCM3 model with HadRM3P as the RCM. Both GCM HadCM3 and RCM HadRM3P used in the study were developed at the Hadley Centre, UK.

Figure 2.8 shows that observation data shared a close similarity with the RCM results whereas a noteworthy difference with the GCM output was seen. The GCM failed to capture high precipitation over high elevations. GCM results showed an overall north-south precipitation gradient but missed the detailed distribution of rainfall. In contrast, RCMs captured the northwest as well as the east precipitation gradient. One of the major reasons behind better simulation results from RCM was its ability to capture fine scale topography details.

Figure 2.8 Simulated precipitation by GCM and RCM over Britain with observation [Reproduced from Jones et al., 2004]

Apart from rainfall, RCMs are also seen to have simulated temperature better than GCM models. A study was done using the parallel climate model, Mesoscale Model version 5 (MM5) developed by the National Center for Atmospheric Research (NCAR). Leung et al. (2004) studied temperature changes for the year 2040-2060 over the Rocky Mountains. The study found that GCM did not correctly resolve temperature changes over coastal mountain ranges. The results suggested that GCMs were not able to adequately capture the snow-albedo feedback effect that was responsible for the additional warming of 1°C (Figure 2.9). In contrast, the MM5 model could capture the feedback mechanism considerably well and thus showed the expected warming.

Figure 2.9 (a) GCM (PCM) and(b)) RCM (MM5) simulations over the Rocky Mountains for temperature change over the Rocky Mountains [Reproduced from Leung et al., 2004]

2.6 Climate Change Impacts

As greenhouse gas levels are increasing and are the primary focal point for framing the global climate debate, one of the significant effects of climate changes has always been the potential change in hydrological cycles and water supplies. The anthropogenic activities, directly and indirectly, affect the water distribution. Change in temperature and precipitation as arising from climate change is one of the leading causes of water vulnerability in some regions, and floods or excess water in other parts (Barnett et al., 2008; Vörösmarty et al., 2000). There has been a redistribution of rainfall because of climate changes. However, the change in precipitation was quite uncertain, i.e., RCMs predicted that the rain increased during the summer over the USA and the Gulf coast of Mexico because of intense rainfall and storms. However, there was a decrease in the runoff because the increase in temperature also resulted in an increase in evapotranspiration over most of the regions in the USA (Mulholland et al., 1997).

Also, Lenderink et al. (2008) indicated that the GCMs outputs showed that every 1-degree change in global temperature resulted in an increase in precipitation by 7% on a globally averaged basis. The projected changes in rainfall intensities will lead to increased risks and environmental destruction. These destructions include excess soil erosion, growth in cost and revision of existing urban design for infrastructure and drainage and damaged coastal ecosystems (Greaver et al., 2010). Molénat et al. (2013) explicitly mentioned in their research that there is a need to revise existing rainfall Intensity-Duration-Frequency (IDF) curves in many regions. The revision of IDF curves is done because of changing the non-stationary characteristics of the rainfall under changing climatic conditions.

Some of the recent rainfall studies that implemented climate models over the WMC were reported by the Center for Environmental Sensing and Modeling (CENSEM) of the Singapore-MIT Alliance for Research and Technology (SMART). An initial study that analyzed historical (1986–2005) climate simulations using a ten-member subset of the Coupled Model Inter-comparison Project Phase 5 (CMIP5) global climate models over Southeast Asia (SEA) was reported by Srivatsan et al. (2016). This study evaluated the subset of GCM performance in simulating the mean state of climate over the less-studied SEA region, with a focus on seasonal precipitation via a comparison with TRMM observations supplemented with GSMaP data. The results indicated that while most of the models are unable to reproduce the observed annual and seasonal precipitation due to large model uncertainties, a multi-model ensemble approach was found to represent observations better. The study emphasized on the need for the next generation of models to simulate regional climates with greater accuracy for impact studies. In a related study Gianotti et al. (2013a, b) developed a new method for parameterizing convective cloud fraction which was applied in the regional climate model RegCM3 driven using historical ECMWF data. The convective parameterization simulated the essential convective–radiative feedback,

absent in existing convective schemes. The new convection scheme showed improvement in model performance, not only concerning convective rainfall (when compared with TRMM 3B42 data) but also in shortwave radiation (when compared with Surface Radiation Budget dataset). Another study by researchers at TMSI and CENSEM (Liew et al. (2014)) provided a 3-step methodology to derive IDF curves for ungauged sites using WRF driven with ERA-40 reanalysis and CCSM 3.0 GCM data over the SEA. The study reported that WRF/CCSM under future SRES-A2 forcing for would result in an increment in maximum rainfall intensity by 70% relative to present at 24hr duration for 50yr return period in the year 2100. A further study by from the CENSEM group (Im et al. 2015) assessed the regional climate projection within the national downscaling project framework of South Korea. This study aimed to obtain fine-scale climate information (12.5 km) by dynamical downscaling the HadGEM2-AO global projections forced by Representative Concentration Pathways (RCP) RCP4.5 and RCP8.5 scenarios using the WRF model. The study reported that a shift in the distribution that directly affected the magnitude of extremes with enhanced extreme hot days but reduced extreme cold days. Changes in return levels of annual maximum precipitation suggested an increased probability of extreme precipitation with 20 yr and 50 yr return period

The RCPs was chosen to highlight foundation behind their use. RCPs are referred as a pathway to emphasize that their main purpose is to provide a time-dependent concentration of atmospheric greenhouse gas. The scenarios that were used in the first IPCC report did not include the rise in greenhouse gases from natural causes, and only human sources were considered. In the AR5 2013 report, the four new scenarios termed as RCPs are proposed. Their classification is dependent on the net relative radiative forcing in the year 1770 to 2100 specifically 2.6 W is for RCP 2.6, 4.5 W for RCP 4.5, 6 W for RCP 6 and 8.5 W for RCP 8.5. However, these values of RCPs should be considered as indicative only because the values of RCPs vary and as

dependent on the model used to generate them. Further, different models have different characteristics and treat short-term climate forcing differently (Alexander *et al.*, 2013).

RCP 2.6, is considered the least forcing methods by greenhouse gas concentration while RCP 4.5 and RCP 6 are also known as stabilization scenarios. The worst-case scenario is RCP 8.5 where the forcing of greenhouse gases is very high. These RCPs can thus be used as representative of 21st-century climate policies, unlike the Special Report Emission Scenarios (SRES) given in IPCC's fourth assessment scenarios. The IPCC fourth assessment scenarios further had no climate policy for controlling the emission of greenhouse gases.

It is noted that the SRES scenarios are based on socio-economic futures; this setup requires each scenario (emission scenario, radiative forcing scenario, climate model scenario) to be formed in sequence with the previous scenario. Thus, if there is a change in any previous scenario, the whole sequence must be restarted. This makes the sequential approach time-consuming. However, RCP Scenarios are based on radiative forcing projections; this allows for socio-economic, emissions and climate scenarios to be developed in parallel with each other. This way, changes can be made to one individual scenario without having to restart the whole sequence.

RCPs consider climate change mitigation policies to bound emissions. The scenarios are named after the radiative forcing relative to the pre-industrial period achieved by the year 2100, or at stabilization after 2100. Changes in land use and air pollution emission are considered sector wise. These factors specified anthropogenic emission and annual greenhouse concentration during the year 2100 and used in the amalgamated carbon cycle models, atmospheric chemistry, and climate models. However, RCP's do not consider the aerosol emissions.

In AR5, the GCM Coupled Model Intercomparing Project Phase 5 (CMIP5) and Earth system model runs were carried out with concentration reaching to the following

values by the year 2100. The corresponding results from the model run are shown in Figure 2.10 (Alexander et al., 2013).

Figure 2.10 Global Average Surface change under different scenarios. Reproduced from: (Alexander et al., 2013)

2.7 Climate Downscaling

As per IPCC (<http://www.ipcc.ch/ipccreports/tar/wg1/377.html>), the global scale coarse information is less informative than the regional level fine-scale data. The IPCC also indicated that when downscaling is done, some additional value and information is added to the data, which is being downscaled. The downscaling technique is of great importance when the requirement is to model precipitation data over a complex terrain and orography or include temperature effect from local atmospheric structures or for a study on mesoscale convection. The downscaling has also been found to be useful to inculcate effect of the sea and land breezes on the local circulations during downscaling (Alexander *et al.*, 2013).

Precipitation is often much localized to a region and time, and NWP's are run at resolution 5km to 30 km. This forecast is inadequate, especially when NWP's ability to correctly capture convective phenomenon is involved. Hence, it is advisable to add information (observed station data) to the NWP forecast. Adding more details improves the quality of downscaled and forecasted precipitation outputs (Kidson et al., 1998; Micheas et al., 2007). Data assimilation techniques are one such method to introduce such additional information.

Micheas *et al.* (2007), reported the use of advection routines for forecasting precipitation using radar data as input and found that hindsight forecast for rainfall was very accurate. However, the forecast results were limited to the short time range of seven days. An alternate method is using statistical methods, i.e., by calibrating a model using statistical relationships between rainfall and the meteorological variables affecting it (Altava-Ortiz et al., 2006). Some of the methods to carry out this are linear regression models, Kalman filters, neural networks and techniques for classification analogs (Jones et al., 2004; Wilby et al., 2004; Wilby et al., 2013).

The principle of the analog method is to use the past; long time interval observed data for rainfall for several stations with the numerical weather forecast outputs for the corresponding time interval. Most of the times as the rainfall follows a nonlinear pattern; a neural network approach can be useful because of the capability of neural networks to handle nonlinear problems. There are many reported works done as mentioned by Kuligowski et al. (1998) for forecasting wind and precipitation with these forecasts are within the time range of 6 hours.

There are two fundamental ways of performing downscaling:

- Dynamic Downscaling
- Statistical Downscaling

2.7.1 Dynamic Downscaling

Dynamical downscaling is a method of using an RCM model to bring coarse resolution GCM outputs to a finer resolution in the dimensions of time and space. A finer nest is built inside a coarser resolution input data using the RCM. The fundamental working principle of the RCM is to solve the governing equations of atmospheric processes using numerical methods, with boundary and initial conditions provided by GCM output directly. Giorgi (1990) was the first to use this method since then many research studies have been carried out on Regional Climate Models (RCMs) because of their unique advantage (Skamarock et al., 2005; Wilby et al., 2004).

The grid spacing of RCMs ranges from the medium (20-25km) to fine (5-10km) depending on the aims and objectives of the study and computational resources available. However, it is widely debatable which resolution gives the best outputs. There are reported circumstances (Gilliland et al., 2007; Madala et al., 2014; Zhang et al., 2011) where large grid sizes sufficiently captured the large-scale and sub-grid atmospheric processes via cumulus parameterization. However, in general, fine grid scale proved to be much better, because it captures microphysics processes correctly. The microphysics process has a crucial role in determining the quality of output variables such as precipitation.

The use of proper physical parameterization is the main challenge when using an RCM model. The physical parameterization is set by the location, i.e., location specific (Christensen et al., 2001; Skamarock et al., 2005). Further, it is important to use appropriate boundary conditions such as soil moisture that is utilized by these models, and these conditions have the greatest role in shaping the output of the RCM. The biggest advantage of RCMs models is that they can resolve the mesoscale and small-scale features such as orographic precipitation and low-level jets better than GCMs models. In addition to that, RCMs can also be used to study climate by

applying different external forcings. These forcings include changes in land use or changes in atmospheric chemistry via implementing different scenarios. Even future changes can be studied using RCM (Murphy, 1999; R.L. Wilby a & 2002; Wilby *et al.*, 2002) by applying future scenarios of the GCM runs.

The atmospheric models such as GCMs simulate physical processes by applying the governing equation the responsible motion of fluids. However, this way of simulating is feasible only where the phenomena are either mesoscale or at large scale. However, cloud formation is a much complex process and is dependent on relatively small spatial scale. The cumulus cloud formation, which is one of the primary reasons for thunderstorms, occurs at 0.1 -10 km grid resolution (Nakamura *et al.*, 2008). It will be highly impractical to represent such small-scale model with a low resolution of 25-100 km. The cumulus parameterization is to amalgamate the collective processes for cloud formation covering radiation budget and rainfall within a single grid and thus simulate the convective part of the precipitation. However, it should be noted that the primary reason for the cumulus scheme is to make the model stable rather than simulating rainfall (Willems, 2012).

On the other hand, if the model runs at a smaller grid (less than 5 km) resolution such as in RCM, the cumulus process is represented explicitly by the microphysics scheme, and the precipitation output is “explicit” in WRF. This rainfall is commonly called as explicit precipitation. However, it is recommended to have a cumulus scheme for the model running at a resolution higher than 10 km to ensure model stability and to achieve reasonable rainfall. There is a common consensus that models are running with a grid resolution lesser than 3km does not require cumulus scheme and can simulate the explicit precipitation well (Willems, 2012).

2.7.2 Statistical Downscaling Method (SDSM) and Automated Downscaling Method (ADM)

The principal objective of statistical downscaling is to develop a quantitative relationship between the local observed variables (predictand) and the GCMs large-scale variables (predictors). The GCM variables represent a larger grid and capture the synoptic phenomenon of a broader region. These relationships are set by linear or nonlinear regression models to provide the link between the large-scale atmospheric processes and the local predictand.

Some of the commonly known regression models are canonical correlation analysis (Storch, 1995), multiple regression methods (Murphy, 1999) and artificial neural networks which are based on nonlinear regression. These models have almost equal variance when compared to the observed data. However, the problem in the prediction of daily precipitation is that the large-scale atmospheric forcing has low predictability for local precipitation amounts. Precipitation is one of the most complex phenomena in the atmosphere that is dependent on several another forcing. Table 2.1 highlights the strength and weaknesses of various downscaling approaches. The regression methods are more commonly used because of their ease, and low computational cost (Wilby R. L. & Dawson, 2013).

Table 2.1 Table summarizing different statistical downscaling methods and tools used in downscaling literature (Wilby R., Charles, et al., 2004; Wilby R., Charles, S., Zorita, E., Timbal, B., Whetton, P., Mearns, L., 2004).

Methods	Strengths	Weaknesses
Weather Typing (example: analog method, hybrid approaches, fuzzy classification, self-organizing maps, Monte Carlo methods)	<ul style="list-style-type: none"> The outputs have interpretable linkage to the surface climate They are much versatile; e.g., can be applied to flooding, erosions, air quality. It involves the analysis of the extreme events. 	<ul style="list-style-type: none"> It requires an additional task for the weather classification Circulation based schemes, and can be insensitive to the forcing of future climate The surface climate for the intra type variation may not be captured.
Weather Generators (e.g., Markov chains, stochastic models, spell lengths methods, storm arrival times, mixture modeling)	<ul style="list-style-type: none"> It produces the large assemblies for uncertainties analysis or long simulation for extremes The spatial interpolation of model parameters is done, in other words; a continuous surface is created from a set of points It can generate sub-daily information. 	<ul style="list-style-type: none"> It makes an arbitrary adjustment of parameters for the future climate. It also creates an unanticipated effect on dependent variables of changing precipitation parameters.
Regression Models (example linear regression, neural networks, canonical correlation analysis, kriging)	<ul style="list-style-type: none"> It is relatively straightforward to apply when compared to other two approaches It employs the full range of available predictor variables “Off the shelf,” i.e., in the form of ready to be used solution and software available 	<ul style="list-style-type: none"> It has a poor representation of the observed variance. These models may assume linearity and normality of the data. It has a comparatively weak representation of the extreme events.

To overcome problems associated with precipitation downscaling several other methods were also used, e.g., weather pattern methods, stochastic weather generators, and limited area regional climate models (Hessami *et al.*, 2008). Bürger (1996) came up with a novel approach of an “expanded downscaling” and thus tried to increase the amount of variance in the predictor series simulated. There are some multi-site models

based on a regression that makes use of stochastic processes to portray the unexplained variance (Wilby *et al.*, 2004).

There are various domains in which statistical downscaling approach is found to be very appropriate and valuable. As mentioned in AR5 (2013) IPCC assessment reports “Research on Statistical Downscaling Model (SDSM) has shown an extensive growth in application and includes an increased availability of generic tools for the impact community” (Christensen *et al.*, 2007). For example, SDSM was tested to reproduce the shift in observed climate, thus demonstrating its ability to predict future change in climate. Another study was done by Timbal B. *et al.* (2009) on the observed rainfall and decline in rainfall for Western Australia in the late 1960’s and the SE Australia for mid-1990. It was evident from such studies that SDSMs can propagate the non-stationary climate conditions. A detailed description of using downscaled SDSM outputs can be found in guidelines on statistical downscaling practices (Wilby *et al.*, 2004) as recommended by IPCC, AR4.

As evident from the above studies, linear regression is one of the standard ways to carry on the statistical downscaling. The SDSM was one of the pioneer models that used this method (Wilby *et al.*, 2002). SDSM provides an advanced and single site scenario of daily weather variable under both presents as well as future forcing conditions in a computationally economical way. This method was first applied to model daily temperature and precipitation scenarios in Toronto, Canada for the period 2040-2069 (Wilby *et al.*, 2002). On the other hand, “weather generator” models such as WGEN and LARS-WG have widely used for hydrological as well as an agricultural research community.

Table 2.2 summarizes the advantages and disadvantages of two basic approaches of downscaling, i.e., dynamic and statistical. It can be seen in Table 2.2 that both the downscaling methods have their pros and cons. The advantage of statistical

downscaling is fundamentally related to its ease, and less computational requirement whereas dynamical downscaling is more robust and has physics backing.

Table 2.2 Pros and cons of statistical and dynamical downscaling (Wilby R. L. et al., 2002).

	Statistical Downscaling	Dynamic Downscaling
ADVANTAGES	<ul style="list-style-type: none"> • It can generate station or point data output from global climate model output • It is very efficient regarding computational resources requirement and time is taken. • Ensembles are given the climate change IPCC scenario and allow examining risks and uncertainty assessment. • It is comparatively more flexible than dynamic downscaling. 	<ul style="list-style-type: none"> • It outputs a grid level output of resolution 10-50km by taking GCMs outputs as input • The response is dependent on the external forcing. • Resolves the local processes effects because of orography. • It has good consistency with GCMs data
Disadvantages	<ul style="list-style-type: none"> • Depends on how realistic the GCMs boundary forcing is. • It is highly dependent on the size of domain location. • To calibrate the data, station data is required. • The -stationary relationship in between predictor and Predictand. • As the transfer scheme used is empirical, even that can affect the rainfall • Even a variability in low frequency can cause problems. 	<ul style="list-style-type: none"> • Dynamical downscaling is highly dependent on how realistic the boundary forcing is. • Highly sensitive to the size of domain and location • Computing resources required are very high. • Sensitive to the initial condition • Precipitation outputs are highly dependent on the choice of the convective scheme • A new physical parameterization is limited to a region and cannot be transferred to a new region readily.

3 An Integrated Dynamical and Statistical Approach to Downscale Precipitation over Western Maritime Continent

3.1 Introduction

Information extracted from high-quality weather and climate data is required by policymakers, and stakeholders to make right and informed mitigation decisions (Mearns et al., 1999; Fowler et al., 2007a; Bronstert et al., 2007). Rainfall is an important climate variable as the extreme events of rainfall can lead to natural disasters such as floods and droughts. An incomplete or a poor information on rainfall can result in loss of life and property, while a quality and accurate information on rainfall help to minimize loss resulting from floods and droughts (Wang and Zhou, 2005). However, most of the regions in the world lack sufficient rainfall records and hence a modeling approach is used to simulate rainfall. However, modeling precipitation has been found to be one of the most challenging problems in many research studies (McCarthy, 2001; Fowler et al., 2007b; Frost, 2007). The difficulty in participation modeling arises because of complex atmospheric circulation and process that estimates precipitation intensity and occurrence. The complexity of atmospheric processes acerbates in the convection dominant tropical region and hence modeling precipitation is more challenging over tropics (Xiong et al., 2006; Houze Jr, 1997; Gettelman et al., 2002).

WMC is an important part of tropics as it encompasses several important cities and receives high solar heating. The problem faced in modeling precipitation over tropics is also prominent over WMC (e.g., Ramage, 1968; Neale and Slingo, 2003). Solving the rainfall modeling issue over WMC is very important as it is densely populated and a large part of the population faces losses resulting from extreme precipitation events.

Some examples of the loss as resulting from extreme events over WMC includes: Thailand flood in year 2011 resulting in 54 deaths, and Labuan flood in the year 2014 that led to 815 deaths and displaced around 1000 people (Brinkman and Hartman, 2008; Osti and Nakasu, 2014; Gale and Saunders, 2013).

A modeling approach to generate daily station data with better reliability is demonstrated in this chapter via using both DD and SD. To avail the benefit of both DD and SD, an integrated downscaling framework is developed. Thus, the integrated downscaling approach broadens the scope of both DD and SD as it incorporates regional scale predictors extracted from DD and generates station level rainfall with SD. When the two downscaling methods are integrated together, the method becomes computationally less expensive compared to solely running DD to a very high resolution. In this approach, WRF was used as an RCM to implement DD, while SD was achieved using ASD model. The WRF model was driven using Reanalysis-2 (R-2) data while the predictors for ASD were provided from the DD output of WRF. The SD method was further refined by using “white-box” decision tree approach for selection of predictors (Goyal et al., 2012; Maraun et al., 2015).

