Simple and yet novel approach in flood assessment to overcome data scarcity : high quality DEM and rainfall proxies

Dong Eon Kim

To cite this version:

Dong Eon Kim. Simple and yet novel approach in flood assessment to overcome data scarcity : high quality DEM and rainfall proxies. Risques. COMUE Université Côte d'Azur (2015 - 2019), 2019. English. NNT : 2019AZUR4029 . tel-02492281

HAL Id: tel-02492281
https://theses.hal.science/tel-02492281

Submitted on 26 Feb 2020

UNIVERSITÉ CÔTE D'AZUR

ÉCOLE DOCTORALE
SCIENCES ET TECHNOLOGIES DE L'INFORMATION ET DE LA COMMUNICATION

THÈSE DE DOCTORAT

Approche simple et novatrice pour l'évaluation des inondations dans un contexte pauvre en données:

Solutions alternatives aux MNT haute résolution et aux données locales de précicipitation

Dong Eon KIM
Polytech Lab

Présentée en vue de l'obtention
du grade de docteur en discipline
d'Université Côte d'Azur
Dirigée par : Philippe GOURBESVILLE
Co-encadrée par : Shie-Yui LIONG
Soutenue le : 10 Juin 2019

Devant le jury, composé de :
Reinhard HINKELMANN, Pr., TU Berlin, Allemagne
Manuel GOMEZ VALENTIN, Pr., Université Politecnica de Catalunya, Espagne
Shie-Yui LIONG, Pr., Université Nationale de Singapour, Singapour
Philippe AUDRA, Pr., Université Nice Sophia Antipolis
Philippe GOURBESVILLE, Pr., Université Nice Sophia Antipolis
Ludovic ANDRES, Métropole Nice Côte d'Azur

ACKNOWLEDGEMENT

Pursuing a PhD program is both enjoyable and painful experience in my life journey. At the beginning of my study, I was very confident about what to do and what will have to be done. However, it was not as straightforward as I initially thought. I met obstacles even on the first step and had difficulty to solve the problems. When the issues were settled after various twists and turns, I regained the confidence that I could settle anything in the future. However, these up and down cycles repeated until the end of this PhD study. The experience gained from this PhD program surely prepare me to face challenges in both my future carrier and personal life.

Whenever I was steeped in anxiety, there were always people around me who comforted and strengthened me. I would like to express my deepest appreciation to all these people. First and foremost, I would like to express my sincere thanks to my supervisors, Professor Philippe Gourbesville and Professor Shie-Yui Liong, for their contributions in providing suggestions, feedbacks and many other countless supports. Without their support, I would never have completed my PhD program.

Special thanks are also extended to my colleagues since I started my PhD: Dr. Jiandong Liu, Dr. Srivatsan Vijayaraghavan, Dr. Yabin Sun, Dr. Dadiyorto Wendi, Dr. Chi Dung Doan, Dr. Qiang Ma and Dr. Lian Guey Ler. They, aside from being my good friends, helped me in providing the necessary data and ideas leading to the present quality of my study. I had excellent time with them; I surely treasure our friendship.

I also wish to extend my appreciation to all those organizations and persons who made their invaluable data available to me: (1) Dr. Ludovic Andres of Nice Côte d'Azur Metropolis for providing the extremely high resolution DEM of Nice; (2) German Aerospace Center (DLR) for TanDEM-X; (3) National Aeronautics and Space Administration (NASA) for SRTM; and (4) European Space Agency (ESA) for Sentinel-2

Last, but not least, I am greatly indebted to my wife, Soo Yeon, and my kids, Do Yeon and Min Geon. They are absolutely the source of my happiness. Their love and support have encouraged me to fully focus on and complete this PhD program with excellent results. Thank you.

ABSTRACT

Simple-and-yet-novel approach in flood assessment to overcome data scarcity: High quality DEM and rainfall proxies

Many urban cities in Southeast Asia witness severe flooding associated to increasing rainfall intensity and rapid urbanization often due to poor urban planning. Two important inputs required in flood hazard assessment are: (1) high accuracy Digital Elevation Model (DEM), and (2) long rainfall record. High accuracy DEM is both expensive and time consuming to acquire. Long rainfall records for areas of interest are often not available or not sufficiently long to determine the probable extremes. This thesis presents a notably cost-effective and efficient approach to estimate high-resolution and accuracy DEM, and suggests proxies for long rainfall data.

DEM data from a publicly accessible satellite, Shuttle Radar Topography Mission (SRTM), and Sentinel 2 multispectral imagery are selected and used to train the Artificial Neural Network (ANN) to improve the quality of the DEM. In the training of ANN, high quality observed DEM is the key leading to a well-trained ANN. The trained ANN will then be ready to efficiently and effectively generate high quality DEM, at low cost, for places where DEM data is not available.

The performance of the DEM improvement scheme is evaluated in places of various landuse types (e.g. dense urban areas, forested areas), and in different countries (Nice, France; Singapore; Jakarta, Indonesia) through various criteria, e.g. whenever possible visual clarity, scatter plots, Root Mean Square Error (RMSE) and drainage networks. The DEM resulting from the latest version of improved SRTM (iSRTM_v2 DEM) performs (1) significantly better than the original SRTM DEM, a 34 % to 57 % RMSE reduction; (2) the visual clarity is so much better; and (3) much closer drainage network with the actual. The much improved DEM allows flood modelling to proceed with high confidence.