The integrated approach in this study is presented for precipitation downscaling at four locations within WMC (see

Table 3.1). The four locations are Labuan, Malaysia (5.32°N, 115.21°E), Singapore (1.35° N, 103.98°E), Hat Yai, Thailand (7.01°N, 100.46°E) and Jakarta, Indonesia (6.17°S, 106.82°E). The downscaling time period in this study was equal to available station data, i.e., 29 years (January 1980-December 2008). These four locations are chosen partly due to needed long-term precipitation data being available. These four locations are on four different R2 grids which allowed a regional evaluation of the integrated approach. Lastly, these stations lie in economically important countries in WMC, which needs good quality precipitation data for mitigation of flood damage

Table 3.1 Station location and data interval used for calibration and validation

Location	Latitude	Longitude	Calibration period	Validation period	Missing Data (%)
Singapore	1.35° N	103.98° E	1980 – 1998	1999 - 2008	0
Labuan (MY)	5.32° N	115.21° E	1980 – 1998	1989 - 2008	10.8
Hat Yai (TH)	7.01° N	100.46° E	1980 – 1998	1989 - 2008	13.9
Jakarta (ID)	6.17° S	106.82° E	1984 – 1998	1989 - 2006	0.89

The evaluation of the integrated approach is conducted by comparing predicted rainfall with observations using various metrics covering ensemble quality and mean monthly rainfall statistics covering mean, standard deviation (i.e., variability), the number of wet days and excess kurtosis (i.e., volatility). Section 3.2 of this chapter presents a model description and implementation. The WRF model and its application are described in Section 3.2.1. Section 3.3.1 provides further verification of the suitability of WRF output variables as predictors in ASD. This is done via spatial correlations between these variables as extracted from the CFSR dataset with the TRMM measured rainfall. Improvements obtained in the downscaled ensemble rainfall at the four stations are lastly described in Section 3.3.3. The evaluation of the integrated approach is conducted by comparing predicted rainfall with observations using various metrics covering ensemble quality and mean monthly rainfall statistics covering mean, standard deviation (i.e., variability), the number of wet days and excess kurtosis (i.e., volatility).

3.2 Model description and implementation

Figure 3.1 shows overall methodology of the integrated downscaling approach (left) as compared to a simple SD approach (right). While SD is common in both methods, different input predictors are used. The integrated approach involves the additional

step of DD using an RCM which further allows an analysis of the downscaled regional physics and precipitation patterns. The DD is achieved using WRF which requires a correct physics parameterization as described below.

Figure 3.1 Flow chart of integrated approach (left) and conventional SD method (right)

3.2.1 WRF Model

The WRF model (Skamarock et al., 2008) was used to dynamically downscale the R2 data over the study region shown in Figure 4.5 to a 50-km resolution grid. The configuration and parameterization of WRF model for the integrated downscaling study are described in this section.

Nudging is used as is typically performed when using WRF. The nudging can be analyzed or spectral. In both cases, the horizontal winds, as well as potential temperature perturbation, were relaxed towards a reference state (Fonseca et al. 2015). However, the vapor mixing ratio is also relaxed in analysis nudging while the perturbation within geopotential height was relaxed in spectral nudging. Given the

importance of the water vapor distribution in the simulation, analysis nudging is employed here at 6 hr. Intervals and on a timescale of ~ 1 hr. as reported by Hoskins et al., (2012) for the timescale that reproduces synoptic atmospheric flows over tropics. The nudging is implemented only above 800 hPa. The Planetary Boundary Layer (PBL) layer was not nudged considering the best configuration recommended by J. Lo, personal communication, 2011.

The WRF physics parameterizations here are shown in Figure 3.2 along with their interactions with each other. In this study, the double-moment five-class was the microphysics scheme. The initial set of parametrizations was achieved by Fonseca et al. (2015) who reported several sensitivity tests to determine the best physics parametrization of WRF over WMC. Here a further sensitivity test was carried out to see if the above parametrization scheme could be improved. It was discovered that WRF simulation for R2 data were very sensitive to nudging and the cumulus scheme adopted over WMC. Compared to the modified BMJ cumulus scheme of Fonseca et al. (2015), all tested cumulus schemes were found to significantly over predict seasonal precipitation and the Kain- Fisher scheme was found to be second best cumulus scheme for R2 runs. However, for WRF climate runs using the CMIP5 dataset, the modified BMJ was found to underestimate precipitation significantly. For these runs, the New Tiedtke cumulus scheme was found to be the best physics scheme in reproducing the CHIRPS (observed data) in several sensitivity tests. Kain- Fisher was again found to be the second-best cumulus scheme. Overall WRF was found very sensitive to input data, cumulus scheme and nudging (grid or spectral) over the WMC. The shortwave (SW) and longwave (LW) radiation modeling were parameterized using Rapid Radiative Transfer Model for Global (RRTMG). Yonsei University (YSU) was the planetary boundary layer scheme (Hong et al., 2006), surface layer parameterization was achieved by Monin-Obukhov scheme (Monin and Obukhov, 1954), the land surface model was parametrized by 4-layer Noah land surface model

(Chen and Dudhia, 2001). Betts-Miller-Janjić (BMJ) scheme (Janjić, 1994) with further modifications by Fonseca et al. (2015) was used. The BMJ modifications by Fonseca et al. was adopted to remove bias in the precipitation by the precipitating convective cloud (PCC) scheme and gives a more realistic cloud cover and reduced biases resulting from surface radiation in deep convective regions. The parameterizations also used time-varying sea surface temperatures (SSTs), vegetation fraction and albedo.

Figure 3.2 WRF parametrization scheme in this study and their interdependence (adapted from Skamarock et al., 2008)

3.2.2 ASD Model

The ASD model (Hessami et al., 2008b) uses an inbuilt backward stepwise regression (McCuen, 2016) for predictor selection. Pre-screening of the predictors is employed via correlation with the station precipitation data and via cross-correlations with other predictors. For example, the relative humidity and specific humidity were correlated, and both have high correlations with station precipitation data. However, only relative humidity was chosen as it has a higher correlation with precipitation. Predictors having cross correlations more than 0.5 were discarded which was stricter

than the criteria of 0.7 reported by Pallant (2007). This avoids overfitting of data during calibration which would lead to substantial biases and error in validation.

Table 3.2 List of predictors extracted from WRF output and R2 data in statistical downscaling as preliminary selected as based on literature reviews on SD and the meteorological variables driving convection, i.e., likely correlation with the station level precipitation. Each of these predictors was bias corrected by a normalization approach with a subtraction of the mean followed by division by the standard deviation (Wilby et al., 2004; Mahmood et al., 2013).

Table 3.2 List of predictors extracted from WRF output and R2 data in statistical downscaling

Predictor	Predictor Code	Predictor	Predictor Code
500 hPa Relative humidity	rh_500	1000 hPa divergence	p_z_1000
850 hPa Relative humidity	rh_850	500 hPa geopotential height	gp_500
1000 hPa Relative humidity	rh_1000	850 hPa geopotential height	gp_850
500 hPa zonal wind velocity	u_500	1000 hPa geopotential height	gp_1000
850 hPa zonal wind velocity	u_850	500 hPa vorticity	z_500
1000 hPa zonal wind velocity	u_1000	850 hPa vorticity	z_850
500 hPa meridional wind velocity	v_500	1000 hPa vorticity	z_1000
850 hPa meridional wind velocity	v_850	the temperature at 2m	t2m
1000 hPa meridional wind velocity	v_1000	surface specific humidity	shum
850 hPa divergence	p_z_850	Mean sea level pressure	mslp

A further two-step approach was used to diagnose and select the predictor variables. Firstly, the predictors in Table 3.2 were sorted by their correlation with the station rainfall. However, the correlations were found to be very low for most of the variables. Secondly, the limitation of a “black box” correlation approach was mitigated by using a white-box decision-tree which was found to be effective in downscaling of atmospheric processes (Goyal et al., 2012a; Goyal et al., 2012b; Maraun et al., 2015). This two-step predictor selection approach was applied as follows:

Firstly, a correlation matrix was formed between the rainfall with all the predictors as shown in Table 3.2 over the time interval of 1st Jan 1980 to 1st Jan 2005 (i.e., 25 years with 9133 daily records) for both WRF and R2 predictors. A list of predictors with a p-value less than 0.05 for station precipitation, were selected as significant predictors.

These daily correlations while low enable the best-correlated predictors to be determined. It was found that the WRF predictors had notably higher correlations compared to those of R2 predictors for the 4 stations. This is seen in Figure 3.3 for the top five correlated predictors at, e.g., the Singapore station were used as predictors. Thus, the change in resolution due to DD had improved on the correlations of the significant predictors.

Figure 3.3 Correlations of the top 5 predictors extracted from WRF (left) and R2 (right) with rainfall for Singapore station (Significant predictors with p-value less than 0.05)

However, as the physics was not clearly understood, there was a possibility that some predictors with lower correlation also played a significant role. Thus, in the second step, a decision tree was generated (Figure 3.4) to understand the influence of predictors on the precipitation amount. In the decision tree, the parent node containing all the observations was subdivided into two new nodes, constructed on the variable that categorized a node better than the present node; this was done iteratively until the entire dataset was explained and the terminal nodes achieved. The stopping criteria of a node was that the new node emerging should have at least 20 observations to divide the preceding node. In other words, a decision tree is a flowchart-like structure in which each internal node represents a "test" on an attribute (e.g., whether a coin flip comes up heads or tails), each branch represents the outcome of the test, and each leaf node represents a class label (decision taken after computing all attributes). The paths

from the root to leaf represent classification rules. The decision tree was achieved using the “rattle” package in R. The outcome derived from the decision tree for the case when of WRF predictors were:

- i. If daily relative humidity at 1000 hPa was greater than 78, it resulted in wet days with precipitation mean of 9.8 mm 35% of the times. However, if rh_100 was less than 78, it resulted in dry day or less rainfall with mean 3.7mm
- ii. Relative humidity at 500 hPa (rh_500) decided on the extreme rainfall events. In case rh_500 was greater than 87, it resulted in a mean precipitation of 16 mm while rh_500 less than 87 results in comparatively lower precipitation mean of 8.1mm.
- iii. Finally, the zonal velocity at 850 hPa greater than 0.098 m/s (westward) could result in extreme events.

Figure 3.4 Decision tree corresponding to the two grids WRF (top right) and R2 (bottom right). The first parent is the top-most node in the decision tree shown. Time range covers 1st Jan 1980 – 1st Jan 2005 or 9133 daily records

The decision tree model for the R2 predictors was slightly more complex with 5 end nodes to explain 100% of the data as compared to 4 end nodes for WRF predictors. It is noted that a decision tree model with lesser end nodes is considered more robust. Specifically, for the R2 predictors,

- i. Relative humidity at 100 hPa was again the primary variable that differentiates the low and high precipitation. rh_1000 hPa is greater than 98 results in extreme events 5% of the times with mean rainfall of 16 mm. rh_1000 values less than 98 results in rainfall mean of 5.8 mm 95% of the time.
- ii. Furthermore, rh_500 greater than 71 results in precipitation with a higher mean value of 8.2 mm 22% of the time.
- iii. Most of the data were concentrated for rh_500 hPa less than 71, resulting in dry or less precipitating events with a mean of 4.2 mm.
- iv. Lastly, the extreme precipitating events were decided by the zonal velocity at 850 hPa. u_850 between 0 to 1.7 mm/s (eastward) or greater than 0 (westward) results in extreme events with mean precipitation 20 mm.

Lastly, the union of step 1 (5 top correlated predictors) and step 2 (node distributing variables) were used in statistical downscaling.

3.2.3 Data Used

Large-scale atmospheric data

The National Centers for Environmental Prediction (NCEP) R2 2.5° x 2.5° data is employed to provide boundary and initial condition to WRF model in DD. The R2 data has the same resolution as in the Reanalysis (R1) data (T62 Gaussian grids and 28 vertical sigma levels) and is archived four times daily. The R2 dataset provided for error correction in the R1 dataset and with a parametrization update for better representation of the physical processes (NCEP, 2000).

Observed gridded data

The TRMM multi-satellite precipitation analysis (TMPA) 3B42 v7 product is employed for verification of WRF rainfall output. TRMM data is widely used in hydrological modeling and case studies (e.g., Tong et al., 2014; Zhu et al., 2016; Zulkafli et al., 2014). Its accuracy has been assessed with ground-based radar estimates or gauge precipitation (e.g., Maggioni et al., 2014; Moazami et al., 2013; Zulkafli et al., 2014) with favorable comparisons.

Climate Forecast System Reanalysis (CFSR) data

The CFSR dataset, having a spatial resolution of 38 km and with 64 vertical levels, is a Reanalysis dataset from NCEP (Kistler, 2010). It is used here to independently assess the capability of WRF output variables in capturing the regional climate when used as predictors in ASD. This is achieved by computing the correlations between these variables as reported in CFSR dataset with the TRMM observed rainfall. CFSR is used in this assessment because of its fine regional grid size, good temporal resolution and sufficient historical coverage (1979-2009) (Fuka et al., 2013).

Local observed precipitation data

The application of ASD requires two sets of data comprising the local predictand (precipitation here) and atmospheric predictors. Local station rainfall data for Singapore (Changi station) is obtained from the Meteorological Service of Singapore while daily precipitation data for other three locations are obtained from the Global Historical Climatology Network (GHCN) archive.

3.3 Results and Discussion

3.3.1 Verification of dynamically downscaled output from WRF

Figure 3.5 shows the spatial comparison for seasonal precipitation rate (in mm/day) between the WRF output and TRMM data for the 9-year period (from Jan 2000 to Dec 2008). The spatial patterns are very similar even though the grid size of the

WRF output is twice that of TRMM. The WRF model captures the general spatial pattern of rainfall as observed TRMM dataset. However, WRF had a lower magnitude to that of TRMM over WMC. Furthermore, a higher precipitation amount is seen over elevated terrain in the western Sumatra mountainous areas where the daily mean rates are twice higher TRMM. The coastal regions having low elevation were observed to have fewer biases in the WRF output. The WRF output had the closest agreement with TRMM during months of DJF, MAM, and JJA. A significant bias for SON was observed where WRF underpredicts the excess precipitation resulting from night time diurnal cycle SON over the coastal Sumatra region (Fujita et al., 2013; Yokoi et al., 2017).

The higher precipitation seen in the WRF results on the elevated terrains or mountain partially resulted from the horizontal grid spacing of 50 km used in WRF model. The resolution with its smoothed topography was comparatively coarser and thus do not take account of distinct convective updrafts and can result in substantial biases. However, TRMM precipitation measurements over steep or elevated topography should be taken care while assessing for biases within the model over steep topography or elevated terrain. For example, TRMM under-estimated precipitation by around fifty percent when compared to observational data in the New Guinean highlands (Matthews et al. 2013; Hassim et al. 2016). However, as a precautionary measure, all the study locations have been chosen far from the western boundary where significant differences between WRF and TRMM were observed. e.g. Western Sumatra where there is 50% under prediction by WRF compared to TRMM.

The above comparison with TRMM indicates that WRF model was satisfactory in simulating the meso-beta scale resolution precipitation. It is observed from Figure 3.5 that WRF predicted precipitation values are close to TRMM values near Singapore, Labuan, Jakarta and Hat Yai which also ascertains to the accuracy of the predictors,

as the precipitation output from the WRF cumulus scheme is also dependent on the accuracy of these variables for its parametrization (Gilliland et al., 2007).

Figure 3.5 Simulated WRF (50km x 50km) (left) and observed TRMM (25km x 25km) (right) from Jan 2000 – Dec 2008, 9-year seasonal precipitation (mm/day) comparison.

3.3.2 Comparison of WRF and CFSR variables with TRMM

A correlation comparison is used to demonstrate that the parameterizations in WRF are appropriate so that WRF captures the regional climatology as in the CFSR dataset. This is done via correlation between TRMM precipitation with the predictors as extracted from WRF and with the predictor values from the assimilated CFSR dataset. Figure 3.6 shows correlations between the TRMM precipitation and three predictors, gp_100, u_1000 and rh_500 which were the significant predictors from the predictor selection process described in Section 3.2.2. Here the 38km grid resolution CFSR dataset and 25km TRMM dataset was regridded to 50 km to match the WRF grid resolution in the correlation analysis. It is seen that the WRF variables not only replicated the positive and negative correlations trends but also had similar magnitudes as compared to CFSR variables. This favorable comparison of the correlations supports the ability of WRF predictors to capture the physics needed for modeling precipitation events.

Comparison of WRF and CFSR variables with TRMM” provides a further confirmation that WRF captured the regional climatology beyond the comparison with the TRMM rainfall data. It shows that the WRF predicted atmospheric variables, as with the observed CFSR atmospheric have comparable correlations to the observed TRMM rainfall. This confirms that the regional climatology, particularly the atmospheric variables selected as predictors for ASD are correctly modelled via the WRF dynamic downscaling. This additional WRF verification step using historical CFSR data is important to demonstrate the applicability of WRF for future runs where no observational data exists

1(a) WRF-Geopotential at 1000hpa

1(b) CFSR-Geopotential 1000hpa

2(a) WRF : u_vel_100

2(b) CFSR : u_vel_1000

3(a)WRF : Relative humidity at 500hPa

3(b) CFSR: Relative humidity at 500hPa

Figure 3.6 Correlation between predictors extracted from WRF (left) and CFSR (right) with observed TRMM data over the WMC (time range from Jan 2000 to Dec 2008)

3.3.3 Calibration and Validation of ASD model using Integrated Approach

Figure 3.7 shows a time series plot of the ensemble and observed datasets for the mean monthly rainfall at the Singapore station after SD using ASD. It was seen that the ensemble data could encompass the observed data within its range. Four statistical metrics, the mean, standard deviation, number of wet days and excess kurtosis, was used to check on the accuracy of the downscaled results. The mean demonstrates the ability of the models to capture the monthly and annual rainfall magnitudes while the standard deviation, the variance of the observed data. The number of wet days is included as precipitation being a dependent variable (threshold of 0.1mm used here) as a good model should predict the monthly wet day distribution. Lastly, excess kurtosis measures the fatness of the tail of the distribution which indicates how efficiently extreme rainfall is captured.

Figure 3.7 Simulated 100 realization of calibrated precipitation monthly mean (time range Jan 1980 - Dec 1998) ensemble for Singapore region using (a)WRF predictors (top) & (b)R2 (bottom)

The metrics were evaluated using the indicators of absolute error; root mean square error (RMSE), Nash–Sutcliffe efficiency (NSE) and the bR^2 . bR^2 was chosen over the usual R^2 as a determinant of correlation as it avoids the situation whereby a model can systematically over- or under-predicts while still having an R^2 close to unity (Krause et al., 2005). Instead, the bR^2 coefficient (range 0 to 1) accounts for biases in the observed data and thus more appropriate here. The NSE value ranges from -1 to 1 with NSE closer to 1 indicating a better fit while NSE values < 0 signifying the mean of the data is a better predictor than the model itself.

Figure 3.8 shows simulated monthly mean, standard deviation, the number of wet days and excess kurtosis over Singapore driven from both WRF (left) and R2 (right) predictors for all months over the calibration period (1980-2000). It is observed that in general ASD forced by WRF predictors showed much better performance in reproducing the rainfall statistics when compared to R2 predictors. The median of the ensemble is closer to the observed monthly median, and the observed median is contained by interquartile range (IQR). This better performance when using WRF predictors is attributed to the WRF DD step in deriving predictors for ASD. Figure 3.9 shows a similarly good comparison for the validation period.

For the Singapore location, the NSE and bR^2 values for the mean annual rainfall with WRF predictors are 0.58 and 0.78 , respectively, for the validation dataset. This is significantly improved over the R2 predictors where the NSE was -0.57 and bR^2 0.68 . The higher predicted precipitation when using R2 predictors is a result of increased residual daily rainfall. This is consistent with R2 predictors underestimating the number of wet days (Figure 3.9(c)). The excess kurtosis showed in Figure 3.8(d), and Figure 3.9(d) indicates that volatility in the rainfall was better captured by the integrated approach.

Figure 3.8 25th and 75th quantile of the calibration results for the Singapore location using the WRF predictors (left) and R2 predictors (right) along with observed median in dotted line. The statistics plotted in the figures are (a) mean, (b) standard deviation (c) wet days (d) excess kurtosis

Figure 3.9 25th and 75th quantile of the validation results for Singapore location using the WRF predictors (left) and R2 predictors (right) along with observed median in dotted line. The statistics plotted in the figures are (a) mean, (b) standard deviation (c) wet days (d) excess kurtosis

However, it is also seen that ASD forced by R2 predictors showed comparable or better statistics at various periods. For example, standard deviation showed fewer improvements in the integrated approach compared to pure SD as can be seen in Figure 3.8(b) and Figure 3.9(b). Furthermore, both models failed to capture the

January excess kurtosis for the validation period as seen in Figure 3.9(d). Overall, the WRF predictor generated precipitation is seen to be closer to the observed data, as shown mainly by the interquartile range (IQR) of the box plots. The observed value was contained by IQR box for the mean precipitation, standard deviation, number of wet days and excess kurtosis for the ensemble generated by WRF predictors for the calibration period, and except for January and December, for the validation period as well.