Rainfall data resulting from a high spatial resolution Regional Climate Model (RCM), Weather Research and Forecasting driven by ERA-Interim (WRF/ERAI) dataset, is extracted, analyzed, and compared with high quality observed rainfall data of Singapore with regard to accuracy. The comparisons are performed, among others, on their Intensity-Duration-Frequency (IDF) curves, the essential design curves for flood risk assessment; they matched quite well. The rainfall data (from the RCM) are then used as proxies for Greater Jakarta (Indonesia), where no rainfall data were available, to derive the IDF curves required for the flood analysis.

MIKE 21 Flow Model Flexible Mesh (MIKE 21 FM) is applied to Greater Jakarta, with input data from the above mentioned much improved DEM and precipitation proxy data, for flood simulations of 2 return periods (50- and 100-years). Qualitative agreement of model results and observation of the 2013 Jakarta flood were obtained. This demonstrates the applications of the approaches/methodologies, proposed in this thesis, on catchments where most essential data for flood risk assessment (high resolution and high accuracy DEM and long and high accuracy rainfall data) are not available.

This thesis should be of interest to readers of the areas of remote sensing, artificial intelligence and flood management, possibly also for the policy makers in proposing relevant flood mitigation measures under climate change with increasing devastating flood damages and casualties.

RESUME

De nombreuses villes d'Asie du Sud-Est subissent de graves inondations liées d'une part à l'intensité croissante des précipitations et part d'autre à une urbanisation rapide souvent due à une planification urbaine non maitrisée. L'évaluation quantitative des risques d'inondation nécessite deux éléments essentiels: (1) un modèle numérique de terrain (MNT) haute définition, et (2) une chronologie de précipitations la plus longue possible. Un MNT haute définition est à la fois coûteux et long à acquérir. Les chronologies de précipitations longues sont fréquemment indisponibles dans de nombreux sites et ne présentent pas toujours une durée suffisante pour une définition pertinente des valeurs extrêmes. Cette thèse présente une approche opérationnelle pour générer des MNT haute définition et suggère une stratégie pour définir des pluies extrêmes en dehors de chronologies de précipitations longues.

Des données pour la production des MNT issues de capteurs satellitaires - mission SRTM (Shuttle Radar Topography Mission) et images multi spectrales Sentinel 2 - ont été utilisées et mises en œuvre. Un réseau de neurones artificiels (ANN) est utilisé afin d'améliorer la qualité du MNT. Dans la phase d'apprentissage du réseau de neurones, la qualité des MNE utilisés comme référence est essentielle et conditionne la performance de l'outil dans son application ultérieure. A la suite de cet apprentissage, le réseau de neurones peut être mis en œuvre pour générer, à faible coût, des MNT haute résolution dans des secteurs où les données sont partiellement indisponibles.

Les performances de la méthode d'amélioration du MNT sont évaluées dans des différents secteurs caractérisés par des occupations du sol variées (secteur urbain dense, secteurs boisés par exemple) et dans différents pays (Nice, France; Singapour; Jakarta, Indonésie). La qualité des résultats est analysées avec différents indicateurs tels que les diagrammes de dispersion, la clarté visuelle, l'erreur quadratique moyenne (RMSE) et l'adéquation avec les réseaux de drainage réels. Le MNT issu des données SRTM améliorées (iSRTM_v2 DEM) montre (1) une qualité nettement supérieure au MNT initial puisque le RMSE passe de 34 % à 57 % du RMSE; (2) la clarté visuelle est largement améliorée; et (3) le réseau de drainage calculé correspond davantage au réseau réel. La production de ce MNT amélioré permet une meilleure modélisation des processus d'inondation et augmente la qualité des résultats des simulations hydrauliques.

Des données de précipitation issues d'un Modèle Climatologique Régional (RCM) haute résolution spatiale ainsi que des prévisions issues de données ERA-Interim (WRF / ERAI) ont été extraites, analysées et comparées avec les observations haute résolution enregistrées à Singapour. Les comparaisons ont également été effectuées avec les courbes Intensité-Durée-Fréquence (IDF) qui sont utilisées pour l'évaluation des risques d'inondation. Les résultats sont très satisfaisants et valident les données produites par le modèle régional. Cette validation permet d’utiliser les données pluviométriques issues du modèle régional pour le site de la métropole de Jakarta (Indonésie) où les enregistrements pluviométriques ne sont pas disponibles pour la production des courbes IDF.

Un modèle hydraulique détaillé a été construit avec le système de modélisation MIKE 21 (MIKE 21 FM) pour toute la métropole de Jakarta à partir d’un MNT amélioré et des précipitations associées à des périodes de retour de 50 et 100 ans. Des cartes d’inondation ont été générées et sont utilisées par les services gestionnaires. Cet exemple démontre que les nouvelles méthodes et approches proposées dans cette thèse sont pertinentes pour produire une évaluation des risques d’inondation pertinente lorsque des données locales (MNT haute résolution et données pluviométriques sur une période longue) sont insuffisantes ou indisponibles.