The absolute error for the mean rainfall and number of wet days showed the largest improvement across the four locations as plotted in Figure 3.10 and Figure 3.11. The improvement is largest for the Singapore station. This is because the coarser grid R2 data covers an area of 277 km x 277 km compared to Singapore's land area of 718 km². Together with the adjacent land mass, this R2 grid for Singapore has a land region spanning 43% of the grid area (see Table 3.3). In contrast, the WRF grid being 50km x 50 km captures more the spatial extent of Singapore with its higher land to grid ratio of 58%. Further, this R2 grid contains two separate land masses separated by a sea region between eastern Sumatra and western Malaysia which implies the predictor data from R2 for Singapore has properties from two landmasses. The grid for Jakarta also has an improved land to grid ratio when going to WRF grids. The R2 grids for Labuan and Hat Yai have a single land mass and have land to sea ratios higher than Singapore, though the WRF grid for Labuan has a much lower land to sea ratio. All these indicate that the Singapore location would show the largest improvement as indeed seen.

The other three locations also showed improvements with that in the mean rainfall being the most noticeable. However, the stations Labuan and Hat Yai have larger missing values at 10.8% and 13.9%, respectively (see Table 3.1) and this likely affected their performance.

Table 3.3 Land/grid percentage for Reanalysis-II grid and WRF grid.

location	Land/grid %	
	(R2)	Land/grid % (WRF)
Singapore	43	58
Labuan	45	13
Hat Yai	67	93
Jakarta	45	83

Figure 3.10 Absolute error between simulated monthly mean (left) and number of wet days (right) over the calibration period at four locations (1) Singapore (2) Hat Yai (3) Labuan and (4) Jakarta, using (a) R2 predictors (dark gray) and (b) WRF predictors

Figure 3.11 Absolute error between simulated monthly mean (left) and number of wet days (right) over the validation period at four location (1) Singapore (2) Hat Yai (3) Labuan and (4) Jakarta, using (a) R2 predictors (dark gray) and (b) WRF predictors

Improvements in all metrics quantified using the RMS, NSE and bR^2 values are tabulated in Table 3.4 and Table 3.5 for all 4 locations. The RMSE values for the average, the number of wet days and excess kurtosis improved at all four locations with the improvements being the most significant in the mean and number of wet days. The most noticeable improvements are for Singapore location where using WRF predictors resulted in an RMSE value of 0.81 for mean monthly rainfall for calibration period while R2 predictors had a much larger RMSE of 3.86. The improvements are greater for the calibration period, as the time span of the calibration data is longer compared to validation data. The prediction performance in the number of wet days also improved significantly. The Singapore location had an improvement of approximately two days for the calibration period while other 3 sites had improvements of one to two days. The excess kurtosis improved almost across all four

locations indicating improvements in the in simulating extreme parameters in the length of the tail in the distribution.

As with the RMSE, Table 3.4 and Table 3.5 also shows improvements via the bR^2 values. The predicted monthly mean rainfall for Singapore using the WRF predictors has bR^2 value 0.90 and 0.78 for calibration and validation, respectively, as compared to using R2 predictors having bR^2 values of 0.64 and 0.68. Similar improvement trends are seen in the bR^2 , RMSE values for the other three locations though at different levels of improvement. The NSE values for the monthly mean values using R2 predictors were negative for all sites except Malaysia during the validation period indicating a bad precipitation prediction. The corresponding NSE values were small and close to zero over the calibration period. In contrast, the NSE values in the integrated downscaling approach using WRF predictors have improved NSE values > 0 and closer to 1 over the calibration period for all the parameters, except for the standard deviation. This resulted in correspondingly better NSE values over the validation period.

Table 3.4 Error parameters for the statistical parameters of the ensemble w.r.t the observed station values in calibration. The RMSE value has units mm for the mean, standard deviation and excess kurtosis, and day for the number of wet days.

Calibration

Parameter	Location	Mean		Std. Dev.		Wet Days		Exc. Kurtosis	
		WRF	R2	WRF	R2	WRF	R2	WRF	R2
RMSE	Singapore	0.81	3.86	1.43	1.48	0.95	2.82	0.47	1.2
<i>bR2</i>		0.9	0.64	0.8	0.77	0.88	0.72	0.67	0.69
NSE		0.84	-1.87	0.7	0.7	0.87	-0.39	0.45	-4.22
RMSE	Thailand	1.26	2.05	3.56	3.42	1.11	2.59	0.95	1.4
<i>bR2</i>		0.75	0.73	0.35	0.39	0.91	0.8	0.67	0.71
NSE		0.65	0.17	-0.84	-0.59	0.95	0.69	0.46	-0.63
RMSE	Malaysia	1.08	1.43	1.52	1.18	0.91	1.11	0.36	0.4
<i>bR2</i>		0.9	0.85	0.88	0.92	0.93	0.9	0.81	0.76
NSE		0.9	0.82	0.85	0.92	0.94	0.91	0.79	0.75
RMSE	Indonesia	2.06	2.7	2.27	2.54	1.41	1.75	0.66	0.98
<i>bR2</i>		0.65	0.63	0.73	0.69	0.93	0.84	0.59	0.47
NSE		0.55	0.3	0.71	0.62	0.92	0.87	0.78	0.32

Table 3.5 Error parameters for the statistical parameters of the ensemble w.r.t the observed station values in validation. The RMSE value has units mm for the mean, standard deviation and excess kurtosis, and day for the number of wet days.

Validation

Parameter	Location	Mean		Std. Dev.		Wet Days		Exc. Kurtosis	
		WRF	R2	WRF	R2	WRF	R2	WRF	R2
RMSE	Singapore	1.37	3.41	4	3.52	1.84	3.98	1.1	1.8
<i>bR₂</i>		0.78	0.68	0.74	0.76	0.7	0.45	0.37	0.26
NSE		0.58	-0.57	-1.14	-0.24	0.56	-2.53	-1.6	-4.9
RMSE	Thailand	2.19	4.06	1.91	2.56	1.84	3.23	0.64	0.99
<i>bR₂</i>		0.54	0.3	0.54	0.41	0.84	0.71	0.69	0.57
NSE		-0.08	-1.03	0.47	0.31	0.85	0.55	0.73	0.27
RMSE	Malaysia	0.73	0.51	3.47	3.57	2.63	2.65	0.96	0.9
<i>bR₂</i>		2.07	3.02	0.51	0.39	0.55	0.54	0.07	0.11
NSE		0.67	0.2	0.33	0.18	0.53	0.51	-0.37	-0.21
RMSE	Indonesia	4.42	5.04	5.79	5.81	2.08	1.89	1.57	1.77
<i>bR₂</i>		0.41	0.3	0.27	0.22	0.8	0.8	0.21	0.15
NSE		-0.3	-2.21	-0.26	-1.49	0.88	0.81	0.1	-1.49

Finally, Figure 3.12 shows a seasonal comparison of mean precipitation rate over the validation period. The plot shows the WRF and R2 ensemble against observed at the four locations. Singapore and Hat Yai shows a definite improvement for all the seasons, where the mean of simulated ensemble precipitation rate generated from integrated approach is closer to the observed mean. The integrated approach at Labuan station also showed improvement for all seasons except MAM which is the dry season receiving the minimum seasonal rainfall where both downscaling methods had similar performances. In contrary to Labuan, Jakarta showed equal performance for DJF, the

season receiving the maximum seasonal rainfall while other seasons showed improvement. Overall, Figure 3.12 indicates that the integrated approach provides better predictors for statistical downscaling across the four distributed sites of the study domain of WMC.

Figure 3.12 Comparison of seasonal daily precipitation rate mean over the four locations for the validation dataset.

3.4 Conclusions of Integrated Downscaling Approach

An integrated downscaling approach comprising of DD using WRF and SD using ASD was developed and compared with a simple SD method for predicting multi-site daily precipitation. Hydrologically relevant metrics such as mean rainfall, standard deviation, the number of wet days and excess kurtosis were used to demonstrate that the WRF downscaled predictors provided improvements prediction accuracy over the

conventional SD approach of using predictors directly from GCM, specifically R2 here. The significant findings are:

- i. A successful parameterization of physics schemes in WRF model which provided good quality precipitation data at 50 km resolution. The resulting spatial comparison of the WRF output precipitation showed good similarity with observed TRMM data.
- ii. The spatial correlation of TRMM precipitation with significant SD predictors, i.e., geopotential height (500 hPa), specific humidity and meridional velocity at 10m as extracted from WRF were compared with similar correlation when such variables extracted from CFSR. Both WRF and CFSR correlations with TRMM were in agreement indicating that good spatial-temporal quality of WRF was achieved and validating the use of WRF outputs as predictors in SD over the WMC.
- iii. The final predictor selection process was made more robust by incorporating a white-box decision-tree procedure.
- iv. The integrated approach wherein better resolution predictors extracted from DD were in SD improved on the predicted station level precipitation statistics covering the mean, number of wet days and excess kurtosis the observed station data, though the standard deviation is not improved. Furthermore, the improvement was seen in the four distributions location of WMC.

The overall ability of the integrated approach to model the mean, wet days, as well as the extreme parameter of excess kurtosis, makes the integrated downscaling approach suitable for impact studies across the WMC. Future work could focus on temporal disaggregation of the daily output to sub-daily or hourly scale as well using GCM for future climate studies.

4 Construction of Future IDF Curves using RCM data over Western Maritime Continent

4.1 INTRODUCTION

Climate change in the last two decades has resulted in significant alterations in frequency, magnitude, and occurrence of extreme weather events (IPCC, 2013). As per IPCC AR5, the projected global surface temperature increase is estimated to be from 0.3° to 4°C by the year 2100. This is likely to affect future precipitation drastically, especially over the tropics. In a study published in Nature, (Wang et al., 2014) reports that the tropical carbon cycle – the uptake and release of carbon dioxide from and back into the atmosphere – has become twice as sensitive to temperature change in the last 50 years. A one degree rise in average tropical temperature leads to a release of around two billion more tons of carbon per year from tropical forests and savannahs, compared with the 1960s and 1970s. (Wang et al., 2014) further reported on a positive feedback loop whereby warmer summers make forests drier and release more carbon dioxide to make summers even more warmer. This reference is added to the thesis next to the statement “This is likely to affect future precipitation drastically, especially over the tropics ((Wang et al., 2014) whereby a 1oC rise in average tropical temperature leads to a release of around two billion more tons of carbon per year from tropical forests and savannahs as compared with the 1960s and 1970s. Further to this, there has been a significant rise in urbanization since last two decades, accelerating temperature rise and rainfall runoff. The combination of changes in precipitation magnitude, duration, the frequency with urbanization would lead to severe flooding and droughts, particularly in developing countries.

IDF curves are currently developed for most of the region using the annual extremes extracted from observed precipitation time series. For example, it is recommended at

least 30 years of time series data be used to derive 100 years of return period. IDF-curves provide rainfall intensity for the various design return periods and a sub-daily duration such as 5, 10, 30 min and 1 to 48 hrs. Designs for hydraulic structures such as culverts, sewers, detention basins are based on IDF curves for the region. However, one of the critical assumptions while constructing IDF curves is that the extreme historical rainfall is stationary and would represent future rainfall as well. The assumption of stationarity is questionable when rapid climate change is observed throughout the world (Sugahara et al., 2009). GCM outputs have been found to resolve the non-stationarity of the precipitation resulting from climate change. However, GCM models have a lower resolution, i.e., greater than 100 km. As such RCMs can be used to downscale which incorporate the local topographic features and circulations.

This part of the Ph.D. work implements the same logic of dynamically downscaling GCM output using RCM model outputs, to obtain precipitation time series and then deriving the resulting IDF curves for future climate. In this study, two RCM outputs were used: (a) CMIP5-WRF (b) HadGEM3-RA-CORDEX. The WRF model was run using CMIP5 NCAR CESM Global Bias-Corrected CMIP5 Output as a boundary condition and have a horizontal grid resolution of 10 km and temporal output of 6 hr. CMIP5 output of CESM1 data was adopted based on various reported studies (Ahn et al., 2013; Chen et al., 2016). The parameterization of the WRF model was different from what was used in the Integrated downscaling approach of Chapter 3. It was observed that WRF model calibrated better by implementing spectral nudging and using the new Tiedtke cumulus scheme (Tiedtke, 1989; Zhang et al., 2011). Along with this, the microphysics and surface scheme for the WRF runs were changed to the Thompson and Unified Noah Land surface model respectively (Tewari et al., 2004; Thompson et al., 2004). The CORDEX data used in this study is the east part of GCM HadGEM2-AO (Hadley Centre Global Environmental Model, version 2) and

downscaled using RCM HadGEM3-RA. CORDEX was available for download from the CORDEX download portal (<https://cordex-ea.climate.go.kr/main/mainPage.do>). The grid resolution for CORDEX data was 0.44° and temporal resolution obtained was 3 hourly.

In this study, three locations in WMC were examined, i.e., Singapore, Kuala Lumpur and Jakarta (Table 4.1). The Changi station data in Singapore was used as a proof of concept for constructing past and future IDF curves. The observed station-IDF curve for Singapore (Station: S22-Changi) was derived from the hourly rainfall data available from the Meteorological Service Singapore (MSS). The observed station-IDF curves for Kuala Lumpur and Jakarta were constructed from rainfall data provided for the UN International Hydrological Programme Document (Daniell et al., 2008). All the RCM-derived IDF curves were constructed from WRF and CORDEX output.

Table 4.1 The locations over which IDF curves are constructed

Location	Latitude	Longitude
Singapore	1.35° N	103.81° E
Kuala Lumpur	3.13° N	101.68° E
Jakarta	6.17° S	106.86° E

The IDF curves presented were derived using the Gumbel Extreme Value-1 (GEV) distribution. This study uses GEV-1, i.e. Gumbel extreme value distribution (shape parameter =0) for fitting annual extremes. The approach of EV1-MOM was found to fit with the observed data. The GEV is distribution described in Section 4.2.3. Biases in the IDF curves derived from the grid data (WRF, CORDEX, and CFSR) and observation comparisons are presented. The biases in $f^{\sim}$ derived IDF curves were firstly compared with the grid level CFSR data. The bias between CFSR grid and observed station data was also calculated. Both the biases were computed using the

baseline period from Jan 1980 to Dec 2005. The overall methodology used in this study is shown in Figure 4.1

Figure 4.1 A flowchart of the methodology for constructing bias corrected future IDF curve.

Section 4.3 presents the IDF curves developed after consideration of the AR5 climate change projections. Results are shown for the Singapore station for the year 2030 to 2060 under RCP 6.0 and RCP 8.5 and both WRF and CORDEX RCMs. This section further draws comparisons between the two projected IDF curves. Finally, Section 4.4 shows the results for the other two regions, i.e., Kuala Lumpur and Jakarta.

4.2 Data and Extreme Value Distribution Used to Construct IDF Curves

4.2.1 NCAR CESM Global Bias-Corrected CMIP5

The initial and the boundary condition for WRF past and future climate runs required for IDF analysis were provided by CMIP5 CESM bias-corrected data (ds316.1 dataset). This dataset was built in accordance with the IPCC (IPCC AR5) (Monaghan et al., 2014). The data was constructed by bias correcting the outputs achieved from Version 1 of CESM that couples GCM which has 4 different model components that

simulate atmosphere, ocean, sea-ice and land surface. The variables within achieved from CESM runs were bias-corrected by Research Data Archive (RDA) using ECMWF Interim Reanalysis (ERA-Interim) fields from 1981 to 2005, following the method in (Bruyère et al., 2014). This dataset consists of all the initial and boundary conditions required for WRF model simulations as provided in an Intermediate file format. The ds316.1 dataset is provided on RDA website (<https://rda.ucar.edu/datasets/ds316.1/>) for 20th-century simulation (1951-2005) termed as historical simulation and three future scenarios (RCP4.5, RCP6.0, and RCP8.5) spanning from the year 2006 to 2100. The temporal frequency of this data is 6 hourly. As per requirements of this study, the ds316.1 dataset was downscaled using WRF to 25 km grid resolution over the year 1980-2005 (historical) and 2030 -2060 (for RCP 4.5 and RCP 8.5).

4.2.2 CORDEX dataset over East Asia

The Coordinated Regional Climate Downscaling Experiment (CORDEX) data for the South-East Asia domain was used as an alternate source for an RCM in this study. The CORDEX facilitates a good resolution RCM data achieved from downscaling efforts via global partnership. The CORDEX community provides high-resolution data required to access information regarding regional extreme events and to form policies for their mitigation. The CORDEX East Asia product was contributed by National Institute of Meteorological Research (NIMR) which implemented HadGEM3-RA RCM build on the global atmospheric HadGEM3 at the Met Office Hadley Centre (MOHC). The domain for East Asia CORDEX run is shown in Figure 4.2. Only the data from parent grid having 220 (west-east) and 183 (north-south) grid points, with a horizontal resolution 0.44 degree (~ 50km) was used in this study.

Figure 4.2 CORDEX East Asia Model domain and orography (m). [Reproduced from <https://cordex-ea.climate.go.kr/main/modelsPage.do>]

HadGEM3-RA has a fully compressible, non-hydrostatic dynamic core and a height-based vertical coordinate. The CORDEX runs were completed for a total of 151 years; 56-year historical (1950-2005) climate while 95-year from 2006-2100 simulations for the future scenarios (RCP 4.5 and RCP 8.5) climate. The historical run was achieved using Atmosphere-Ocean coupled Hadley Center Global Environmental Model version 2 (HadGEM2-AO) simulation (Baek et al., 2013). RCP 4.5 and 8.5 scenarios of HadGEM2-AO model were used for future runs. In the CORDEX experiment, RCP 4.5 was considered as stabilization scenario wherein the radiative forcing stabilizes before 2100 (Thomson et al., 2011). Whereas, RCP 8.5 scenario was categorized by increasing GHGs and thus represents a scenario which leads to highest concentration of GHGs (Riahi et al., 2011)

4.2.3 Extreme Value Distribution

The extreme statistic is one of the most important components of hydrological studies (e.g., the IDF curves). Usually, historical observed data is one of the reliable information used. Fuller, (1914) introduced the study on floods where extreme events

statistics were used. Extreme value theory underwent theoretical development with one of the most important work by Gumbel (1958).

The most important step in the analysis is to select and fit a suitable probability distribution function for the extreme rainfall. The most popular choices distribution function is the Generalized Extreme Value (GEV) fitted using maximum likelihood, method of moments and least square error method. The GEV distribution is a combination of three different distributions, i.e., Fréchet ($k>0$), Weibull ($k<0$) and Gumbel ($k=0$), where k is the shape parameter. In a GEV distribution for non-stationarity, time-varying covariates are used [Martins et al., 2000; Cheng et al., 2014b; Cheng and AghaKouchak, 2014]. However, in this study, non-stationarity is accounted by using RCM data. Hence, a simple stationary GEV function is used, the quantile function for GEV distribution to obtain maximum intensity for T year return period (see equation 4.1). The quantile function for GEV distribution is obtained as

$$\left(1 - \frac{1}{T}\right)$$

$$I_{max} = \mu + \left(\frac{\alpha}{k}\right) \left\{ \left[-\ln \left(1 - \frac{1}{T}\right) \right]^{-k} - 1 \right\} \quad (4.1)$$

k, μ, α are fitted shape, location and scale parameter respectively.

4.3 Construction and bias correction of Historical IDF curves

The overview of the study is presented in Figure 4.1 The construction of historical IDF curves for the baseline period (1980-2005) was done in three steps:

- i. Dynamical downscaling of GCM product (WRF-CMIP5) or extracting precipitation from downscaled CORDEX RCM dataset.
- ii. Comparison of RCM derived IDF curve with Reanalysis CFSR derived and station data derived IDF curve
- iii. Derivation of present-day station IDF curve using bias correction

The above three steps can be used to generate grid IDF curves for any grid on the RCM and CFSR domain. However, it is necessary to have station data or a historical IDF curve data to derive station level IDF curves further.

Step 1 Dynamical downscaling of GCM product (WRF-CMIP5) and extracting precipitation from downscaled CORDEX RCM dataset.

This was the most computationally intensive part of the work where more than 88 years (26 years historical and 31 years each for RCP 4.5 and RCP 8.5) of WRF runs were done using the parameterization scheme shown in Figure 4.3., WRF model was run using CMIP5 GCM outputs the details of which is provided in Section 4.2.1. The parameterization schemes were compared via WRF runs for the historical period (1980-2005) to that with CHIRPS datasets (the seasonal comparison is shown in Figure 4.4) to find the best combination of the physics. As seen in Figure 4.4 the new parametrization scheme used for WRF model is like 5 km resolution CHIRPS data. CHIRPS data incorporates 0.05° resolution satellite imagery and in-situ station data to create gridded precipitation time series. The comparisons for the new parametrization scheme confirms an improved simulated precipitation. Thus, this a better parametrization scheme than the one described in Section 3.2.1, where modified BMJ scheme was used rather than New Tiedtke cumulus scheme. Some of the significant improvements observed while using the new parametrization scheme as observed from Figure 3.5 and Figure 4.4 are:

- a. The new parametrization did not show the excess precipitation over and near the western mountain ranges (Bukit Barisan). The parametrization scheme described in Chapter 3 exaggerated the precipitation over these high mountain ranges.
- b. The new parametrization scheme could have captured the areas with high precipitation with quite more accuracy compared to the previous parametrization scheme for all the seasons (see Figure 3.5 and Figure 4.4)

The comparisons for WRF (25 km grid output) were made with CHIRPS (5 km grid), a smaller magnitude of rainfall at many areas was expected as 25 km grid could not capture subgrid processes. However, it was found that the new parametrization could capture details that could have been only being resolved only by higher resolution data, for example, the excess precipitation resulting from Borneo vortex was quite accurately captured by the WRF new parametrization scheme.