Les résultats de ce travail de recherche devraient intéresser les lecteurs oeuvrant dans les domaines de la télédétection, de l'intelligence artificielle et de la gestion des inondations. La méthodologie proposée est destinée à permettre aux décideurs et gestionnaires de proposer des mesures appropriées afin de réduire les conséquences des inondations sur les biens et les personnes dans le contexte du changement climatique.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

LIST OF ABBREVIEATION

ACCESS1.3	Australian Community Climate and Earth-System Simulator
AI	Artificial Intelligence
AMR	Annual Maximum Rainfall
ANN	Artificial Neural Network
APHRODITE	Asian Precipitation - Highly-Resolved Observational Data Integration Towards Evaluation
ASEAN	Association of South East Asian Nations
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
BCA	Building and Construction Authority
BRDF	Bidirectional Reflectance Distribution Function
CDF	Cumulative Distributed Function
CFL	Courant-Friedrich-Lewy
CHIRPS	Climate Hazards Group InfraRed Precipitation with Station
COP35	35 rainfall stations, in Code of Practice
CRU	Climate Research Unit
DEM	Digital Elevation Model
DLR	German Aerospace Center (German: Deutsches Zentrum für Luft- und Raumfahrt e.V.)
DSM	Digital Surface Model
DTM	Digital Terrain Model
ECHAM6	European Centre HAMburg Model
ECMWF	European Centre for Medium-Range Weather Forecasts
EROS	Earth Resources Observation and Science
ESA	European Space Agency
ESRI	Environmental System Research Institute
GCM	Global Climate Model
GDEM	Global DEM
GEV	Generalized Extreme Value
GIS	Geographic Information System
GMES	Global Monitoring for Environment and Security
GPS	Global Positioning System
GPU	Graphical Processing Units
GTM	Global Tide Model
IDF	Intensity-Duration-Frequency
InSAR	Interferometric Synthetic Aperture Radar
iSRTM DEM	Improved SRTM DEM

L-CV	L-Coefficient of Variation
LiDAR	Laser Imaging Detection and Ranging
LM	Levenberg-Marquardt
METI	Ministry of Economy, Trade, and Industry
MIKE21FM	MIKE21 Flow Model Flexible Mesh
MIROC5	Model for Interdisciplinary Research on Climate
MSI	Multispectral Instrument
MSL	Mean Sea Level
NASA	National Aeronautics and Space Administration
NBAR	Nadir BRDF Adjusted Reflectance
NDWI	Normalised Difference Water Index
NIR	Near Infrared
NOAA	National Oceanic and Atmospheric Administration
NSI	National Space Institute
OSM	Open Street Map
P95p	95th Percentile of Precipitation
PDF	Probability Density Function
PUB	Public Utility Board
RCM	Regional Climate Model
RCP	Representative Concentration Pathway
RFA	Regional Frequency Analysis
RMSE	Root Mean Square Error
SDII	Simple Day Intensity Index
SRTM	Shuttle Radar Topography Mission
SWIR	Short Wave Infrared
SWMM	Storm Water Management Model
TC	Time of Concentration
THB	Thai baht
TOA	Top-Of-Atmosphere
USGS	United States Geological Survey
VNIR	Visible and Near Infrared
WRF	Weather Research and Forecast
WRF/ERAI	WRF driven by ERA-Interim

1 Introduction

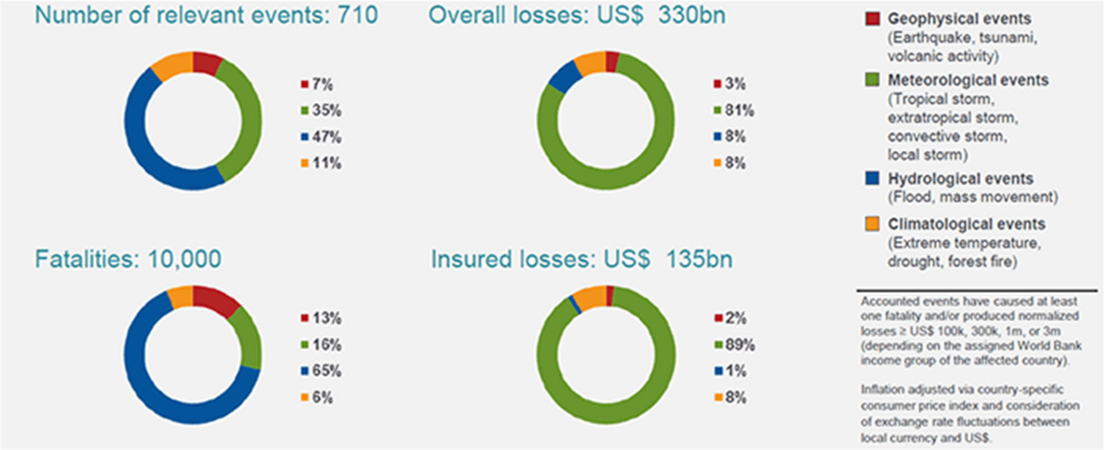
1.1 Background

Flooding is known as one of the most devastating and costliest global phenomena that can seriously disrupt transport and communication, severely damage properties, and cause devastated economic and human losses. Flooding is likely to be even further worsened in the future with changing climate. With clearly increasing global temperature, extreme weather events have been recorded; in Southeast Asia alone in recent years, these extreme events range from typhoon Haiyan (the highest category 5) in 2013, heavy rainfalls (and hence flooding in Jakarta 2007, Hanoi 2008, Bangkok 2011) to resulting landslides. Rainfall events with both higher intensity and frequency have been observed and are likely to continue. A warmer atmosphere is expected to accelerate the hydrological cycle due to the higher humidity content.

In late June through mid-July 2018, heavy downpours in Japan, a country highly ranked in disaster preparedness, resulted in severe floods and landslides. 225 people were confirmed dead in the affected areas, and economic losses reached an estimated US$ 3.66 billion [*Sim*, 2018]. Despite the efforts of many international organizations and countries, these catastrophes continue to occur at alarming rate. In Southeast Asia this is due to the rapid urbanization (hence, increasing migration into the urban areas for economic reasons), poor urban planning and enforcement of storm drainage network designs and maintenance works.

Insurance Information Institute (https://www.iii.org/fact-statistics/facts-statistics-global-catastrophes) showed insightful statistics on various world natural catastrophes by type of events in 2017, and by continent as shown in Figure 1.1 (a) and (b) respectively. Figure 1.1 (a) showed the number of events, fatalities and insured losses mainly caused by meteorological and hydrological events (about 80 % combined) while Figure 1.1 (b) clearly showed how vulnerable Asian continent is in the number of events (44 %), fatalities (65 %) and insured losses (2 %). The message is very alarming for Asia as the vulnerability is high and yet the insured losses are extremely low compared to their North American counterparts. Insurance Information Institute gave further breakdown on world weather-related catastrophes by type of events in 2017, and by continent as shown in Figure 1.2 (a) and (b) respectively. Figure 1.2 (a) showed the number of events (51 %) and fatalities (75 %) caused by hydrological events (flood, mass movement) dominated; the corresponding insured losses, however, was just 1 %. The statistics for Asia was again most alarmingly vulnerable: the number of events was 41 %, fatalities 66 %, and yet the insured losses only 2 %; note that most likely the main portion of the insured losses is in more developed countries such as Japan, Korea, and Singapore. Southeast Asian countries are indeed very vulnerable.

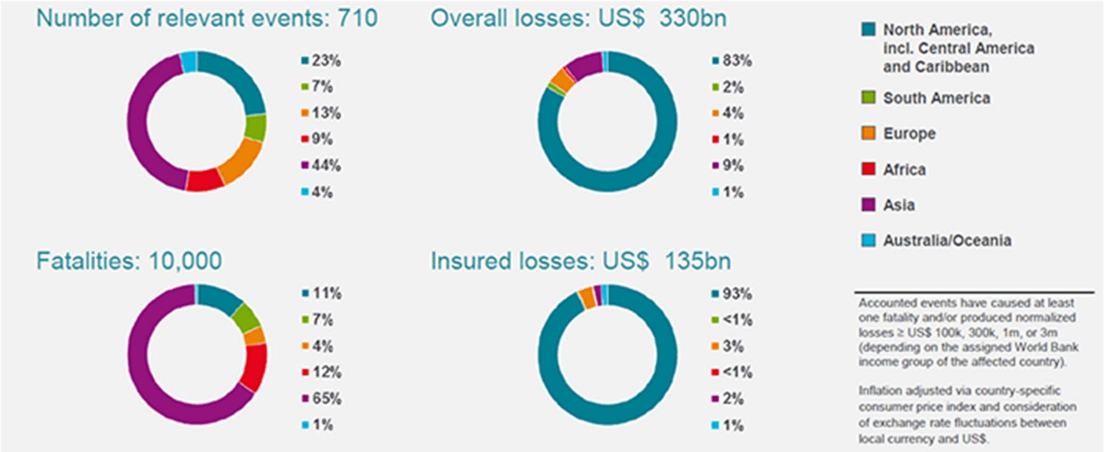
(a) World natural catastrophes by type of event, 2017



Accounted events have caused at least one fatality and/or produced normalized losses ≥ US$ 100k, 300k, 1m, or 3m (depending on the assigned World Bank income group of the affected country).

Inflation adjusted via country-specific consumer price index and consideration of exchange rate fluctuations between local currency and US$.