The future precipitation rate corresponding four seasons (3 monthly) averaged between the year 2030-2060 for RCP 4.5, and RCP 8.5 is shown in Figure 4.5. The distribution of precipitation rate corresponding four seasons (3 monthly) averaged between the year 2030-2060 for RCP 4.5 and RCP 8.5. However, the overall distribution of rainfall rate over four seasons for RCP 4.5 and RCP 8.5 was similar, changes in mean intensity were observed over all the land masses. For example, alterations in precipitation over the west coast of Borneo, where the rainfall intensity observed for RCP 8.5 was higher than that for RCP 4.5 for DJF. Similar increment in mean precipitation was observed over Java Island where the mean rainfall intensity was higher for RCP 8.5 than RCP 4.5 over the northern mountainous region of the Island. It was also noticed that RCP 4.5 had a higher intensity compared to RCP 8.5 over the southern Sumatra for MAM season. However, to study the changes of seasonal precipitation rate for each RCP scenarios a percentage change or bias plot w.r.t to historical seasonal rainfall rate (from CHIRPS averaged to 25 km) was plotted in Figure 4.6. The future seasonal rainfall rate from WRF-CMIP5 was averaged for a year between 2030-2060, while the historical CHIRPS was averaged for a year between 1980-2005. As seen in Figure 4.6 a significant change in rainfall rate was observed for the future scenarios w.r.t historical for all the seasons. The maximum changes in mean seasonal precipitation rate ~ 45% was observed for Peninsular Malaysia for DJF season and Java as well as southern Sumatra for the season JJA. Seasons MAM and SON showed the minimum changes in spatially averaged

precipitation (as seen from Figure 4.6). Interestingly many regions in the WMC might also experience a decline in mean precipitation rate for the future scenarios, for example, south-east Borneo in the season DJF and MAM, western Sumatra and Malay Peninsular for MAM. However, as most of the study, e.g., Donat et al. (2016) and Lehmann et al. (2015) reported that global warming is already increasing the odds of extreme precipitation events. It is of much more relevance to study the changes in extreme precipitation which would be the focus of this study in further sections.

Figure 4.3 WRF physics parameterization for running CMIP5 model

Along with RCM WRF-CMIP5, CORDEX data (described in Section 4.2.2) as belonging to the CORDEX East Asia run and was download from the CORDEX download portal (<https://cordex-ea.climate.go.kr/main/searchPageCdx.do>). The CORDEX HadGEM2-AO dataset was available for the same periods as the WRF run (historical: 1980 to 2005, RCP 4.5 and RCP 8.5 future data: 2030-2060) and hence provided uniformity to the study. The completion of step 1 enables the output data to be used for a companion in Step 2.

Figure 4.4 Comparison plots for seasonal rainfall rate (mm/day) between CHIRPS (left) and WRF (right) for the baseline period (1980-2005).

Figure 4.5 Future mean precipitation as predicted by RCP 4.5 (left) and RCP 8.5 (right) using WRF model for the year 2030-2060.

Figure 4.6 percentage precipitation changes (future modeled precipitation – observed historical precipitation)/observed precipitation) for WRF model RCP 4.5 (left) and RCP 8.5 (right) for the year 2030-2060

Step 2 Comparison of RCM derived IDF curve with Reanalysis CFSR derived and Station data derived IDF curve

Once the RCM outputs from Step 1 were achieved, annual maximum rainfall series was generated from the RCM datasets. After this, the annual maximum series was fitted to Gumbel Extreme Value distribution for constructing IDF curve (e.g., Figure 4.7 where Singapore rain gauge data was used). The Gumbel fit for the annual maximum rainfall corresponding to 1hr, 2hr, 3hr, 8hr and 24hr duration for is shown. Further, the IDF curves published by Public Utilities Board (PUB) Singapore for stations averaged over Singapore was compared against the IDF curves to validate the methodology of constructing IDF curves (see in Figure 4.8). Though the IDF curves for the individual Changi station is slightly higher than average of all the stations in IDF curves were quite close to each other, hence validating the methodology for the constructed IDF curves.

Figure 4.7 Gumbel fit with observed annual extreme values for Singapore rain gauge from the year 1980 to 2005.

As the main objective is to construct IDF curves for the three regions in WMC, the concept was first analyzed with the sites having station IDF curves. Such historical IDF curves were generated for Singapore, Jakarta and Kuala Lumpur stations. These IDF curves (1980-2005) were used for comparisons and in bias corrections. The derived bias factors would also provide corrections to IDF curves under future climate scenarios. This bias correction was necessary as WRF and other RCM grid models tend to underestimate rainfall over the WMC when compared to station data.

Figure 4.8 Constructed IDF curve for this study using Changi (Singapore) station data (left) and the IDF curve published by Public Utilities Board (PUB) Singapore for station averaged over Singapore (right)

The baseline period in this study was from the year (1980-2005). The steps for bias correction of IDF curves were followed in following steps:

a) Firstly, the IDF curves corresponding to the Singapore, Jakarta and Kuala Lumpur CFSR grids were plotted along with the WRF generated IDF curves grids. These plots for Singapore station corresponding to 100, 50 and 25-year return period are shown in Figure 4.10a.

(c)

Figure 4.9 Present station IDF curves (a) Singapore (b) Jakarta (c) Kuala Lumpur

b) The historical IDF curves for the three stations were used to provide a comparison between the grid (CFSR, WRF) and the station level of IDF curves and were constructed using data listed in Table 4.2. These IDF curves were used to determine scaling factor to downscale from grid level data to station level.

It can be seen in Figure 4.10 that WRF-IDF under-estimated the rainfall when compared to CFSR-IDF, though the trend is constant.

Table 4.2 Table showing IDF values corresponding to which historical-station level IDF curves were constructed.

STATION DATA	Hour\ R.P	100 yr	50 yr	25 yr	5 yr	CFSR	Hour\ R.P	100 yr	50 yr	25 yr	5 yr
	Singapore						Singapore				
	1hr	158.14	137.13	118.07	79.55		1hr	72.23	61.4	51.59	31.74
	2hr	102.25	88.67	76.35	51.44		2hr	57.44	48.83	41.02	25.24
	6hr	45.37	39.34	33.87	22.82		6hr	35.61	30.27	25.43	15.65
	8hr	36.13	31.33	26.97	18.17		8hr	30.9	26.27	22.07	13.58
	12hr	26.04	22.58	19.44	13.1		12hr	25.12	21.35	17.94	11.04
	18hr	18.67	16.19	13.94	9.39		18hr	20.28	17.24	14.48	8.91
	24hr	14.72	12.76	10.99	7.4		24hr	17.37	14.77	12.41	7.63
	48hr	8.25	7.15	6.17	4.15		48hr	11.89	10.11	8.49	5.23
Jakarta					Jakarta						
1hr	85	67	50	32	1hr	54.42	46	38.37	22.94		
2hr	60	46	37	23	2hr	42.84	36.21	30.2	18.06		
6hr	35	26	20	15	6hr	26.03	22.01	18.35	10.97		
8hr	30	23	18	12	8hr	22.46	18.99	15.84	9.47		
12hr	26	19	16	9	12hr	18.11	15.31	12.77	7.63		
18hr	20	16	13	7	18hr	14.5	12.25	10.22	6.11		
24hr	18	14	10	5	24hr	12.34	10.43	8.7	5.2		
48hr	13	12	7	4	48hr	8.32	7.04	5.87	3.51		
Kuala Lumpur					Kuala Lumpur						
1hr	107.2	96.4	83.9	67.9	1hr	62.89	54.91	47.68	33.06		
2hr	65	58	50	42	2hr	49.03	42.81	37.17	25.77		
6hr	27.5	24.8	21.6	17.5	6hr	29.16	25.46	22.11	15.33		
8hr	25	22	19	16	8hr	24.99	21.82	18.95	13.14		
12hr	15.6	14	12.2	9.9	12hr	19.95	17.42	15.13	10.49		
18hr	12	10	8.8	7	18hr	15.81	13.81	11.99	8.31		
24hr	8.8	7.9	6.9	5.6	24hr	13.37	11.67	10.13	7.03		
48hr	4.9	4.4	3.8	3.1	48hr	8.85	7.73	6.71	4.65		

IDF curve bias correction

Singapore's Changi station (1.345°N and 103.98°E) located in the eastern part of Singapore was selected to illustrate the procedure for bias correction in the IDF curve.

The Singapore IDF curves were derived using hourly station data as shown in Figure 4.9a.

Firstly, a present IDF curve was derived for the Singapore station using the rainfall extracted from WRF-CMIP5 output along with CFSR-IDF. Figure 4.10 shows IDF curves corresponding to CFSR and WRF-CMIP5 for 100 and 50 year RP. The IDF curve derived from WRF-CMIP5 was lower when compared to CFSR data. For RP 100 year, a range of bias correction of 38% (1-hour), 34% (6-hour) and 29% (24 hours) is seen. However, for RP 50year, bias correction is 40% (1-hour), 36% (6-hour) and 31% (24 hour). The RCM simulations tend to underestimate rainfall over WMC is consistent with the reported works (Davis et al., 2009; Febri et al., 2016; Liew et al., 2014). Thus, as a part of the first step, the scaling factor for intensity was calculated for each duration and return periods between WRF-IDF and CFSR-IDF (grid IDF). The scaling factors would serve as a guideline to derive a bias-corrected grid IDF curve required for Step 2.

Figure 4.10 Singapore WRF-IDF curves along with CFSR IDF for time range 1980-2005, RP=100(left) and RP=50(right).

In the second step, the IDF curve achieved from step 1 (for historical bias corrected WRF-IDF via comparison with CFSR-IDF) was plotted with observed station IDF curves. As a demonstration using Singapore, Figure 4.11 shows IDF curves corresponding to CFSR and station for 100 and 50 year RP. For RP 100 year, the range of bias correction was 54% (1-hour), 14% (6-hour) and -18% (24 hours). However, for RP 50 year, the range was 55% (1-hour), 23% (6-hour) and 15% (24 hours). It is noted that while Step 1 was a pure bias correction of WRF-IDF, step 2 was downscaling from grid level CFSR-IDF to station level IDF. Generally, in statistical downscaling, a relationship is found between the predictors and the predictand. In Step 2 here, however, a simple scaling factor was considered and used between the grid and the station data. Thus, the predictand in Step-2 was the station extreme rainfall intensity, while the predictor was grid extreme rainfall Intensity. The difference between the Step 1 and Step 2 regarding bias correction and downscaling is quite evident from Figure 4.10 and Figure 4.11. As seen in Figure 4.10 the IDF curves for the two grids, i.e., WRF and CFSR were consistent with each other though with WRF underestimating compared to CFSR. Whereas, in Figure 4.11 a different trend between the IDF curve derived from CFSR grid-IDF and station-IDF is seen with the bias percentage changing from positive at lower duration to negative at higher durations. The reason behind lower precipitation at smaller durations for grid rainfall is because grid rainfall represents a spatial mean rainfall of all the “stations” within the grid. Because of the spatial variability of rainfall, not all the stations receive rain at the same time and thus resulting in a lowering of the average intensity when compared to a specific individual station. However, when the duration increases, grid rainfall overtakes the station rainfall (Figure 4.11). This is because the continuous rainfall over a single station is not intense for longer durations. However, over a grid, some stations would be having higher intensity rainfall while other might have lower. The result of this implies a possible higher mean for grid compared to a single station.

Figure 4.11 Singapore CFSR-IDF curves along with station IDF for time range 1980-2005, RP=100(left) and RP=50(right).

4.4 Comparison of Projected Future IDF curve for Singapore, Kuala Lumpur and Jakarta using bias-corrected WRF and CORDEX data

The updated and bias-corrected grid IDF curves for RCP 4.5, and RCP 8.5 over Singapore, Kula Lumpur and Jakarta are presented next. These IDF curves were derived from the future outputs from WRF and CORDEX over the year 2030 to 2060. As seen in Figure 4.12(for WRF) and Figure 4.13(for CORDEX) future IDF curves were derived for durations of 1, 2, 6, 12, 24, 48hr over return periods of 100, 50, 25 and 5 years. The IDF curves were bias corrected using past CFSR grid data. In the bias correction, IDF curves for the past (1980-2005) were plotted for both WRF and CFSR data set. Scaling bias correction factor was calculated between WRF and CFSR IDF curves for 1, 2, 6, 12, 24, 48hr durations. This scaling factor was multiplied to the future (2030-2060) derived IDF curves. The result of multiplying was a bias-corrected grid IDF for the future climate corresponding to RCP 4.5 and RCP 8.5.

The next step was to downscale bias corrected grid level future IDF curve to station IDF curve. To achieve station level bias correction for future RCP scenarios, historical grid bias corrected IDF curve was plotted with observed station IDF curve. The downscaling factor was calculated Corresponding to each duration and return period. This downscale factor achieved from historical data was then multiplied to the future grid level bias corrected IDF to achieve future station level IDF curves.

The final IDF curves achieved from above steps are shown in Figure 4.12(for WRF) and Figure 4.13(for CORDEX). It was observed that almost all the location showed an upward shift of future station IDF curves compared to baseline period IDF curves shown in Figure 4.9.

Figure 4.12 Future bias corrected IDF curves generated from WRF dataset for RCP 4.5 (left) and RCP 8.5 (right) over (a) Singapore (b) Kuala Lumpur and (c) Jakarta

Figure 4.13 Future bias corrected IDF curves generated from CORDEX dataset for RCP 4.5 (left) and RCP 8.5 (right) over (a) Singapore (b) Kuala Lumpur and (c) Jakarta

4.4.1 WRF derived IDF curve and historical IDF curves at station level

The changes in extreme precipitations were quite evident from the WRF generated IDF curves. For example, Figure 4.14 shows the IDF curve for Singapore corresponding to 100-year return period. A clear increase in extreme rainfall events is seen corresponding to all durations for 100 year RP, as also tabulated in Table 4.3.

Figure 4.14 Comparison plot for Singapore (RP=100 year) showing IDF points corresponding to historical, RCP 4.5 and RCP 8.5

Table 4.3 100-year return period IDF curve and the relative change intensity percentage corresponding to RCP 4.5 and RCP 8.5 at Singapore, Jakarta, and Kuala Lumpur

100-year Return Period: WRF DATA					
Hour	Scenarios			Change in % to historical	
	Historical	RCP 4.5	RCP 8.5	RCP 4.5	RCP 8.5
Singapore					
1hr	158.14	175.27	208.05	10.83	31.56
2hr	102.25	113.1	135.58	10.61	32.6
6hr	45.37	49.97	61.2	10.13	34.88
8hr	36.13	39.74	49	9.98	35.61
12hr	26.04	28.6	35.59	9.83	36.68
18hr	18.67	20.48	25.74	9.7	37.86
24hr	14.72	16.13	20.42	9.59	38.7
48hr	8.25	9.03	11.63	9.44	40.93
Jakarta					
1hr	85	107.63	124.63	26.63	46.63
2hr	60	79.69	91.69	32.82	52.82
6hr	35	51.29	58.29	46.55	66.55
8hr	30	45.29	51.29	50.96	70.96
12hr	26	40.96	46.16	57.55	77.55
18hr	20	32.87	36.87	64.37	84.37
24hr	18	30.53	34.13	69.64	89.64
48hr	13	23.84	26.44	83.38	103.38
Kuala Lumpur					
1hr	107.2	111.42	115.05	3.93	7.32
2hr	65	65.67	68.68	1.03	5.65
6hr	27.5	26.19	28.13	4.75	2.3
8hr	25	23.4	25.31	6.4	1.26
12hr	15.6	14.24	15.59	8.74	0.06
18hr	12	10.67	11.81	11.09	1.57
24hr	8.8	7.67	8.57	12.85	2.6
48hr	4.9	4.07	4.65	16.85	5.12

100 Year Return Period WRF-IDF

As seen in

Table 4.3 and for RP 100 year and 1hr duration, the Singapore rainfall increase over historical was observed to be 10.83 % (under RCP 4.5) and 31.56% (under RCP 8.5). Over daily 24 hr. duration increments of 9.44% (RCP 4.5) and 40.93% (RCP 8.5) was observed. Rainfall intensity had an increasing trend at different durations for RCP 8.5 and decreasing trend for RCP 4.5 (see

Table 4.3). An average increase in the intensity of 10.01% for Singapore was observed for RCP 4.5 and 36.10% for RCP 8.5 as averaged over all the calculated durations.

Over Jakarta, both RCP 4.5 and RCP 8.5 had an increasing intensity trend with duration (

Table 4.3). The change in 1-hour intensity was observed 26.63% (RCP 4.5) and 46.63% (RCP 8.5). The most noticeable difference was seen for RCP 8.5 precipitation where the rainfall intensity for 48hr rainfall approximately doubled to that of historical rainfall. A plot for the IDF point corresponding to RCP 4.5 and RCP 8.5 and historical is shown in Figure 4.15.

Figure 4.15 Comparison plot for Jakarta (RP=100 year) showing IDF points corresponding to historical, RCP 4.5 and RCP 8.5

The results for Kuala Lumpur were found to be different from Singapore and Jakarta. The changes in extreme precipitation were much less compared to other two locations. For instance, as seen

Table 4.3 and Figure 4.16 the intensity change for 1 hr. duration was 3.93 % (RCP 4.5) and 7.32 % (RCP 8.5) over the historical. It was also observed that the rainfall intensity for RCP 4.5 was lower than historical intensity for a duration greater than 1 hr. However, the percentage change was very small. The average intensity change magnitude (over 100 year RP) for RCP 4.5 was 8.20 % while for RCP 8.5 it was 3.23 %.

Figure 4.16 Comparison plot for Kuala Lumpur (RP=100 year) showing IDF points corresponding to historical, RCP 4.5 and RCP 8.5

Summarizing the results for three stations together, the largest rainfall intensity change was found at Jakarta station where the intensity approximately doubled for RCP 8.5 and 48 hr durations. For Singapore, a consistent rise in extreme precipitation event was observed for both RCP 8.5 and RCP 4.5. Lastly, Kuala Lumpur had a small increase in the 100 year IDF curve compared to the historical station IDF curve.