(b) World natural catastrophes by continent, 2017

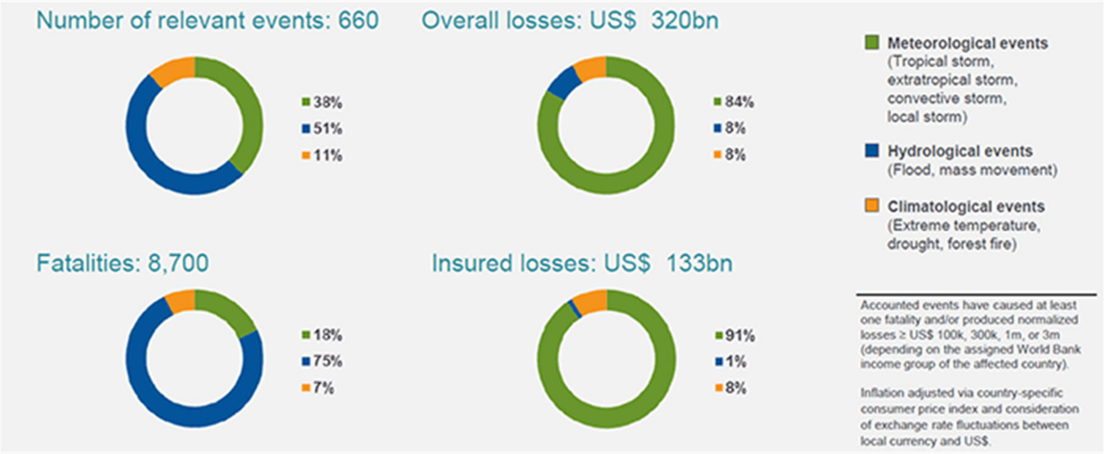


Accounted events have caused at least one fatality and/or produced normalized losses ≥ US$ 100k, 300k, 1m, or 3m (depending on the assigned World Bank income group of the affected country).

Inflation adjusted via country-specific consumer price index and consideration of exchange rate fluctuations between local currency and US$.

Figure 1.1 World natural catastrophes by (a) type of event and (b) continent in 2017 [*III*, 2018]

(a) World weather-related catastrophes by type of event, 2017



(b) World weather-related catastrophes by continent, 2017

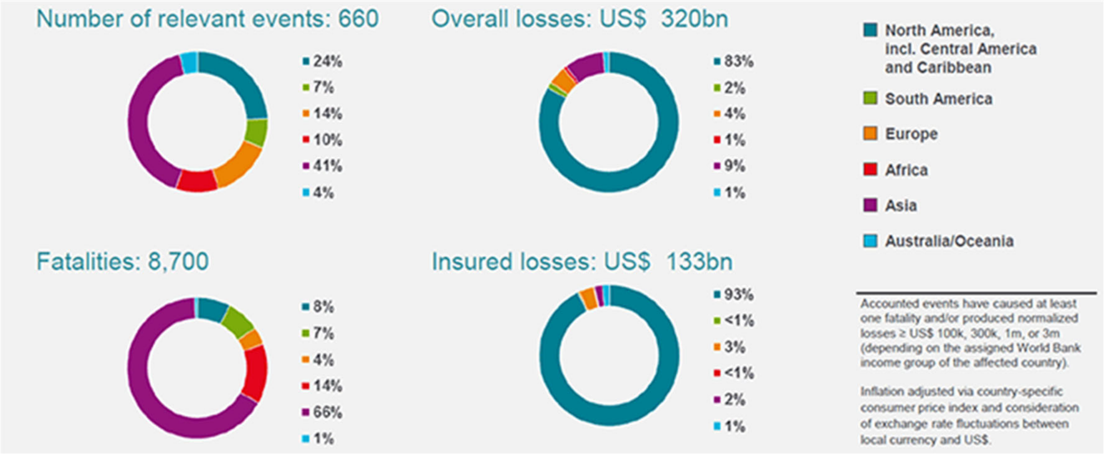


Figure 1.2 World weather-related catastrophes by (a) type of event and (b) continent in 2017 [*III*, 2018]

Southeast Asia comprises of 11 nations - Indonesia, Malaysia, Singapore, Vietnam, Brunei, Laos, Myanmar, Philippine, Thailand, Cambodia and Timor Leste. This region covers an area of about 4.5 million km^2 which is 3 % of the earth's land area. Its population is more than 641 million which is about the 8.5 % of world's population. The economic growth rate of Southeast Asia is the fastest growth in the world since 1990. With abundant resources and high population growth, the overseas investments concentrate in rapidly developing coastal megacities. However, Southeast Asia is geographically located in one of the most disaster-prone regions of the world; many countries in the region have a history of devastating disasters that have caused colossal economic and human losses. According to ASEAN (Association of South East Asian Nations) Disaster Risk Management Initiative, flood is the most reported disaster in ASEAN countries between 1970 and

2009, as listed Table 1-1. Out of all reported disasters, 36 percent were floods, 32 percent were cyclonic storms [*UNISDR*, 2010].

Indonesia, one of the most vulnerable countries to natural disasters, experienced in year 2018 almost 2,000 natural disasters (tsunami, earthquake and flooding) which claimed nearly 4,000 lives and displaced around 3 million people. As recent as in December 2018, Sunda Strait tsunami, triggered by a volcanic activity on the island of Anak Krakatoa, caused 429 deaths and 1,485 injured [*Renaldi and Shelton*, 2018].

The damage of natural disasters is expected to become more severe in the future due to rapid developments and population growth over the study region compounded with climate change. Challenges faced by ASEAN are huge. ASEAN needs to incorporate future potential risks into disaster risk reduction.