50 and 25 Year Return Period WRF-IDF curves

Similar observations for year 50 year and 25 year RP as for 100 years RP was seen at all three locations. For Singapore, as seen in Figure 4.17, Table 4.4 and Table 4.5 with 50 years RP showed an increment in extreme rainfall intensity for both RCP 4.5 and RCP 6.5 scenarios. The rainfall intensity for hourly rainfall for 50 year RP increased by 11.81 % (RCP 4.5) and 32.5% (RCP 8.5). The average intensity change (over all the durations shown) for 50 year RP for RCP 4.5 was 11.06% whereas it was 37.21% for RCP 8.5 this being around 1% more than 100 year RP. For 25 year RP over

Singapore, it is expected the IDF curve for future scenarios came much closer (compared to 50 and 100 year RP) to the observed historical IDF curves as climate change intensifies precipitation for longer RPs. It was observed that the % change in the intensity was small for short durations but higher for a longer duration. Overall it was noted from Table 4.5 for RP 25 year over Singapore, the average intensity increase (over all durations showed) was 13.04 % (RCP 4.5) and 16.775 % (RCP 8.5)

Figure 4.17 A Comparison plot for Singapore RP=50 year (left) and RP=25 (right) showing IDF points corresponding to historical, RCP 4.5 and RCP 8.5

Over Jakarta (RP=50 year), both RCP 4.5 and RCP 8.5 had an increasing intensity trend with duration (Table 4.5). The average intensity magnitude change was 47.81% (RCP 4.5) and 67.81% (RCP 8.5). Furthermore, Jakarta (RP=25 year) IDF curve was much closest to the observed historical IDF curves in comparison to other two larger RPs. The average of intensity magnitude change (over 50 year RP) was observed 40.58% (RCP 4.5) and 60.48% (RCP 8.5)

Lastly, over Kuala Lumpur (RP=50 year) both RCP 4.5 and RCP 8.5 scenarios had an increasing intensity trend with duration (Table 4.5). The average intensity magnitude change was observed 7.78% (RCP 4.5) and 3.36% (RCP 8.5). At RP=25 year the

absolute average intensity change (over 50 year RP) was observed 7.24% (RCP 4.5) and 3.53% (RCP 8.5)

Figure 4.18 Comparison plot for Jakarta RP=50 year (left) and RP=25 (right) showing IDF points corresponding to historical, RCP 4.5 and RCP 8.5

Figure 4.19 Comparison plot for Kuala Lumpur RP=50 year (left) and RP=25 (right) showing IDF points corresponding to historical, RCP 4.5 and RCP 8.5

Table 4.4 50-year return period WRF-IDF curve and the relative change intensity percentage corresponding to RCP 4.5 and RCP 8.5 at Singapore, Jakarta and Kuala Lumpur

50-year Return Period: WRF DATA					
Hour	Scenarios			Change in % to historical	
	Historica	RCP 4.5	RCP 8.5	RCP 4.5	RCP 8.5
	I				
Singapore					
1hr	137.13	153.33	181.7	11.81	32.5
2hr	88.67	98.94	118.44	11.58	33.57
6hr	39.34	43.73	53.49	11.16	35.96
8hr	31.33	34.77	42.82	10.99	36.68
12hr	22.58	25.04	31.13	10.91	37.85
18hr	16.19	17.95	22.51	10.87	39.06
24hr	12.76	14.13	17.86	10.72	39.95
48hr	7.15	7.9	10.17	10.49	42.18
Jakarta					
1hr	67	81.39	94.79	21.48	41.48
2hr	46	58.64	67.84	27.47	47.47
6hr	26	36.57	41.77	40.64	60.64
8hr	23	33.34	37.94	44.97	64.97
12hr	19	28.73	32.53	51.19	71.19
18hr	16	25.26	28.46	57.91	77.91
24hr	14	22.81	25.61	62.9	82.9
48hr	12	21.11	23.51	75.93	95.93
Kuala Lumpur					
1hr	96.4	101.16	103.99	4.94	7.88
2hr	58	59.19	61.59	2.05	6.19
6hr	24.8	23.85	25.5	3.83	2.83
8hr	22	20.79	22.4	5.49	1.81
12hr	14	12.91	14.06	7.81	0.44
18hr	10	8.97	9.89	10.29	1.1
24hr	7.9	6.96	7.74	11.91	2.05
48hr	4.4	3.7	4.2	15.98	4.62

Table 4.5 25-year return period WRF-IDF curve and the relative change intensity percentage corresponding to RCP 4.5 and RCP 8.5 at Singapore, Jakarta and Kuala Lumpur

25-year Return Period: WRF DATA					
Hour	Scenarios			Change in % to historical	
	Historical	RCP 4.5	RCP 8.5	RCP 4.5	RCP 8.5
I					
Singapore					
1hr	118.07	133.54	157.98	13.1	33.8
2hr	76.35	86.23	103.01	12.94	34.92
6hr	33.87	38.13	46.55	12.59	37.44
8hr	26.97	30.33	37.28	12.47	38.22
12hr	19.44	21.85	27.1	12.4	39.4
18hr	13.94	15.66	19.61	12.32	40.69
24hr	10.99	12.33	15.57	12.2	41.63
48hr	6.17	6.92	8.89	12.09	44.04
Jakarta					
1hr	50	57.7	67.7	15.39	35.39
2hr	37	44.84	52.24	21.19	41.19
6hr	20	26.78	30.78	33.91	53.91
8hr	18	24.8	28.4	37.8	57.8
12hr	16	23	26.2	43.76	63.76
18hr	13	19.54	22.14	50.32	70.32
24hr	10	15.51	17.51	55.11	75.11
48hr	7	11.75	13.15	67.86	87.86
Kuala Lumpur					
1hr	83.9	89.19	91.11	6.3	8.6
2hr	50	51.66	53.46	3.33	6.92
6hr	21.6	21.05	22.36	2.54	3.52
8hr	19	18.19	19.48	4.24	2.54
12hr	12.2	11.37	12.33	6.77	1.05
18hr	8.8	8	8.76	9.06	0.4
24hr	6.9	6.16	6.81	10.78	1.33
48hr	3.8	3.23	3.65	14.97	3.95

4.4.2 Comparison between CORDEX derived IDF curve and historical IDF curves at station level

100 Year Return Period CORDEX-IDF

As seen in

Table 4.3 and Figure 4.14 Comparison plot for Singapore (RP=100 year) showing IDF points corresponding to historical, RCP 4.5 and RCP 8.5, 100 year RP and hourly duration had the Singapore rainfall intensity decrease over historical by 3.08 % (RCP 4.5) and increased by 2.27% (RCP 8.5). Over 48-hour duration increments of 22.78% (RCP 4.5) and 31.65% (RCP 8.5) was observed. Rainfall intensity had an increasing trend with duration hours for RCP 8.5 and RCP 4.5 (except for 1-hour duration). An average (overall durations shown and RP = 100 years) increase in the intensity magnitude of 13.01% for Singapore was observed for RCP 4.5 and 16.775 for RCP 8.5 over all the calculated duration.

Over Jakarta, both RCP 4.5 and RCP 8.5 had an increasing intensity trend with duration. The change in 1-hour intensity was 11.5% (RCP 4.5) and 54.93% (RCP 8.5). The intensity change for RCP 8.5 precipitation over Jakarta for 48hr rainfall was approximately 50%.

A similar analysis shows the results for Kuala Lumpur were similar to Singapore and Jakarta. For instance, as seen in Table 4.6 the intensity change over historical data for 1 hr. duration was 8.29 % (RCP 4.5) and 22.5 % (RCP 8.5). However, the percentage difference was small for RCP 4.5 and larger for RCP 8.5. The average intensity magnitude change (over 100 year RP) for RCP 4.5 was 3.375 % while for RCP 8.5 it was 15.96 %.

Table 4.6 100-year return period CORDEX- IDF curve and the relative change intensity percentage corresponding to RCP 4.5 and RCP 8.5 at Singapore, Jakarta, and Kuala Lumpur

100-year Return Period: CORDEX DATA					
Hour	Scenario			Change in % to	
	Historical	RCP 4.5	RCP 8.5	RCP 4.5	RCP 8.5
Singapore					
1hr	158.14	163.01	161.73	3.08	2.27
2hr	102.25	107.74	107.95	5.37	5.58
6hr	45.37	50.15	51.27	10.54	13
8hr	36.13	40.6	41.71	12.36	15.45
12hr	26.04	29.79	30.9	14.4	18.68
18hr	18.67	21.83	22.85	16.94	22.4
24hr	14.72	17.5	18.43	18.88	25.17
48hr	8.25	10.13	10.86	22.78	31.65
Jakarta					
1hr	85	75.22	131.69	11.5	54.93
2hr	60	52.35	92.62	12.75	54.36
6hr	35	29.59	53.55	15.47	52.99
8hr	30	25.11	45.82	16.3	52.75
12hr	26	21.43	39.52	17.58	52.02
18hr	20	16.32	30.34	18.38	51.71
24hr	18	14.52	27.2	19.32	51.14
48hr	13	10.23	19.5	21.34	50
Kuala Lumpur					
1hr	107.2	116.09	131.32	8.29	22.5
2hr	65	69.41	78.58	6.79	20.88
6hr	27.5	28.46	32.28	3.48	17.38
8hr	25	25.63	29.07	2.51	16.29
12hr	15.6	15.81	17.92	1.35	14.86
18hr	12	11.94	13.6	0.46	13.3
24hr	8.8	8.7	9.91	1.15	12.64
48hr	4.9	4.75	5.39	2.97	9.9

50 and 25 Year Return Period CORDEX-IDF curves

The observations at 50 year and 25 year RP were similar to the 100 year RP for all the three locations. Singapore as seen in Figure 4.13, 50 year RP showed an increment in rainfall intensity for both RCP 4.5 and RCP 8.5 forcing scenarios. The rainfall intensity for an hourly rainfall was observed to increase by 2.64 % (RCP 4.5) and 1.68% (RCP 8.5) at 50 year RP. The average change in intensity magnitude for RCP 4.5 for all durations shown at 50 year RP was 12.68% whereas it was 16.25% for RCP 8.5. For 25 year RP over Singapore, the expected IDF curve for future scenarios showed a significant increment. As observed from Table 4.5. The 25 year RP results for Singapore had the average intensity for all durations shown and 50 year RP) increasing by 12.51 % (RCP 4.5) and 38.76 % (RCP 8.5)

Over Jakarta for RP=50 year, both RCP 4.5 and RCP 8.5 had an increasing intensity trend with duration (Table 4.5). The absolute average intensity change (for all durations shown and 50 year RP) was observed 17.08% (RCP 4.5) and 50.78% (RCP 8.5). At 25 year RP, the absolute average intensity for all durations was observed to increase by 40.66% (for RCP 4.5) and 60.66% (for RCP 8.5)

Lastly as seen in Table 4.5, for Kuala Lumpur at RP=50 year, both RCP 4.5 and RCP 8.5 had an increasing intensity trend with duration (Table 4.5). The absolute average intensity change (for all durations shown and 50 year RP) was observed 3.56% (RCP 4.5) and 14.67% (RCP 8.5). At RP=25 year the absolute average intensity change (for all durations shown) calculated from Table 4.5 was 3.91% (RCP 4.5) and 14.47% (RCP 8.5)

Table 4.7 50-year return period CORDEX-IDF curve and the relative change intensity percentage corresponding to RCP 4.5 and RCP 8.5 at Singapore, Jakarta and Kuala Lumpur

50-year Return Period: CORDEX DATA					
Hour	Scenarios			Change in % to historical	
	Historical	RCP 4.5	RCP 8.5	RCP 4.5	RCP 8.5
Singapore					
1hr	137.13	140.74	139.43	2.64	1.68
2hr	88.67	93.04	93.14	4.93	5.05
6hr	39.34	43.38	44.25	10.27	12.47
8hr	31.33	34.97	35.93	11.62	14.68
12hr	22.58	25.75	26.71	14.04	18.3
18hr	16.19	18.79	19.66	16.07	21.43
24hr	12.76	15.1	15.88	18.32	24.43
48hr	7.15	8.84	9.43	23.61	31.94
Jakarta					
1hr	67	60.74	101.03	9.34	50.78
2hr	46	41.1	69.07	10.66	50.14
6hr	26	22.52	38.75	13.39	49.04
8hr	23	19.75	34.19	14.14	48.65
12hr	19	16.08	28.12	15.36	47.98
18hr	16	13.34	23.58	16.61	47.35
24hr	14	11.54	20.55	17.6	46.78
48hr	12	7.25	13.19	39.58	9.9
Kuala Lumpur					
1hr	96.4	104.9	116.91	8.82	21.28
2hr	58	62.24	69.42	7.31	19.69
6hr	24.8	25.85	28.85	4.25	16.33
8hr	22	22.66	25.32	3.01	15.07
12hr	14	14.26	15.91	1.85	13.65
18hr	10	10.05	11.21	0.5	12.06
24hr	7.9	7.85	8.74	0.63	10.63
48hr	4.4	4.3	4.78	2.17	8.7

Table 4.8 25-year return period CORDEX-IDF curve and the relative change intensity percentage corresponding to RCP 4.5 and RCP 8.5 at Singapore, Jakarta and Kuala Lumpur

25-year Return Period: CORDEX DATA					
Hour	Scenarios			Change in % to historical	
	Historical	RCP 4.5	RCP 8.5	RCP 4.5	RCP 8.5
Singapore					
1hr	118.07	133.54	157.98	13.1	33.8
2hr	76.35	86.23	103.01	12.94	34.92
6hr	33.87	38.13	46.55	12.59	37.44
8hr	26.97	30.33	37.28	12.47	38.22
12hr	19.44	21.85	27.1	12.4	39.4
18hr	13.94	15.66	19.61	12.32	40.69
24hr	10.99	12.33	15.57	12.2	41.63
48hr	6.17	6.92	8.89	12.09	44.04
Jakarta					
1hr	50	57.7	67.7	15.39	35.39
2hr	37	44.84	52.24	21.19	41.19
6hr	20	26.78	30.78	33.91	53.91
8hr	18	24.8	28.4	37.8	57.8
12hr	16	23	26.2	43.76	63.76
18hr	13	19.54	22.14	50.32	70.32
24hr	10	15.51	17.51	55.11	75.11
48hr	7	11.75	13.15	67.86	87.86
Kuala Lumpur					
1hr	83.9	91.76	75.35	9.36	10.2
2hr	50	53.93	45.93	7.86	8.14
6hr	21.6	22.6	18.6	4.63	13.87
8hr	19	19.68	16.82	3.59	11.46
12hr	12.2	12.5	10.24	2.42	16.04
18hr	8.8	8.9	7.19	1.1	18.3
24hr	6.9	6.9	5.65	0	18.16
48hr	3.8	3.71	3.06	2.35	19.6

4.4.3 Comparison between WRF and CORDEX derived IDF curves

It was observed that that extreme rainfall intensity increased for both WRF-IDFs and CORDEX-IDFs for all the three regions. The overall increments for WRF-IDFs and RCP 8.5 IDFs were larger than CORDEX-IDFs and RCP 4.5, respectively. For Singapore (100 year RP and RCP 8.5) average intensity increase was 36.10% for WRF-IDF at RCP 8.5 while it was 16.77% for CORDEX-IDF. Over Jakarta (100 year RP) the RCP 8.5 WRF-IDF showed 52.48% average increment while the WRF-IDF had a 73.98% increase in average intensity. However, over Kuala Lumpur, the intensity increase demonstrated by CORDEX-IDF was greater than WRF-IDFs for most of the durations. Over Kuala Lumpur (100 year RP) the RCP 8.5 CORDEX-IDF showed 15.96% average increment while WRF-IDF had a smaller 3.23% increase in average intensity. Overall the maximum intensity for WRF-IDF was observed for Singapore under RCP 8.5 as 208.05 mm/hr. While the maximum intensity for CORDEX-IDF was also noted for Singapore under RCP 8.5 as 161.73 mm.

4.5 Concluding Remarks

The primary objective of this chapter was to demonstrate an approach to derive future IDF curves from the outputs of two different RCMs. The construction of IDF curve was described for Singapore while the future changes and amplifications of IDF curves were described for all the three locations Singapore, Kuala Lumpur and Jakarta studied. The study also generated historical IDF curves that provided a baseline for downscaling of grid level IDF curves to station level. The author believes this is the first such study to apply a bias correction approach where a bias correction (from CFSR grid) and downscaling (from station data) is adopted and compared IDF curves from two different RCMs over the WMC.

The results presented in this study were based on two RCM models. In the subsequent work, more RCMs should be used to verify the results. An ensemble of many other

RCM derived IDFs could reduce the uncertainty and would represent a much better mean trend of future IDF curves within a region. Also, different GCM inputs can be input to a single RCM (e.g., WRF), and downscaled result ensemble can be analyzed, and this would further give more certainty to the analysis.

5 Cluster Analysis of Spatial and Temporal Monthly Precipitation over Western Maritime Continent

5.1 Introduction

Precipitation over the WMC is highly influenced by the prominent convection happening in the region. Also as the WMC has homogeneous air mass, and evenly distributed pressure and sea surface temperatures, the weather, and climate in this region rely heavily on local and mesoscale factors (Neale et al., 2003). The objective of this work is to identify the spatial and temporal patterns of monthly precipitation over this region, using Ward's clustering algorithm (Murtagh et al., 2014; Ward Jr, 1963). Ward's hierarchical clustering is one of the most appropriate ways of clustering for quantitative variables as it gives a clear guidance of what the number of clusters should be. Ward's clustering gives a hint about k , the number of clusters based on merging cost function (Shalizi, 2009). Furthermore, Blashfield (1976) discovered that Ward's method generated most accurate result among the 4-hierarchical clustering method tested by Blashfield (1976) and recommended it over other methods for hierarchical clustering.

Grouping of spatial sites having a similar climatology is a common practice done by meteorologists. Grouping of the sites provides a concise and structural summary of the climate data in the study region. The climatology of a region is also highly related to the precipitation pattern and distribution within a region. Clustering or Cluster Analysis (CA) is an effective method for data mining that helps to identify groups. The CA contributes to the identification of an unusual pattern or groups and links within the data. CA is recognized as an effective tool for climatology. Some of the early use of the method to identify a homogeneous grouping of station w.r.t to their climatology are DeGaetano et al. (1990), Ahmed (1997) and DeGaetano (2001).

Precipitation is also one of the primary meteorological variables in the WMC. Understanding precipitation patterns and grouping is an important aspect of mitigation measures to prevent floods and droughts as well as to understand redistribution of rainfall arising from climate change. Here it is noted that the redistribution refers to the redistribution or reallocation of rainfall stations amongst the derived spatial clusters under climate change forcing and not that of a fixed total amount of rainfall with a changed spatial distribution pattern. However, understanding rainfall over WMC is of great challenge, because of its complex topography and weather system. The WMC encompasses four large-scale islands- Java, Indonesia, Peninsular Malaysia, Borneo and most of them consist of a multitude of small islands many of which are mountainous.

Previous studies primarily grouped the climatology of these four major islands within the WMC based on the geographical location. For example, Peninsular Malaysia was divided into four regions, i.e., Northern, Central, Eastern and Southern by Varikoden et al. (2010) and Suhaila et al. (2011). The Northern region was reported to receive the highest rainfall from September to October, Central from October to November, Eastern from November to December and Southern from December to January. The Indonesian Bureau of Meteorology, Climatology, and Geophysics (BMKG) had classified Indonesia into 293 areas with different rainfall characteristics to reflect the complexity and heterogeneity of the rainfall distribution (<http://www.bmkg.go.id/?lang=EN>).

The goal of this work is to apply hierarchical clustering on the monthly precipitation data over WMC using Ward's algorithm. The methodology adopted for this study is shown in Figure 5.1. Monthly rainfall for over 43 approximately evenly distributed stations (locations) over four major islands of Java, Sumatra, Borneo and Peninsular Malaysia are used. The objectives behind clustering the rainfall data are: (i) Firstly, to discover spatial grouping of rainfall distribution WMC. (ii) Secondly, to uncover the

monthly and seasonal rainfall cycle for each cluster and (iii) understanding the future redistribution of rainfall in WMC using WRF-CMIP5 data. The author believes this is the first-time the CA approach is applied over WMC and with observations related to distribution and redistribution of rainfall w.r.t future IPCC forcing's patterns of precipitation over WMC.

Figure 5.1. A flowchart showing clustering framework in this study

The spatiotemporal cluster is analyzed and evaluated over WMC in this chapter. Section 5.2 outlines the study area and data used. Section 5.3 presents the methods used to achieve clustering. Section 5.4 shows Results and Findings for 5 Km CHIRPS dataset. Section 5.5 documents the Results for Cluster Analysis on 25 km CHIRPS and future WRF-CMIP5 Dataset to Investigate Redistribution of Rainfall. Finally, conclusions are presented in Section 5.5.

5.2 Study Area and Data

Study area

The WMC is a western part (90⁰E to 120⁰E and 10⁰S to 10⁰N) of the Maritime region defined by Rampage (Ramage, 1968; Neale and Slingo, 2003). Figure 5.2 shows the complex topography and mixed land use patterns over WMC. 43 homogeneously distributed “stations” (locations) were selected for the cluster analysis. Note that stations here refer to grid rainfall, though the grid is high resolution being 5 km for CHIRPS and 25 km for WRF-CMIP5. The stations were strategically chosen from the

4 major islands in the WMC, specifically from Sumatra (12), Java (7), Malaysia (10), Borneo (12) and Independent Islands (2). Table 5.1 categorizes the islands from where the stations are selected, coordinates and altitude of the stations which were the part of the study.

Figure 5.2 Spatial map showing a) topography and b) land use category, for the study region Western Maritime Continent

Dataset used

a) CHIRPS Dataset

The Climate Hazards Groups InfraRed Precipitation with stations (CHIRPS) dataset is a long period dataset that employs “smart” interpolation technique on infrared cold cloud distribution to estimate precipitation. CHIRP is built on 0.05° climatology that includes daily as well as monthly precipitation from the year 1981 to present. CHIRP was used as the observed data-set in this study, as it facilitated a 0.05° x 0.05° resolution data enabling a high-resolution study. Further, CHIRPS also have the long-time range (Jan 1981- Dec 2016).

b) WRF dataset

WRF-CMIP5 run for the future period 2030-2060 as driven with CMIP5 data (see Section 4.2.1) was used in this study. The physics parametrization was shown in Figure 4.3. Use of WRF model enabled the study to use a high-resolution dataset (WRF grids at 25 km) for explaining the future distribution and redistribution of precipitation.

Table 5.1 Stations, coordinates (latitude and longitude), and altitude of the locations used in this study.