Table 1-1 Statistic of reported disaster between 1970 and 2009 [*UNISDR*, 2010]

Disaster type	No. of disasters /year	Total no. of deaths	Deaths/ year	Relative vulnerability (deaths/year/million)	Average annual economic loss ($ million)
Flood	10.85	17,800	445.0	0.75	312.1
Storm	9.65	184,063	4,601.6	7.76	339.4
Epidemic	2.28	7,294	182.4	0.31	-
Landslide	2.05	5,058	126.5	0.21	4.4
Forest Fire	0.45	310	7.8	0.01	511.9
Drought	0.98	1,337	33.4	0.06	45.8
Tsunami	0.15	92,021	2,300.5	3.88	214.2
Volcano	1.33	1,380	34.5	0.06	32.1
Earthquake	2.58	105,735	2,643.4	4.46	243.9

Focusing further on Southeast Asia, a study of *Yusuf and Francisco* [2009], relevant to present conditions, highlighted the climate hazards over domain using a multi climate index; the study conducted a vulnerability mapping based on tropical cyclones, floods, landslides, droughts, and sea level rise. Figure 1.3 and Figure 1.4 focus on the overall climate hazard and the annual flood frequency maps for Southeast Asia. Figure 1.3 showed, for example, how vulnerable the Philippines is; its vulnerability is mainly caused by destructive typhoons which hit the Philippines about 5 times annually. Figure 1.4 showed, for example, the vulnerability of the city of Jakarta (on Java island of Indonesia) to floods.

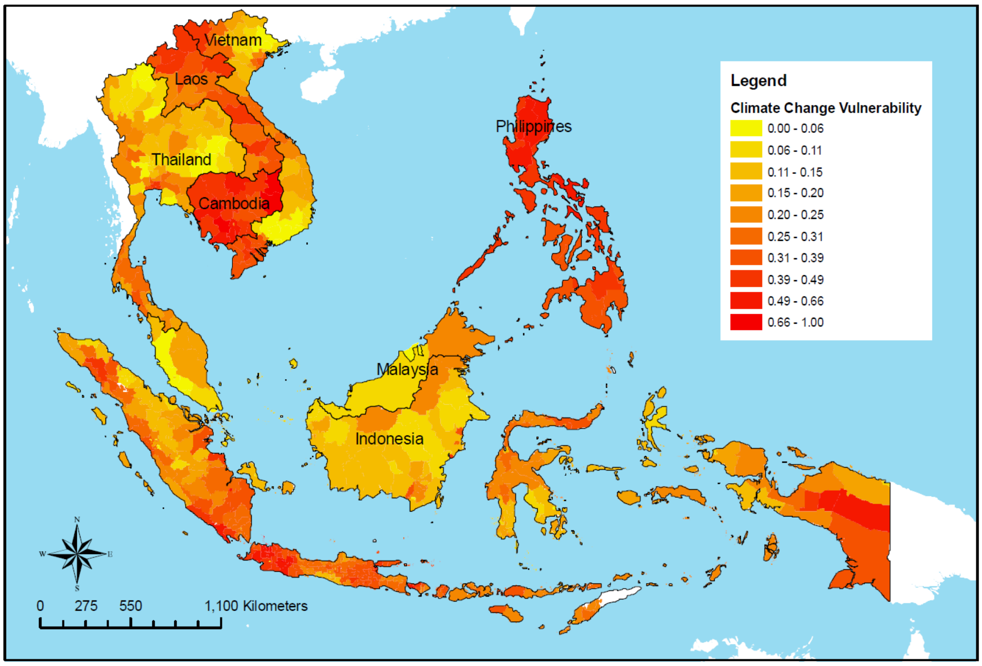


Figure 1.3 Multiple climate hazard map of Southeast Asia [*Yusuf and Francisco*, 2009]

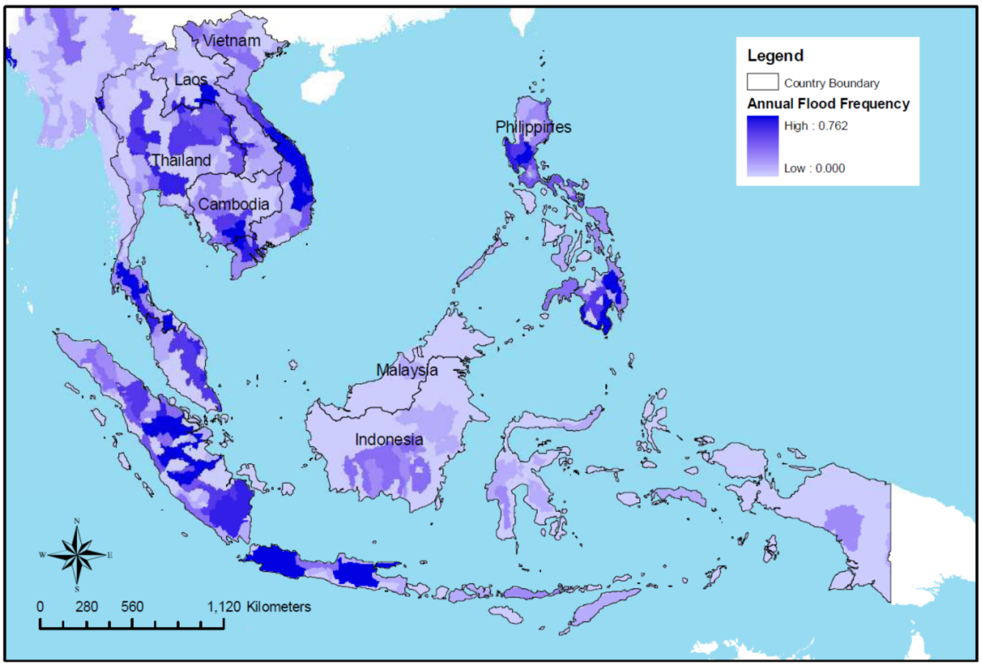


Figure 1.4 Annual flood frequency map of Southeast Asia [*Yusuf and Francisco*, 2009]

To briefly highlight the severity of floods in Southeast Asia in recent years, two countries, Thailand and Indonesia, are singled out here. In July – Dec 2011, Central and Northern Thailand received 300 to 500 mm (that is about 3 times) more rainfall than the normal condition as the Asian monsoon started early with an extraordinary heavy rainfall [*Bevere et al.*, 2012], as shown in Figure 1.5. Duration of flooding was 175 consecutive days. It resulted in 815 deaths and affected more than 13.6 million people. With many industrial areas severely affected, [*Bhoochaoom and Dixon*, 2012] estimated the total damage and economic losses of about Thai Baht (THB) 1.43 trillion (US$ 46.5 billion).

Figure 1.5 Thailand Flood in July to December 2011 [*Bangkokpost*, 2012]

In January – February 2007 torrential rains pounded Jakarta, the capital city of Indonesia. This caused floods that buried 36 % of the city under as much as 5 m in some areas, affected 2.6 million people, and forced 340,000 people to flee their home, as shown in Figure 1.6. Over 70 people died and outbreaks of disease affected over 200,000 people, with losses estimated at US$ 900 million. Jakarta's notorious subsidence, estimated at 7.5 cm/year, worsens the flooding during heavy rainfall combined with high tide. This will be even of greater concern with climate change which projected increasing rainfall intensity and rising sea level.

Figure 1.6 Jakarta Flood in January 2007 [*Floodlist*, 2008]

1.2 Gaps in Flood Analysis

Developed countries apply, for example, sophisticated rainfall nowcasting and forecasting, well calibrated and validated flood models, and flood early warning systems to disseminate anticipated water levels in the river/stream networks and low lying areas, and thus, to minimize damages/losses caused by floods. However, developing countries may not have enough information to solve the aforementioned issues due to its economic and technical constraints. Effective preventive and mitigation measures can only be undertaken when good quality data (e.g. Digital Elevation Model (DEM) and rainfall records) are available for flood model.

There are several well established numerical flood models, such as SWMM [*Rossman*, 2010], SOBEK [*Deltares*, 2019] and MIKE Flood [*DHI*, 2017]. In principle, the governing equations of these models are the same, St. Venant and Navier-Stokes equations, coupled to various sub-components (infiltration, evapotranspiration, etc.), and solved by different numerical schemes. Challenges in developing countries are, among others, limited project funding to acquire the aforementioned good quality, high temporal resolution and long rainfall record, and high spatial resolution and high accuracy DEM; their availabilities are often even doubtful.

In recent years, there is a neologism, called ‘the flood of Big Data’, which means huge data publicly accessible so people can develop immense opportunities in various ways. Remote sensing, as an example, is the process of detecting and monitoring the physical characteristics of an area by measuring its reflected and emitted radiation at a distance from the targeted area. This technology has been used, for example, in taking images on the earth’s surface, tracking clouds to predict the weather, tracking the growth of an area and changes in landuses etc.

This study focuses on two important input data required in flood modelling and analysis. They are high accuracy DEM and a long rainfall record. Remote sensing data and an artificial intelligence technique, Artificial Neural Network (ANN) are proposed to significantly improve the original remote sensing DEM data, for areas where high spatial resolution and high accuracy DEM is not available. For areas where observed rainfall data are either not available or not sufficiently long, the study proposes rainfall proxy products from various gridded observation data such as Climate Hazards Group InfraRed Precipitation with Station (CHIRPS), Climate Research Unit (CRU), Asian Precipitation - Highly-Resolved Observational Data Integration Towards Evaluation (APHRODITE) or rainfall outputs from a Regional Climate Model (RCM). The question remains how to assess flood risk in the region of interest where the aforementioned data are not available or sufficient. This study will offer (1) a methodology to derive improved DEM from publicly accessible remote sensing data, and (2) an approach to select a highly accurate rainfall proxy, from a RCM, required to construct the much needed Intensity-Duration-Frequency (IDF) curves for flood model and analysis.

1.3 Study Region – Greater Jakarta, Indonesia

As shown in Figure 1.4 and Figure 1.6, Java Island is vulnerable to flooding in Southeast Asia. Java Island is one of the large islands and most densely populated. More than half of Indonesia's 226 million populations live in Java (141 million in 2014 Census). The island also hosts major industrial complexes.