Time range of the monthly data is from Jan 1981 to Dec 2016

number	Location	Latitude (°)	Longitude (°)	Altitude (m)	Mean Annual rainfall (mm/day)
Java					
1	Jakarta	-6.17	106.87	9	4.87
2	Bardung	-6.9	107.61	704	5.61
3	Surabaya	-7.25	112.75	9	4.66
4	Bali	-8.3	115.03	1491	7.38
5	Malang	-7.9	112.6	458	6.25
6	Tegal	-6.879	109.12	9	5.5
7	Kebumen	-7.6681	109.65	57	10.13
Sumatra					
8	Riau	0.2933	101.7	6	7.71
9	Jambi	-1.6101	103.613	45	5.9
10	Lampung	-4.55	105.4	32	7.45
11	Pagar alam	-4.04	103.22	794	7.54
12	Padang	-0.947	100.41	59	10.05
13	Tapan	-2.14	101.025	18	10.66
14	Pekanbaru	0.507	101.447	10	7.36
15	Aceh	4.6951	96.749	1061	4.94
16	Dumai	1.666	101.4	14	6.63
17	Palembang	-2.97	104.77	2	7.09
18	Babahrot	3.93	96.7	67	7.7
19	Tut tut	4.49	96.13	104	9.69
Independent Islands					
20	Singapore	1.3521	103.81	58	6.33
21	Pulau Bintan	1.136	104.42	27	7.28
Malaysia					
22	Kuala lumpur	3.13	101.68	59	7.26
23	Kluang	2.03	103.318	23	6.37
24	Mersing	2.43	103.83	9	6.84
25	Temerloh	3.448	102.41	59	5.79
26	Kampar	4.3	101.15	23	9.41

27	Gua Musang	4.88	101.96	87	8.84
28	Jitra	6.264	100.42	7	6.45
29	Langkawi	6.35	99.8	49	6.85
30	Phuket	7.95	98.33	326	6.71
31	Hat Yai	7	100.47	11	5.57
Borneo					
32	Singkawang	0.9	108.98	19	8.27
33	Labuan	5.28	115.23	27	8.78
34	Kapit	1.99	112.93	39	11.05
35	Amutai	-2.4166	115.23	16	7.22
36	Melak	0.286	115.82	8	8.92
37	Belaga	2.7	113.78	48	10.36
38	Beluran	5.62	117.13	97	10.88
39	Tarakan	3.327	117.57	25	10.46
40	Sintang	-0.137	112.81	579	9.87
41	Balai Beukuak	-0.48	110.38	203	8.99
42	Berapi	-2.25	111.75	70	8.48
43	Tewah	-1	113.7	65	9.56

5.3 Methods

5.3.1 Cluster Analysis (Ward's Method)

The method proposed by Ward (1963) the “Ward’s Hierarchical Clustering” or the “method of minimum variance” was used in this study. The method was used to group stations and their monthly rainfall distribution based on their similarity. The Ward’s method looks at cluster analysis as a variance problem. The Ward’s method computes the Euclidean distances to evaluate dissimilarity between the clusters. Equation (5.1) gives the formula for calculating Euclidian distance d_e :

$$d_e = \left[\sum_{j=1}^n (P_{a,j} - P_{b,j})^2 \right]^{0.5} \quad (5.1)$$

$P_{a,j}$ and $P_{b,j}$ are the quantitative variable j (monthly precipitation in this study) from station a and b respectively.

The Wards algorithm, when implemented on a dataset, establishes groups by minimizing the dissimilarity or the total sum of squares (TSS) (as defined in Equation 5.2). The final clusters have minimum TSS within each cluster. The Wards algorithm achieves this by minimizing TSS within each group and maximizing TSS across groups. The algorithm though calculates several clusters at a level when inter group similarity is maximized while intra group similarity is minimized.

$$TSS = \sum_{j=1}^n (P_{a,j} - P_{b,j})^2 \quad (5.2)$$

Ward's method, when implemented for precipitation, uses the coefficient of variation (CV) and the mean value for the monthly rainfall. The values for CV and mean are organized in two configuration matrix p_{np} . The matrix p_{np} has mean (or CV) in its rows(n) and the station(p) in the columns. These configuration matrices were used to identify clusters of similar stations. Further, rainfall within these clusters was analyzed to discover the monthly (i.e., temporal) patterns or cluster seasonality. For the second step, the mean (or CV) were arranged in rows for matrix p_{np} and months were setup in columns (p).

The Ward's clustering method used in this paper was implemented directly using the "stats" package in R (<https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html>). The basic algorithm for Wards clustering is following:

Basic algorithm

Compute the distance matrix between the input data points

Let each data point be a cluster

Repeat

Merge the two closest clusters

Update the distance matrix

Until only a single cluster remains

The method is applied over WMC that has a varied topography and land use patterns (see Figure 5.2) which controls the climate and variability of rainfall within the region. Figure 5.3 shows that the Ward's clustering starts with each point provided to the algorithm as a single cluster. At each step, the merging of the closest pair starts (based on the distance matrix), the pairing continues until only one cluster is left.

Figure 5.3 The steps involved in wards clustering (reproduced from cs-people.bu.edu/evimaria/cs565-12/lect5.pptx)

5.4 Results and Findings for 5 Km CHIRPS dataset

Figure 5.4 Spatial distribution of CHIRPS accumulated seasonal rainfall (mm) over WMC. During Dec, Jan and Feb (DJF, boreal winter), strong easterlies persist over the western Pacific. The main region of convection as originating from the East Indian Ocean, across the Maritime Continent and into the South Pacific Convergence Zone (SPCZ), is characterized by a band of westerly wind. Although the observed seasonal

mean low-level wind field is also mostly westerly during the boreal autumn, winter, and spring, there is considerable variability in the direction of the low-level zonal wind across the MC, particularly on small spatial scales around the coasts of the islands that make up the region. Multiple influences on the MC introduce complexities into its mean climate. DJF, the season which is the most convectively active have atmospheric models systematically underestimate precipitation over the MC. A relatively dry season with prevailing easterly winds is experienced during boreal summer, coinciding with the Asian summer monsoon over the Indian Subcontinent. The dry season peaks during. SON and MAM are regarded as transition seasons. With prevailing westerly winds (mean seasonal winds up to 6m/sec), is experienced in most of Indonesia and parts of Southeast Asia during the boreal winter, conditions are particularly conducive to rainfall along the coasts and mountain ranges as being exposed to the NE monsoon. However, because of significant variability within WMC, it necessary to study regions in specific groups and not as an overall Maritime region. The clustering technique would give more insights into this analysis.

Figure 5.4 Spatial distribution of CHIRPS accumulated seasonal rainfall (mm) over WMC averaged from the year 1980 to 2016,

5.4.1 Grouping in Space

As seen in Figure 5.5(a) for the dendrograms, the Wards CA discovered 5 different Spatial Groups (SG) from the 43 monthly station data location. Figure 5.5(b) shows a spatial map where each group is shown with the same color. SG5 was the most populated group and had 15 members in it. Most of the members were located north of the equator, over North and N.E Sumatra and Southern Malaysian Peninsular. The cluster was also present over some of the higher topography regions (see Figure 5.2), such as Aceh and Babahrot in Sumatra, Bardur in Indonesia and Singkawang in Borneo. SG5 thus completely dominated the region central-east Sumatra and southern

Malaysian peninsular, and including the independent islands lying in between. This group had an annual mean rainfall (mean over all the stations within the group) in the range of 5 to 8 mm/day (see Table 5.1).

The second most populated group was SG2 (see Figure 5.5(a)) with 11 members within it. The disparity of this group from other groups was evident as 7 out of 11 members in this cluster were mainly located on Borneo Needle leaf forest (see Figure 5.2(b) and Figure 5.5) and other three locations were Kebumen, located at the Western-coast of Java and Padang, Tapan and Tut Tut at the western-coast of Sumatra (Figure 5.5(b)).

The third most populated is SG3 (see Figure 5.5(a)). This encompasses most of the rainfall stations in southern Sumatra and almost all the stations in Java. There was also one station Amutai from Borneo and none from Peninsular Malaysia. All the stations in SG3 were located south of the equator.

Lastly, the least populated were found in SG1 and SG4. SG1 was found to have 5 members while SG4 had 3 members. Three members of SG1 were located on the eastern coast of Borneo in the proximity of Gunung Kinabalu mountain ranges with higher topography. The other two members were in the central region of Peninsular Malaysia near Gunung Than mountains. Lastly, the members of SG4 are in the northern Malaysian peninsular coastlines (Figure 5.5).

(a)

(b)

Figure 5.5 Cluster Analysis groups (Group1 – Group 5) (a) dendrograms based on their spatial similarity and (b) spatial representation of the spatial clusters.

5.4.2 Monthly Rainfall Over Spatial Clusters

Further to the spatial grouping, temporal grouping (TG) of the monthly distribution of rainfall over the SGs were studied. Box plots in Figure 5.6 show the monthly distribution of rainfall (over Jan 1981- Dec 2016) for each SG along with variations in the monthly distribution of rainfall. For example, as seen in Figure 5.6(a), a uniform rainfall over the months was observed for SG1. The mean rainfall over SG1 is highest compared to other groups as these locations were located near mountain- sheds that received heavy rainfall. The larger Inter Quartile Range (IQR) of box plots shows that Jan, Feb, Mar, and Apr were seen to have the largest variability (over the year) of rainfall compared to other months (with more compact IQR). The highest mean rainfall was observed for Oct while the lowest amount mean rainfall was observed for February. No clear trend of the dry and wet season was observed for SG1.

In contrast, SG2 monthly boxplots (see Figure 5.6(b)) had a clear trend of wet and dry seasons. The stations within SG2 received a mean rainfall greater than 10 mm/day for Jan-Apr as well as from Oct-Dec. However, May-Sep was comparatively dry as the mean rainfall was below 10mm/day. The lowest rainfall for SG2 was observed for Jul, while the highest rainfall was seen for Nov. Also, a noticeable difference between SG1 and SG2 monthly precipitation boxplots was in the monthly variability (see Figure 5.6(a and b)). The IQR for boxplots were more homogenous for SG2 than SG1 with SG1 had higher variability for months Jan to Apr.

SG3 also showed a clear pattern of seasonality (Figure 5.6(c)). The Dec-Jan-Feb(DJF) was the wettest for this SG. A decline in rainfall was observed from Mar to Aug, with Aug being the driest month. The second driest month was Sep. However, an acceleration in daily precipitation intensity was observed from Sep with a continuous increase over Sep to Feb. This SG had less variability for most of the months except the months for which acceleration in precipitation was observed (Oct-Nov-Dec). The

low variability in the cluster signifies a more homogenous cluster while a larger variability indicates the presence of different elements within the cluster.

Interestingly, the monthly rainfall in SG4 was found to be opposite to SG3. The months where SG4 saw a decline in rainfall, had an increase of rainfall for SG3 and vice versa (see Figure 5.6(c and d)). Sep was the wettest month for SG4 while Jan was the driest month. SG4 is also quite different from other SGs in that SG4 was arid over DJF while other SGs were observed to receive a good amount of rain over DJF however,

Lastly, SG5 was the most populated SG was analyzed. Though this group had more stations, it had a lesser variability (more compact IQR) compared to other groups such as SG1 and SG2. The small variability signified that the station in SG5 had a very similar climatology. Nov and Dec were the wettest months for the SG5 while JJA were the driest months for the group (see Figure 5.6(e)).

The analysis of temporal clustering of the five SGs would also be informative towards the understanding of the temporal distribution within each SG done next in Section 5.4.3. The TGs would give further insights into the seasonality experienced by each of the SG.

a) SG1

b) SG2

c)SG3

d)SG4

e) SG5

Figure 5.6 The box-plots denote the 25% and 75% of the monthly rainfall(mm/day) for the Spatial Groups 1 to 5. (a to e)

5.4.3 Temporal Clusters

To understand seasonality (e.g., w.r.t. to dry and wet seasons), a temporal cluster analysis was applied to each spatial cluster using the Ward's method of clustering. Figure 5.7 shows a dendrogram for the TG within each SG discovered. The subsequent analysis is aimed at understanding the TGs achieved from monthly rainfall records and to identify wet and dry seasons.

Figure 5.7(a) shows the TG dendrogram for SG1. As seen in box plots in Figure 5.6(a) that SG1 had a uniform rainfall over all the months and identifying seasonal variation would be difficult. However, Ward's algorithm found four different seasons in SG1. The first TG was formed for the month Mar and April (TG1), these two months had high variability and an increasing trend w.r.t their previous month. The TG2 was established from Jan and Feb; these two months had high variability and a decreasing trend of rainfall during their preceding month see Figure 5.6(a). TG3 composed of JJA, this season had a lesser variability and an increasing trend. The most populated season was TG4 encompassing from months May and Sep to Dec.

a) SG1

b) SG2

c) SG3

d) SG4

e) SG5

Figure 5.7 Cluster analysis applied to monthly rainfall for SG 1 to 5 based on temporal grouping for each SG.

Contrary to SG1, SG2 has just two distinct TGs. The TGs for SG2 (see Figure 5.7-b)) were formed from months May to Sep (TG1) while Jan to Apr and Nov, Dec were in

TG2. As seen from Figure 5.6(b) the months within TG1 showed characteristics of the dry season while TG2 months the wet season.

Three different TGs within SG3 were discovered. TG1 had months from May to Oct. TG2 grouped Jan and Feb, and TG3 was formed from Mar, Apr, and Nov, Dec. Within the SG3 TGs, TG1 was associated with the dry months, TG2 with the wet months while TG3 was identified as a transition period.

SG4 (see Figure 5.7-d) comprised four distinct TGs. May, Jun, Jul (TG1) and Aug to Oct (TG2) were the wet period for SG4, with TG2 receiving more rainfall compared to TG1. Dec to Feb (TG3) was the extreme dry months with precipitation less than 2 mm/day (see Figure 5.6(d)). The final cluster TG4 was formed from the months Mar, Apr and Nov. TG4 was identified to represent transition period for SG4.

Finally, in the most populated SG5 was composed of three distinct TGs. Dec and Jan (TG1) were identified as the wettest period while Jan, Oct, Apr, Mar (TG3) was designated as another wet period. TG2 composed of months Feb, Aug, and May to Sep and was identified as the dry period for SG5.

5.5 Results for Cluster Analysis on 25 km CHIRPS and future WRF-CMIP5 Dataset to Investigate Redistribution of Rainfall

The Ward's clustering method described in Section 5.3.1 was further implemented on the WRF simulation driven using CMIP5 GCM over the year 2030 to 2060. This WRF data is same as that used in Chapter 4. However, aggregation from hourly to monthly was done as per the monthly CA is in this chapter. The spatial resolution of the WRF monthly precipitation data was 25 km x 25 km. Hence unlike 5 km CA using CHIRPS data in the previous section, this part of the study was done for 25 km grids around the same locations listed in Table 5.1. The clusters formed using the WRF data needed to be evaluated with an observed historical data to monitor changes. The CHIRPS data

was then bilinearly averaged (to 25 km) from 5 km grid resolution to match the WRF resolution.

Quantile Mapping of the WRF dataset

The gridded WRF-CMIP5 data used to derive IDF curve in Chapter 4 was aggregated monthly for CA in this study. However, as the comparison for WRF-CMIP5 was done with historical, it was important to normalize the model biases for the future data extracted from WRF. The bias correction of 25 km monthly WRF-CMIP5 model using the Gamma Quantile Mapping or GQM (Piani et al. 2010). The GQM bias correction method is based on the initial assumption that both observed and simulated intensity distributions are well approximated by the gamma distribution, therefore is a parametric q-q map that uses the theoretical instead of the empirical distribution.

The GQM method is an efficient bias correction method as it considers both mean and extreme rainfall method while bias correcting rainfall (Wilcke. et.al 2013; Amengual et.al 2012; Piani et.al 2009). It was observed that GQM filters out the extreme values that are results of errors while keeping the extreme values that are genuine (Yuanchao Xu, 2016). This makes GQM preferable over the conventional quantile mapping and scaling methods that do not distinguish between the extreme values resulting from the noise or a result of extreme climate conditions (Gutjahr et. al. 2013). However, the accuracy of GQM model is highly dependent on the time period over which the data is trained, i.e., larger the training period; more is the accuracy. In this study, the monthly historical data extracted from WRF was bias corrected with historical 25 km CHIRPS data using GQM method for a time period of 26 years (1980-2005). The bias correction factor achieved from the historical datasets were mapped to the WRF-CMIP5 future dataset. Hence in this chapter WRF-CMIP5 stands for bias corrected WRF-CMIP5 dataset.

5.5.1 Spatial Grouping of 25 km Historical CHIRPS data

The CHIRPS data was bilinearly averaged to 25 km, and Wards CA applied to the 42 grids that surrounded the locations listed in Table 5.1. This provided a historical CHIRPS rainfall to compare with the future scenarios (RCP 4.5 and RCP 8.5) driven WRF-CMIP5 data. For a fair comparison, only three SGs close to height 300 (distance between clusters) were compared in this study. This is because three SGs were formed corresponding to height 300 for all the datasets and thus a fair comparison was achieved.

As seen in Figure 5.8 (a), three SGs, i.e., SG1, SG2, SG3 have formed from the CHIRPS 25 km data. SG1 consist of a maximum number (25) of members while SG2 has a minimum number of members (4) and SG3 had 13 members. The distribution of the clusters on a spatial map is shown in Figure 5.8(b). It is observed from Figure 5.8(b) that the most populated cluster SG1 encompasses all the grid stations over Borneo. Also, Borneo having the largest tropical rainforest (Figure 5.2b) receives the maximum rainfall compared to all other regions in WMC. Beside Borneo, SG1 also captured grid stations with similar land use patterns over Sumatra and Malaysia that were near Borneo landmass (see Figure 5.2b). It was observed that most the of stations within SG1 had similar land use pattern, i.e., rainforest except the three grids over Java, i.e., Jakarta, Bardung and Kebumen that has an urban land use pattern (see Figure 5.2b). The possible similarity of these stations with other members of SG1 is the high intensity of rainfall received by this region throughout the year as they lie towards the windward side of mountains (see Figure 5.4a and Figure 5.2).

SG2 consist of four grid stations that lie over Java landmass (Figure 5.8). All these stations lie near the western coast of Java and had urban land use pattern. These locations received moderate to heavy rainfall depending on the season. DJF is the wettest season for SG2 (see Figure 5.4)

(a)

(b)

Figure 5.8 Cluster Analysis groups (SG1 – SG3) for CHIRPS historical (25 km) (a) dendrograms based on their spatial similarity and (b) Spatial representation of the clusters on a map

Finally, SG3 the second most populated cluster mostly included the grid stations north of the equator. Most of these stations were located over Peninsular Malaysia while the

other points were over Sumatra (see Figure 5.8). Almost all the stations in SG2 had a proximity to the sea. Unlike SG1 and SG2, SG3 experiences maximum rainfall for SON season (see Figure 5.4).

5.5.2 Spatial Grouping for RCP 4.5 WRF-CMIP5 data (2030 – 2060)

SGs corresponding WRF-CMI5 (RCP 4.5) were observed to be different compared to CHIRPS (historical) SGs (see Figure 5.8 and Figure 5.9). The three clusters (see Figure 5.9a) were SG1, SG2, SG3. SG2 was the most populated member with 19 members. However, SG1 had 18 members. SG3 was the least populated member with five members.

SG1 for WRF-CMIP5 (RCP 4.5) was similar to SG1 of CHIRPS-25km (historical) in terms of SGs and their TGS. However, SG1 for WRF-CMIP5 (RCP 4.5) gained some members over Java in comparison to SG1 of CHIRPS-25km (historical) making Java homogeneous and prevailed by SG1. Furthermore, SG1 for WRF-CMIP5 (RCP 4.5) lost some of its members in Borneo in comparison to SG1 of CHIRPS-25km (historical) that encompassed all the grid-stations over Borneo. The stations over Borneo by SG1 was lost by a new pattern of SG3 that emerged over the eastern part of Borneo.

SG2 for WRF-CMIP5 (RCP 4.5) was similar to SG3 of CHIRPS-25km cluster. However, it was seen that SG2 of WRF-CMIP5 (RCP 4.5) gained more stations over Sumatra compared to its similar counterpart SG3 of CHIRPS-25km (historical).

Lastly, SG3, the least populated group was constrained to the area under eastern Borneo.

Figure 5.9 Cluster Analysis groups (SG1 – SG 3) for RCP 4.5 (a) dendrograms based on their spatial similarity and (b) SG representation on a map

5.5.3 Spatial Grouping for RCP 8.5 WRF-CMIP5 data (2030 – 2060)

The SGs for WRF-CMIP5 (RCP 8.5) was observed to be very similar to SGs for WRF-CMIP5 (RCP 8.5) for Java, Sumatra, and the Malay Peninsula. However, a significant difference was seen over Borneo. A new pattern SG3 emerged over eastern which encompassed the entire Borneo landmass by replacing SG1 patterns over Borneo.

SGs corresponding WRF-CMI5 (RCP 8.5) were observed to be different compared to CHIRPS (historical) SGs (see Figure 5.8 and Figure 5.10). Three clusters Figure 5.10a were SG1, SG2, SG3 with SG2 was the most populated member with 20. SG1 and SG3 both had 11 members within them.

SG1 for WRF-CMIP5 (RCP 8.5) was similar to SG1 of CHIRPS-25km (historical). However, SG1 for WRF-CMIP5 (RCP 8.5) gained some members over java in comparison to SG1 of CHIRPS-25km (historical) making Java more homogeneous and prevailed by SG1. Furthermore, SG1 for WRF-CMIP5 (RCP 8.5) lost all except one member over Borneo in comparison to SG1 of CHIRPS-25km (historical) that encompassed all the grid-stations over Borneo.

SG2 for WRF-CMIP5 (RCP 8.5) was similar to SG3 of CHIRPS-25km cluster however it was seen that SG2 of WRF-CMIP5 (RCP 8.5) gained more stations over Sumatra compared to its similar counterpart SG3 of CHIRPS-25km (historical).

SG3, group was mainly constrained over locations in eastern Borneo, with all except for one member being in Borneo.

(a)

(b)

Figure 5.10 Cluster Analysis groups (SG11 – SG3 3) for RCP 8.5 (a) dendrograms based on their spatial similarity and (b) spatial representation of the clusters on a map.

5.5.4 Temporal Groupings of 25 km Historical CHIRPS data

Once the SGs were discovered, Ward's CA was next applied to the monthly rainfall data for each SG to find TGs or seasonality for each SGs. Box plots were used to represent the monthly intensity as seen in Figure 5.11(a). TGs for every SGs is shown in Figure 5.11 (b).

TGs for SG1

Two seasons TG1 and TG2 were predicted for SG1 (Figure 5.11b). TG1 was a dry season for SG1 with all the months with rainfall median less than 10 mm/day. TG2, on the other hand, was the wet season for SG1 with rainfall median of 10mm/day or more for all the months within TG2.

TGs for SG2

Three temporal clusters were formed for SG2. TG1 depicted a dry season over SG1 with rainfall intensity less than 8 mm/day. TG2 composed of two months Apr and Nov are the transition months between wet to dry and dry to wet season respectively. Lastly, TG3 depicted rainy season over SG2 with rainfall intensity greater than 8mm/day over the months (see Figure 5.11b).

TGs for SG3

Three TGs were classified for SG3. TG1 was formed from six months having their rainfall IQR below 10 mm/day. Two of the wettest months, i.e. Jan and Dec for the other two TGs were observed to have the lowest median for TGs in SG3. Also, the variability in Jan and Dec rainfall was found to largest in comparison to other months.

Figure 5.11 The box plots (left) and Dendrograms of TGs (right) for different 3 SGs discovered in the CHIRPS historical data (25 km)

5.5.5 Temporal Grouping of RCP 4.5 WRF-CMIP5 data (2030 – 2060)

The TGs corresponding to RCP 4.5 WRF-CMIP5 data for all SGs is shown in Figure 5.12.

TGs for SG1

Three TGs were discovered for SG1. TG1 (as seen in Figure 5.12) corresponded to the dry season within SG1. TG2 appeared to be the transition period between wet and dry months, i.e., between DJF and May to Sep with the transition month in TG2 of April while Oct and Nov is a transition period from dry to rainy season, i.e., from May to Sep and DJF. Finally, TG3 corresponds to the rainy season with median rainfall close to 10 mm/day (see Figure 5.12).

TGs for SG2

Two distinct TGs were found for SG2. As this SG was comparatively dry to other SGs a median below 10mm/day was observed for most of the months (Figure 5.12b). The classification was based comparatively on wet and dry seasons. As seen in Figure

5.12b TG1 encompassing months from Dec to March while other months were part of TG2.

TGs for SG3

The rainfall for the TGs corresponding to SG3 was uniform and steady (see Figure 5.12). The two clusters formed for SG2 had a very similar characteristic to each other. As seen in boxplot in Figure 5.12b, the minimum value for the TG1 cluster was lower compared to the TG2 cluster. Also, all the members of TG1 cluster had higher variability with median equal or less to 10 mm/day.

Figure 5.12 The box plots (left) and Dendrograms of TGs (right) for different 3 SGs discovered in the WRF-CMIP5 (RCP 4.5)

5.5.6 Temporal Grouping of RCP 8.5 WRF-CMIP5 data (2030 – 2060)

The TGs corresponding to RCP 8.5 WRF-CMIP5 data for SGs is shown in Figure 5.13.

TGs for SG1

As observed from SGs for RCP 4.5 and RCP 8.5 (see Figure 5.9 and Figure 5.10) members in SG1 for RCP 4.5 WRF CMIP5 were similar to SG1 for RCP 8.5 WRF

CMIP5. Then as to be expected, three TGs were formed for SG1 (RCP 8.5) as for SG1 (RCP 4.5). However, interestingly as seen in Figure 5.13a the monthly rainfall intensity median for SG1 (RCP 8.5) was close to the median for months with SG1 (RCP 4.5).

TGs for SG2

A significant change in the median monthly rainfall and TGs were observed for the TGs between SG2 (RCP 4.5) and SG (RCP 8.5) (see Figure 5.13). Also, SG2 (RCP 8.5) was discovered to have four TGs compared to two in SG2 (RCP 8.5). TG1 here depicted two of the driest months for SG2 while TG2 appeared to be the transition intervals. TG3 was composed of months that were dry but had smaller IQR (variability) compared to TG1 dry months. Finally, TG4 was a grouping of the wettest months within SG2.

TGs for SG3

SG3 was the most dominant cluster over Borneo as it grouped most of the stations in the Borneo area. The most significant difference between SGs (RCP 8.5) and SGs (RCP 4.5) was seen because of SG3 and especially over the Borneo region. Though the monthly boxplots for SG3 (4.5) and SG3 (8.5) are similar, the TGs formed are now entirely different. SG3 (RCP 8.5) has 3 TGs while SG3 (RCP 4.5) had 2 TGs. TG1 encompassed the wet months within the cluster with median close to 10 mm/day and with a large IQR. TG2 had a comparatively lesser IQR compared to TG1 and was composed of wet months. TG3 was the dry season for SG3.

Figure 5.13 The box plots (left) and Dendrograms of TGs (right) for different 3 SGs discovered in the WRF-CMIP5 (RCP 8.5)

5.6 Summary

A detailed cluster analysis over 43 stations locations in WMC using CHIRPS (5 km) data and for 42 stations using Chirp (25km), WRF (resolution 25km for each of RCP 4.5 and RCP 8.5) was done in this study. Two different types of cluster analysis, i.e., a) spatial cluster analysis b) temporal cluster analysis was studied and reported. The spatial cluster analysis was too implemented to discover different spatial patterns of precipitation over the WMC regions via SGs. In the further analysis, the SGs was further analyzed using temporal cluster analysis to discover different TGs within each SG. This study presented both the distribution of monthly rainfall using high-resolution CHIRPS (5 km data) over WMC along with redistribution of rainfall that is likely to be a result of climate change using WRF-CMIP5 (RCP 4.5 and RCP 8.5) data and CHIRPS (25 km data).

5.6.1 Distribution of rainfall using Cluster Analysis of 5 Km CHIRPS dataset over WMC

The work here relates the clustering patterns with other meteorological variables as well as topography and land uses within the region. 37 year of monthly precipitation data over 43 approximately evenly distributed stations (locations) over four major large islands: Java, Sumatra, Borneo and Peninsular Malaysia was used. Overall, five major spatial rainfall clusters were identified within the region using the Ward's algorithm. The stations were mostly clustered with other stations within the same Island. Other clusters were also observed, where the stations rather than grouping to the nearest stations, grouped with stations having similar topography and vegetation. This study discovered five different SGs across the WMC region based on the monthly rainfall data. It was observed that the stations within an individual SGs were highly dependent on which among the four landmasses the grid stations were located. The Java island had a homogeneous mix of 2 SGs a) located just west of the mountains (Mt. Salak and Mt. Gede) and b) located near the eastern coast with an urban land use pattern. It was observed that the most dominant SG in Java, SG3 was confined to the south of the equator. These stations also had similar characteristics to the coastal stations in Java island and the latter were part of SG3. In Sumatra, a clear difference was seen between the stations lying left to the Barinas mountain range. The most dominant cluster was SG5 which was mainly present in northern Sumatra, the Independent islands near Sumatra and southern Peninsular Malaysia. SG1 formed in central Peninsular Malaysia and eastern Borneo. These stations had similar topography and land use patterns. SG4 had the minimum number stations; all these were from northern Malaysia with most in coastal regions.

Most of the studies considered standard seasons as based on the prevalent wind directions. The temporal clustering, however, revealed that WMC is comprised of the SGs that have different seasonality. Differences in seasonality for the SGs was

observed. For example, the seasonality of SG3 was found to be opposite of in SG4, i.e., the wet period in SG3 corresponds to the dry period in SG4 and vice versa.

A further study on this region that could consider clustering of other meteorological variables such as wind, temperature, and specific humidity. These could further reveal the reasons behind the different SGs and the TGs within these SGs.

5.6.2 Redistribution of Rainfall over WMC

This same study also examined the retribution of SGs using the 25 Km CHIRPS (historical) and the WRF-CMIP5 (RCP 4.5. RCP 8.5) for the year 2030-2060. The dataset consists of monthly data for 42 grid stations distributed across four major land masses within WMC. The results showed a noteworthy difference in SGs under future rainfall from RCP 4.5 and RCP 8.5 as compared with CHIRPS (historical- 25 km) and with a further difference in TGs. For all the SGs an increment in median rainfall for RCP 4.5 and RCP 8.5 WRF-CMIP5 when compared to CHIRP (historical) was observed.

Table Table 5.2 shows a clear trend in redistribution of SGs over WMC is observed in this study. Individual landmasses over WMC would tend to have a more homogeneous rainfall distribution in the future (as subjected to RCPs projected). Apart from Sumatra, all the landmasses would mostly have a single SG under RCP 8.5. Though Sumatra was consistent with the number of SG within its spatial boundary, a change in characteristics, i.e., TGs and intensity of monthly rainfall were observed. Thus, a change in internal as well as overall structure of SG for monthly precipitation is noted in this study

Table 5.2 The distribution of SGs over four landmasses for Historical and future time period (RCP 4.5 and RCP 8.5)

Landmass	Number of SGs (with minimum than two locations in a SG)		
	Historical	RCP 4.5	RCP 8.5
Sumatra	2	2	2
Java	2	1	1
Borneo	1	2	1
Malaysian Peninsular	2	1	1

Comparison of RCP clusters over Borneo

Table 5.3 shows the distribution of locations among the SGs for each landmass covering historical and the future conditions. Borneo of all the landmasses experienced the most significant changes going from RCP 4.5 to RCP 8.5. For both the RCPs locations in Borneo either belonged to SG1 or SG3 (see Table 3). However, it was observed that for RCP 4.5 the number of locations in SG1 (SG3) was six (five), whereas for RCP 8.5 SG1 (SG3) changed to one (ten). This signified that the stations had almost equal weightage within SG1 and SG3 for RCP 4.5 but shifted to SG3 for RCP 8.5. Thus, these observations confirm that SG over Borneo becomes homogeneous for RCP 8.5.

Table 5.3 Distribution of locations among the clusters over the five different landmasses in WMC for RCP 4.5 and RCP 8.5 monthly rainfall data

	SG1		SG2		SG3		
Island	RCP	RCP	RCP	RCP	RCP	RCP	Total stations
Landmass	4.5	8.5	4.5	8.5	4.5	8.5	per
							Landmass
Sumatra	4	3	8	8	0	1	12
Java	7	7	0	0	0	0	7
Malaysia	1	1	9	10	0	0	10
Borneo	6	1	0	0	5	10	11
Independent							
Island	0	0	2	2	0	0	2

Further studies in this area can be done by considering other RCMs and at a higher resolution (e.g., 5 km). The study at a higher resolution would reveal a more detailed redistribution patterns over WMC. The use of different RCMs would make the finding more robust.

6 Summary and Conclusion

6.1 Summary

Before concluding on key findings of this Ph.D. research thesis, it is useful to recall three major challenges faced by climate researchers and policy makers over the WMC. The first is the complex atmospheric system that leads to significant uncertainties when downscaling rainfall. The second and the third challenges are to model the changes in rainfall intensity and redistribution because of climate change. This study makes use of RCM data (WRF and CORDEX in this study) to addresses these issues via three different studies.

Firstly, in Chapter 3 the uncertainties in daily precipitation are reduced by employing an integrated downscaling approach that uses predictors from high-resolution WRF output rather than coarse resolution GCM predictors. This method was compared against the conventional statistical downscaling approach. It was observed that the integrated downscaling approach simulated precipitation which had mean and extreme statistics closer to the observed value in comparison to conventional downscaling approach.

The second part which is in Chapter 4 presents the past, and future RCM-derived and bias-corrected IDF curves at three stations over the WMC. The future RCM-IDF curves were simulated considering the impact of climate change. The study compared two RCM derived IDF curves specifically from WRF and CORDEX data for RCP 4.5 and RCP 8.5 scenarios. The results provide estimates of the increment in future extreme precipitation because of climate change.

The third part of the study presented in Chapter 5 used agglomerative clustering method specifically Ward's method to model distribution and redistribution of

monthly rainfall over the WMC. Clustering of precipitation over the WMC using 5 km CHIRPS data for 43 locations was first performed. Five different SGs for 43 approximately evenly distributed stations over WMC was discovered. Further to this, the seasonality within each of these SGs was examined. Clustering was then applied to the 42 grid locations of WRF-CMIP5 (for RCP 4.5 and RCP 8.5) data over the year 2030-2060. The spatial cluster for the future monthly precipitation was discovered and compared against historical CHIRPS data (bilinearly averaged to 25 km) as well as between the two scenarios. The study revealed changes in the distribution of rainfall and intensity within each SG.

The following Section 6.2 to Section 6.4 elaborates on the conclusions from these three parts of the Ph.D. work.

6.2 Integrated Dynamical-Statistical Approach to Downscale Precipitation over WMC

This part of the study is on the development of an alternate and an efficient downscaling technique as demonstrated at four stations over WMC. An integrated downscaling approach is developed using predictors from WRF model as input in SD rather than using GCM or Reanalysis predictors. Chapter 3 gives the details about the procedures and the results for the approach with calibration and validation performed at four stations in WMC, i.e., Singapore, Jakarta, Labuan and Hat Yai. To the author's knowledge, this is the first study to apply RCM predictors to improve SD over WMC.

6.2.1 Results from Downscaling of daily precipitation

Reanalysis R2 data was used to drive RCM-WRF model over WMC domain. TRMM data was used to evaluate the performance of the WRF model and particularly for the

combination of physics parametrization used. The downscaled WRF data over the year 1980-2005 was then used as the source of predictors for SD. The selection process used “decision tree” and correlation coefficient with daily precipitation. A similar procedure for extraction and selection of predictors were also done for R2 data. These significant predictors for both WRF and R2 data along with station data were provided to ASD model to statistically downscale precipitation. The downscaled output daily precipitation from both WRF and R2 predictors were compared against each other using four hydrological statistics, i.e. (a) mean, (b) standard deviation (c) wet days (d) excess kurtosis that represent important characteristics of daily rainfall.

Boxplots for the four hydrological statistics depicted that the integrated downscaling approach performed better in capturing the observed data with the IQR of the box plots being tighter in range as compared to the conventional method. Further, improvements in the error metrics (R^2 , bR^2 , RMSE, NSE) for these statistics showed that the proposed approach had lesser uncertainty. The integrated downscaling approach was also found to be more accurate in reproducing the mean and extreme rainfall over different seasons.

6.2.2 Recommendations for Future Studies

The results provided a proof of concept using historical data. The study can be more useful when applied to the future climate to study changes in daily as well as seasonal rainfall with greater accuracy. For example, the integrated downscaling approach can be employed with GCM data wherein a dynamically downscaled GCM data predictors are inputted to SD models.

Beyond regional downscaling a further wavelet analysis of e.g. temperature and precipitation data can also be performed over the WMC to identify dominant

oscillations of these climate variables at a regional scale such as performed for southwestern Canada (Gan et al., 2007; Gobena et al., 2006).

6.3 Future IDF Curves using RCM data over WMC

Comparison of IDF curves derived from two RCMs, WRF and CORDEX is presented in this part of the study. This study aims to assess the non-stationarity induced because of climate change while constructing IDF curves. The derivation of the future IDF curves was demonstrated for three stations, i.e., Singapore, Jakarta, and Kuala Lumpur. This study is useful as two different RCM derived IDF curves are compared to the future climate over typical stations in WMC.

6.3.1 Calculating bias and downscaling factors using historical IDF Curves

The approach of bias correcting RCM derived IDF curves using grid (CFSR) and station (observed) IDF curves is implemented in this study. Two important steps involved bias correction using the historical IDF curves. Firstly, the historical IDF curves corresponding to various selected durations and return periods was scaled to match with historical CFSR-IDF, this being a bias correction as the scaling was done between two grid data. The scaling factor achieved was termed as bias correction factor. Then the bias-corrected RCM-IDF was scaled to match the station IDF for each selected return periods and durations, this step being downscaling. The scaling factor achieved from this step was called downscaling factor.

Figure 4.9 as an example, demonstrates the bias correction step where the historical WRF-CMIP5 data was biased corrected based on historical CFSR IDF curve (the year 1980 to 2005). Before bias correction, the WRF-CMIP5 underestimated the rainfall intensity by 38% (upper bound) for one hr. duration and 29 % (lower bound) for 24 hr. duration at 100 year RP. Further, Figure 4.10 demonstrated the downscaling step

for IDF curves where for 100 year RP the upper bound was 54 % (1 hr. duration), and the lower bound was -18% (24 hr. duration). A similar process was successfully repeated to achieve the bias correction and downscaling for historical CORDEX derived IDF curves.

6.3.2 Bias-corrected future IDF curves

Based on the bias and downscaling factors derived using historical rainfall, future IDF curves derived from WRF-CMIP5 and CORDEX were bias adjusted and downscaled for the three stations over WMC. The results obtained were station level IDF curves under RCP 4.5 and RCP 8.5 scenarios for the year 2030 to 2060. Both RCM-IDFs showed a rise in rainfall intensity corresponding to all the RPs and durations. However, the increment was observed to be highest for Jakarta followed by Singapore and lastly Kuala Lumpur. Furthermore, the increments for WRF-CMIP5-IDF was observed to be greater than CORDEX-IDF. Finally, over all the locations and RPs, the IDF curved derived for RCP 8.5 predicted higher rainfall intensity compared to RCP 4.5.

6.3.3 Recommendation for Future Work

The study can be made detailed and robust by considering more GCM inputs to RCMs and RCMs outputs. Further, the study can be expanded to more stations over WMC and other regions to validate further the bias correction approach and the results of climate different climate change forcing scenarios. Lastly, a more robust bias correction can be adopted in the future studies to improve the results.

6.4 Cluster Analysis of Spatial and Temporal Monthly Precipitation over WMC

The third part of the Ph.D. work is to apply Ward's clustering method (agglomerative clustering) over the past and the future grid data over WMC to discover different spatial and temporal groups for monthly rainfall. CHIRPS high resolution (5 km) data was used to study the historical rainfall distribution at 43 stations (locations) in WMC. The study revealed interesting clusters of rainfall characterized by their landmass, topography and land use. The historical time range of this study was from year Jan 1980 to Dec 2016.

Wards clustering was next applied on future monthly rainfall extracted from RCM data to examine the redistribution of rainfall resulting from climate change. WRF-CMIP5 was used for the future precipitation data under RCP 4.5 and RCP 8.5 scenarios over the year 2030 to 2060. The outputs were aggregated monthly and provided as an input to the Ward's algorithm to output SGs and their TGs corresponding to the future climate forcing. CHIRPS data bi-linearly averaged to 25 km was used to assess the changes in SGs and TGs from historical.

6.4.1 Cluster analysis using 5 km CHIRPS data from Jan 1980 to Dec 2016

The Clustering approach discovered five distinct SGs for 43 locations over WMC. Each of these clusters was represented on the dendrogram which represented the height (i.e., the distance) between the clusters. Further, the clusters were plotted on a spatial map to show the distribution of these clusters on the four major land masses in the WMC, i.e., Sumatra, Java, Peninsular Malaysia and Borneo. It was found that the classification of a location in a specific SG was primarily dependent on the landmass that it belongs too. Apart from the landmass, secondary factors comprising topography, land use and vicinity from the sea controlled the grouping of the members

within a cluster. Once the SGs were achieved, monthly rainfall from each SG was further inputted to Wards clustering to discover TGs within the SGs. The TGs showed different seasonality within each SG i.e. the dry months and wet months found to be partially or completely different across the SGs. Box plots were used to assess the magnitude of variability of rainfall within the SGs. The box plots showed not just the seasonality but that the variability and intensity of rainfall between SGs were different.

6.4.2 Cluster analysis using 25 km WRF-CMIP5 data along with 25 km averaged CHIRPS data

The Ward's clustering method was also used to study redistribution of rainfall over WMC for the 42 stations (grids here) using WRF-CMIP5 (25 km) output for the year 2030 to 2060, and CHIRPS (25 km) grids for the year 1980-2005. As this was a comparison study, only three SGs were considered for the study that had a height (between cluster distance) of ~300. It was observed that SGs formed for WRF-CMIP5 (RCP 4.5 and RCP 8.5) were very different. The difference between CHIRPS (historical) and WRF-CMIP5 (future) SGs either signifies the changes in rainfall over next 25 years or may also be the result of model uncertainty. While comparing the SGs for RCP 4.5 and RCP 8.5, it was observed that of the landmasses Borneo would see the most significant redistribution of rainfall.

6.4.3 Recommendations for Future Studies

The future studies can focus on applying clustering analysis on variables such as temperature, the wind, relative humidity to understand the variables affecting the distribution and redistribution of rainfall over WMC. Furthermore, the study can be applied to daily and hourly data to examine the changes distribution of daily climatology as well as diurnal precipitation cycle over stations. Lastly, the

redistribution of rainfall can be studied in more detail by using an ensemble of RCMs that examines model uncertainty as well at a resolution higher than 25 km.

APPENDIX

Plots for bias correction in IDF for Jakarta and Kuala Lumpur

Figure a.1 Percentage error in between WRF and CFSR IDFs (left) and CFSR and Observed station ODFs (right) for RP=100 years (top) and RP= 50 years (bottom) over Kuala Lumpur

Figure a.2 Percentage error in between WRF and CFSR IDF's (left) and CFSR and Observed station ODF's (right) for RP=100 years (top) and RP= 50 years (bottom) over Jakarta

Analysis of the clusters formed for CHIRPS 5 and 25 km grid

As this study used SGs of different resolution, i.e., 25 (bilinear averaged from 5k data) and 5 km CHIRPS, it is important to review the changes in the SGs when resolution of the data is changed (see Fig a.3 and Fig a.4). A dendrogram resulting from these two resolutions, i.e., CHIRPS-25 and CHIRPS-5 signify the following:

- 1) CHIRPS-25 data results in three SGs at a greater height (between cluster distance) when compared to CHIRPS-5. This indicated that the clusters of CHIRPS- 25 km grid were far from each other whereas the clusters from CHIRPS-5 km data were closer to each other.
- 2) It was observed that though the clusters formed had different numbering similar grid stations were grouped together. Hence, we concluded that overall the smaller number of cluster was not significantly different as they were being determined largely by locations.

(a)

(b)

Figure a.3 Dendrogram corresponding CHIRPS 25 km data (top) and 5 km data (bottom)

a)

Figure a.4 CHIRPS dataset SG map 25 km (top) and 5km (bottom)

References

- Ahmed, B.Y.M., 1997. Climatic classification of Saudi Arabia: An application of factor--Cluster analysis. *GeoJournal*, 41(1): 69-84.
- Ahn, J.-B., Hong, J.-Y. and Seo, M.-S., 2013. Present-day climate of the Korean Peninsula centered Northern East Asia based on CMIP5 historical scenario using fine-resolution WRF. *Atmosphere*, 23(4): 527-538.
- Altava-Ortiz, V. et al., 2006. Application of the MM5 and the analogous method to heavy rainfall event, the case of 16? 18 October 2003 in Catalonia (NE Spain). *Advances in Geosciences*, 7: 313-319.
- Baek, H.-J. et al., 2013. Climate change in the 21st century simulated by HadGEM2-AO under representative concentration pathways. *Asia-Pacific Journal of Atmospheric Sciences*, 49(5): 603-618.
- Barnett, T.P. et al., 2008. Human-induced changes in the hydrology of the western United States. *science*, 319(5866): 1080-1083.
- Blashfield, R.K., 1976. Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods. *Psychological Bulletin*, 83(3): 377.
- Boer, R. and Faqih, A., 2004. Current and future rainfall variability in Indonesia. An Integrated Assessment of Climate Change Impacts, Adaptation and Vulnerability in Watershed Areas and Communities in Southeast Asia. Report from AIACC Project No AS21. Int. START Secretariat. Washington, DC. p, 47.
- Bruyère, C.L., Done, J.M., Holland, G.J. and Fredrick, S., 2014. Bias corrections of global models for regional climate simulations of high-impact weather. *Climate dynamics*, 43(7-8): 1847-1856.
- Chang, C.P., 2005. Annual Cycle of Southeast Asia—Maritime Continent Rainfall and the Asymmetric Monsoon Transition. *Journal of Climate*, 18(2): 287-301.
- Change, I.P.o.C., 2014. *Climate Change 2014—Impacts, Adaptation and Vulnerability: Regional Aspects*. Cambridge University Press.
- Chen, L. and Frauenfeld, O.W., 2016. Impacts of urbanization on future climate in China. *Climate dynamics*, 47(1-2): 345-357.
- Christensen, O., Gaertner, M., Prego, J. and Polcher, J., 2001. Internal variability of regional climate models. *Climate Dynamics*, 17(11): 875-887.
- Christensen, O.B., 2001. The PRUDENCE project, Description of work. Technical Report. 61.
- Chua, L.H., 2002. *The Singapore Green Plan 2012*. Ministry of the Environment and Water Resource: Singapore.

- Collischonn, B., Collischonn, W. and Tucci, C.E.M., 2008. Daily hydrological modeling in the Amazon basin using TRMM rainfall estimates. *Journal of Hydrology*, 360(1): 207-216.
- Daniell, T.M. and Tabios III, G., 2008. Asian Pacific FRIEND Rainfall Intensity Duration Frequency (IDF) Analysis for the Asia Pacific Region. Regional Steering Committee for Southeast Asia and the Pacific UNESCO Office, Jakarta.
- Davis, N., Bowden, J., Semazzi, F., Xie, L. and Öñol, B., 2009. Customization of RegCM3 regional climate model for eastern Africa and a tropical Indian Ocean domain. *Journal of Climate*, 22(13): 3595-3616.
- DeGaetano, A.T., 2001. Spatial grouping of United States climate stations using a hybrid clustering approach. *International Journal of Climatology*, 21(7): 791-807.
- DeGaetano, A.T. and Shulman, M.D., 1990. A climatic classification of plant hardiness in the United States and Canada. *Agricultural and Forest Meteorology*, 51(3-4): 333-351.
- Development, N.C.P.C.s.g.m.d.d.i.s.o.t.U.A.f.I., NOAA Climate Prediction Center's global meteorological data dissemination in support of the US Agency for International Development.
- Donat, M.G., Lowry, A.L., Alexander, L.V., O'Gorman, P.A. and Maher, N., 2016. More extreme precipitation in the world's dry and wet regions. *Nature Climate Change*, 6(5): 508-513.
- Fan, Y. and Dool, H.V.D., 2004. Climate Prediction Center global monthly soil moisture data set at 0.5 resolution for 1948 to present. *Journal of Geophysical Research Atmospheres*, 109(D10): -.
- Febri, D.H., Hidayat, R. and Hanggoro, W., 2016. Sensitivity of WRF-EMS model to predict rainfall event on wet and dry seasons over West Sumatra. *Procedia Environmental Sciences*, 33: 140-154.
- Firth, P. and Fisher, S.G., 2012. *Global climate change and freshwater ecosystems*. Springer Science & Business Media.
- Fujita, M., Takahashi, H.G. and Hara, M., 2013. Diurnal cycle of precipitation over the eastern Indian Ocean off Sumatra Island during different phases of Indian Ocean Dipole. *Atmospheric Science Letters*, 14(3): 153-159.
- Gabler, R.E., Petersen, J.F., Trapasso, L. and Sack, D., 2008. *Physical geography*. Nelson Education.
- Gan, T.Y., Dlamini, E.M. and Biftu, G.F., 1997. Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling. *Journal of Hydrology*, 192(1-4): 81-103.
- Gan, T.Y., Gobena, A.K. and Wang, Q., 2007. Precipitation of southwestern Canada: Wavelet, scaling, multifractal analysis, and teleconnection to climate anomalies. *Journal of Geophysical Research Atmospheres*, 112(D10): -.

- Gilliland, E.K. and Rowe, C.M., 2007. A comparison of cumulus parameterization schemes in the WRF model, Proceedings of the 87th AMS Annual Meeting & 21th Conference on Hydrology.
- Giorgi, F., 1990. Simulation of regional climate using a limited area model nested in a general circulation model. *Journal of Climate*, 3(9): 941-963.
- Glover, D., 2006. Indonesia's fires and haze: the cost of catastrophe. IDRC.
- Gobena, A.K. and Gan, T.Y., 2006. Low-frequency variability in Southwestern Canadian stream flow: links with large-scale climate anomalies. *International Journal of Climatology*, 26(13): 1843–1869.
- Greaver, T.L. and Sternberg, L.S., 2010. Decreased precipitation exacerbates the effects of sea level on coastal dune ecosystems in open ocean islands. *Global Change Biology*, 16(6): 1860-1869.
- Hänsel, S., Mehler, S. and Matschullat, J., 2013. Evaluating dry and wet period changes using an ensemble of GCMs, ENSEMBLES RCMs and additional higher resolved RCMs. Egu General Assembly, 15.
- Hewitt, C.D., 2004. Ensembles-based predictions of climate changes and their impacts. *Eos, Transactions American Geophysical Union*, 85(52): 566-566.
- Hock, R. and Holmgren, B., 2005. A distributed surface energy-balance model for complex topography and its application to Storglaciären, Sweden. *Journal of Glaciology*, 51(172): 25-36.
- Holland, G.J. and Keenan, T.D., 1980. Diurnal variations of convection over the “Maritime Continent”. *Monthly Weather Review*, 108(2): 223-225.
- Holton, J., Dmowska, R. and Philander, S., 1990. El Nino, La Nina, and the Southern Oscillation.
- Hossain, F. and Katiyar, N., 2006. Improving flood forecasting in international river basins. *Eos, Transactions American Geophysical Union*, 87(5): 49-54.
- IPCC, 2013. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.
- Jones, R. et al., 2004. Generating high resolution climate change scenarios using PRECISReportMet Office Hadley Centre (UK).
- Kang, E.L., Cressie, N. and Sain, S.R., 2012. Combining outputs from the North American regional climate change assessment program by using a Bayesian hierarchical model. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(2): 291-313.
- Keenan, T.D., Manton, M.J., Holland, G.J. and Morton, B.R., 1989. The Island Thunderstorm Experiment (ITEX)—A study of tropical thunderstorms in the maritime continent. *Bulletin of the American Meteorological Society*, 70(2): 152-159.

- Kidson, J.W. and Thompson, C.S., 1998. A comparison of statistical and model-based downscaling techniques for estimating local climate variations. *Journal of Climate*, 11(4): 735-753.
- Knutti, R., Furrer, R., Tebaldi, C., Cermak, J. and Meehl, G.A., 2010. Challenges in combining projections from multiple climate models. *Journal of Climate*, 23(10): 2739-2758.
- Kraus, E.B., 1963. The diurnal precipitation change over the sea. *Journal of the Atmospheric Sciences*, 20(6): 551-556.
- Kuligowski, R.J. and Barros, A.P., 1998. Localized precipitation forecasts from a numerical weather prediction model using artificial neural networks. *Weather and forecasting*, 13(4): 1194-1204.
- Kuo, C.C., Gan, T.Y. and Gizaw, M., 2015. Potential impact of climate change on intensity duration frequency curves of central Alberta. *Climatic Change*, 130(2): 115-129.
- Lehmann, J., Coumou, D. and Frieler, K., 2015. Increased record-breaking precipitation events under global warming. *Climatic Change*, 132(4): 501-515.
- Lenderink, G. and Van Meijgaard, E., 2008. Increase in hourly precipitation extremes beyond expectations from temperature changes. *Nature Geoscience*, 1(8): 511.
- Leung, L.R. et al., 2004. Mid-century ensemble regional climate change scenarios for the western United States. *Climatic Change*, 62(1): 75-113.
- Liew, S.C., Raghavan, S.V. and Liong, S.Y., 2014. How to construct future IDF curves, under changing climate, for sites with scarce rainfall records? *Hydrological processes*, 28(8): 3276-3287.
- Madala, S., Satyanarayana, A. and Rao, T.N., 2014. Performance evaluation of PBL and cumulus parameterization schemes of WRF ARW model in simulating severe thunderstorm events over Gadanki MST radar facility—case study. *Atmospheric research*, 139: 1-17.
- Mailhot, A., Duchesne, S., Caya, D. and Talbot, G., 2007. Assessment of future change in intensity–duration–frequency (IDF) curves for Southern Quebec using the Canadian Regional Climate Model (CRCM). *Journal of hydrology*, 347(1): 197-210.
- McCarthy, J.J., Canziani, O.F., Leary, N.A., Dokken, D.J. and White, K.S., 2001. *Climate Change 2001: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, 2001. *J Appl Meteorol*, 44(5): 7009716Messerli.
- Measey, M., 2010. Indonesia: a vulnerable country in the face of climate change. *Global Majority E-Journal*, 1(1): 31-45.

- Meehl, G.A., 1987. The annual cycle and interannual variability in the tropical Pacific and Indian Ocean regions. *Monthly Weather Review*, 115(1): 27-50.
- Miao, C. et al., 2014. Assessment of CMIP5 climate models and projected temperature changes over Northern Eurasia. *Environmental Research Letters*, 9(5): 055007.
- Micheas, A.C., Fox, N.I., Lack, S.A. and Wikle, C.K., 2007. Cell identification and verification of QPF ensembles using shape analysis techniques. *Journal of Hydrology*, 343(3): 105-116.
- Molénat, J., Gascuel-Oudou, C., Aquilina, L. and Ruiz, L., 2013. Use of gaseous tracers (CFCs and SF 6) and transit-time distribution spectrum to validate a shallow groundwater transport model. *Journal of hydrology*, 480: 1-9.
- Monaghan, A.J., Steinhoff, D.F., Bruyere, C.L. and Yates, D., 2014. NCAR CESM Global Bias-Corrected CMIP5 Output to Support WRF/MPAS Research. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, CO.
- Moriondo, M. and Bindi, M., 2006. Comparison of temperatures simulated by GCMs, RCMs and statistical downscaling: potential application in studies of future crop development. *Climate Research*, 30(2): 149-160.
- Mulholland, P.J. et al., 1997. Effects of climate change on freshwater ecosystems of the south-eastern United States and the Gulf Coast of Mexico. *Hydrological Processes*, 11(8): 949-970.
- Murtagh, F. and Legendre, P., 2014. Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *Journal of Classification*, 31(3): 274-295.
- Neale, R. and Slingo, J., 2003. The Maritime Continent and Its Role in the Global Climate: A GCM Study. *Journal of Climate*, 16(5): 834-848.
- NH, S., BN, G., PN, V. and T, Y., 1999. A dipole mode in the tropical Indian Ocean. *Nature*, 401(6751): 360.
- Qian, J.-H., 2008. Why precipitation is mostly concentrated over islands in the Maritime Continent. *Journal of the Atmospheric Sciences*, 65(4): 1428-1441.
- Qu, T., Meyers, G., Godfrey, J.S. and Hu, D., 1994. Ocean dynamics in the region between Australia and Indonesia and its influence on the variation of sea surface temperature in a global general circulation model. *Journal of Geophysical Research: Oceans*, 99(C9): 18433-18445.
- Rahmat, Boonlert Archevarahuprok, Chai Ping Kang, Chung and Xiang, J., 2014. A Regional Climate Modelling Experiment for Southeast Asia.
- Ramage, C.S., 1968. Role of a tropical "maritime continent" in the atmospheric circulation. *Mon. Wea. Rev*, 96(6): 365-370.
- Randall, D.A. et al., 2007. Climate models and their evaluation, Climate change 2007: The physical science basis. Contribution of Working Group I to the

- Fourth Assessment Report of the IPCC (FAR). Cambridge University Press, pp. 589-662.
- Saha, S. et al., 2010. NCEP Climate Forecast System Reanalysis (CFSR) Selected Hourly Time-Series Products, January 1979 to December 2010. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, CO.
- Sardeshmukh, P.D. and Hoskins, B.J., 1988. The generation of global rotational flow by steady idealized tropical divergence. *Journal of the Atmospheric Sciences*, 45(7): 1228-1251.
- Schumacher, C., 2013. The role of low-level convective heating in tropical weather and climate (Invited), AGU Fall Meeting.
- Shalizi, C., 2009. Distances between Clustering, Hierarchical Clustering. Lectures notes.
- Shan, H., 2011. Rainfall Forecasting And Its Applications In Flood Early Warning Systems.
- Shiklomanov, A., Lammers, R. and Vörösmarty, C.J., 2002. Widespread decline in hydrological monitoring threatens pan-Arctic research. *Eos, Transactions American Geophysical Union*, 83(2): 13-17.
- Skamarock, W.C. et al., 2005. A description of the advanced research WRF version 2, National Center For Atmospheric Research Boulder Co Mesoscale and Microscale Meteorology Div.
- Stocker, T., 2014. Climate change 2013: the physical science basis: Working Group I contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press.
- Stokstad, E., 1999. Scarcity of rain, stream gages threatens forecasts. *American Association for the Advancement of Science*.
- Strachan, J., 2007. Understanding and modelling the climate of the maritime continent, University of Reading.
- Strandberg, G. et al., 2014. CORDEX scenarios for Europe from the Rossby. Centre regional climate model RCA4.
- Sugahara, S., Da Rocha, R.P. and Silveira, R., 2009. Non-stationary frequency analysis of extreme daily rainfall in Sao Paulo, Brazil. *International Journal of Climatology*, 29(9): 1339-1349.
- Suhaila, J., Jemain, A.A., Hamdan, M.F. and Zin, W.Z.W., 2011. Comparing rainfall patterns between regions in Peninsular Malaysia via a functional data analysis technique. *Journal of hydrology*, 411(3): 197-206.
- Sui, C. and Lau, K., 1992. Multiscale phenomena in the tropical atmosphere over the western Pacific. *Monthly weather review*, 120(3): 407-430.
- Tan, B.S.J., 2014. The role of large-scale organisation of convection for tropical weather and climate. *Chest*, 116(4): 935-40.
- Te Chow, V., 1988. Applied hydrology. Tata McGraw-Hill Education.

- Tewari, M. et al., 2004. Implementation and verification of the unified NOAA land surface model in the WRF model, 20th conference on weather analysis and forecasting/16th conference on numerical weather prediction.
- Thompson, G., Rasmussen, R.M. and Manning, K., 2004. Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part I: Description and sensitivity analysis. *Monthly Weather Review*, 132(2): 519-542.
- Tiedtke, M., 1989. A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Monthly Weather Review*, 117(8): 1779-1800.
- Trenberth, K.E., 2011. Changes in precipitation with climate change. *Climate Research*, 47(1/2): 123-138.
- Tucci, C.E., Collischonn, W., Clarke, R.T., Paz, A.R. and Allasia, D., 2008. Short-and long-term flow forecasting in the Rio Grande watershed (Brazil). *Atmospheric Science Letters*, 9(2): 53-56.
- Varikoden, H., Samah, A. and Babu, C., 2010. Spatial and temporal characteristics of rain intensity in the peninsular Malaysia using TRMM rain rate. *Journal of hydrology*, 387(3): 312-319.
- Vörösmarty, C.J., Green, P., Salisbury, J. and Lammers, R.B., 2000. Global water resources: vulnerability from climate change and population growth. *science*, 289(5477): 284-288.
- Wang, X. et al., 2014. A two-fold increase of carbon cycle sensitivity to tropical temperature variations. *Nature*, 506(7487): 212-215.
- Wang, Y., Zhou, L. and Hamilton, K., 2007. Effect of convective entrainment/detrainment on the simulation of the tropical precipitation diurnal cycle. *Monthly Weather Review*, 135(2): 567-585.
- Ward Jr, J.H., 1963. Hierarchical grouping to optimize an objective function. *Journal of the American statistical association*, 58(301): 236-244.
- Wetterhall, F., Halldin, S. and Xu, C.-y., 2005. Statistical precipitation downscaling in central Sweden with the analogue method. *Journal of Hydrology*, 306(1): 174-190.
- Wilby, R. et al., 2004. Guidelines for use of climate scenarios developed from statistical downscaling methods. Supporting material of the Intergovernmental Panel on Climate Change, available from the DDC of IPCC TGCIA, 27.
- Wilby, R.L. and Dawson, C.W., 2013. The statistical downscaling model: insights from one decade of application. *International Journal of Climatology*, 33(7): 1707-1719.
- Williams, M. and Houze Jr, R.A., 1987. Satellite-observed characteristics of winter monsoon cloud clusters. *Monthly Weather Review*, 115(2): 505-519.

- Wu, C.-H. and Hsu, H.-H., 2009. Topographic influence on the MJO in the Maritime Continent. *Journal of Climate*, 22(20): 5433-5448.
- Xu, C.-Y. and Singh, V., 2005. Evaluation of three complementary relationship evapotranspiration models by water balance approach to estimate actual regional evapotranspiration in different climatic regions. *Journal of Hydrology*, 308(1): 105-121.
- Yang, G.-Y. and Slingo, J., 2001. The diurnal cycle in the tropics. *Monthly Weather Review*, 129(4): 784-801.
- Yip, S., Ferro, C.A.T., Stephenson, D.B. and Hawkins, E., 2011. A Simple, Coherent Framework for Partitioning Uncertainty in Climate Predictions. *Journal of Climate*, 24(24): 4634-4643.
- Yokoi, S. et al., 2017. Diurnal Cycle of Precipitation Observed in the Western Coastal Area of Sumatra Island: Offshore Preconditioning by Gravity Waves. *Monthly Weather Review*, 145(9): 3745-3761.
- Yu, M., Wang, G., Parr, D. and Ahmed, K.F., 2014. Future changes of the terrestrial ecosystem based on a dynamic vegetation model driven with RCP8.5 climate projections from 19 GCMs. *Climatic change*, 127(2): 257-271.
- Yusuf, A.A. and Francisco, H., 2009. Climate change vulnerability mapping for Southeast Asia.
- Zhang, C., 2005. Madden-Julian Oscillation. *Reviews of Geophysics*, 43(2): -.
- Zhang, C., Wang, Y. and Hamilton, K., 2011. Improved representation of boundary layer clouds over the southeast Pacific in ARW-WRF using a modified Tiedtke cumulus parameterization scheme. *Monthly Weather Review*, 139(11): 3489-3513.
- Zhu, B. and Wang, B., 1993. The 30–60-day convection seesaw between the tropical Indian and western Pacific Oceans. *Journal of the atmospheric sciences*, 50(2): 184-199.