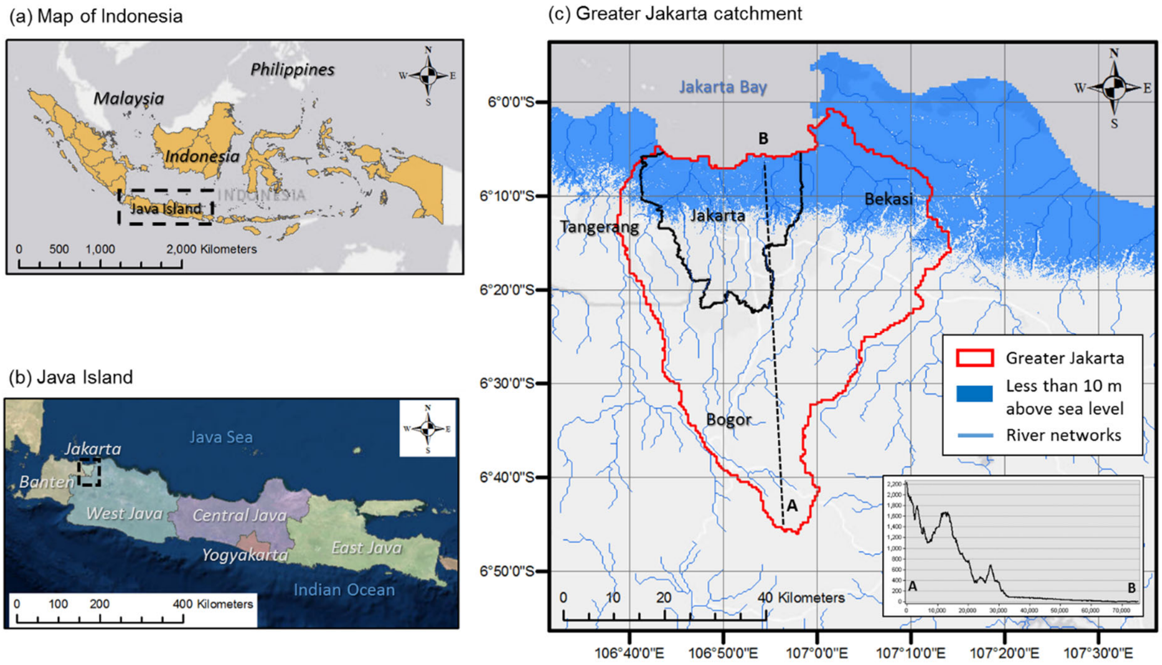
Jakarta, a capital city of Indonesia, is located on the northwest coast of Java Island, at the mouth of the Ciliwung River on Jakarta Bay. Jakarta has a land area of about 660 km^2 with population over 10 million [JakartaOpenData, 2015]. There are 13 main rivers flowing through the city and its vicinity. The longest river is the Ciliwung River. The climate of Jakarta is tropical wet and dry; the rainy season in Jakarta starts in December and ends in March. The rainfall intensity often reaches its peak in January or February (Table 1-2). Jakarta suffers from massive flooding almost yearly mainly due to high rainfall intensities, low lying areas and poorly managed drainages. In 2007, approximately 70 % of Jakarta's area was flooded with water depths up to 4 meters. Area of the catchment contributing water to Jakarta Bay is about 2,976 km^2 (hereafter referred as Greater Jakarta).

Table 1-2 Climate information of Jakarta [*Sun et al.*, 2014]

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Year
Average High (C°)	31.5	32.3	32.5	33.5	33.5	34.3	33.3	33.0	32.0	31.7	31.3	32.0	32.6
Average Low (C°)	24.2	24.3	25.2	25.1	25.4	24.9	25.1	24.9	25.5	25.5	24.9	24.9	24.8
Humidity (%)	85	85	83	82	82	81	78	76	76	77	81	82	81
Rainfall (mm)	389.7	309.8	100.3	257.8	139.4	83.1	30.8	34.2	30.0	33.1	175.0	123.0	1706.2

Although there are many steep mountains in the upstream of Greater Jakarta, most of areas near to the coastal areas are quite flat and less than 10 m above Mean Sea Level (MSL), Figure 1.7 showed (a) map of Indonesia; (b) map of Java Island; and (c) Greater Jakarta with low lying areas along the coast. Figure 1.7 (c) showed also a longitudinal profile from the upstream point A and downstream point B; it shows drastic elevation changes from mountainous (A) to low lying coastal areas (B). This clearly implies that the low lying areas are prone to floods. Despites considerable flood risk management system introduced in the past decades, the flood impacts have worsened;

the main reasons are (1) rapid urbanization with not much of urban planning; (2) poor law enforcement on drainage maintenance (e.g. garbage dumping in rivers/canals). This study provides a baseline of flood risk mapping over Greater Jakarta with innovative cost-effective technology.


(a) Map of Indonesia

(b) Java Island

(c) Greater Jakarta catchment

Figure 1.7 (a) Map of Indonesia; (b) map of Java Island; and (c) Greater Jakarta catchment with low-lying areas shown in blue, less than 10 m above MSL

1.4 Research Motivations

To implement flood adaptation measures, it is imperative to couple topographical, hydrological and hydraulic understanding with model analysis. High spatial resolution and high accuracy DEM, and high temporal resolution and long rainfall records are important in flood hazard assessment using numerical model. Challenges in developing countries, such as Indonesia, are obtaining: (1) high accuracy DEM which is very costly and time consuming to acquire; and (2) good quality and long rainfall record required to derive Intensity-Duration-Frequency curves, design curves for drainage.

Figure 1.8 showed the schematic diagram of the process how the most relevant data for the flood simulations and analysis are cost effectively obtained. The research motivations and processes are summarized as follows.

1. DEM data from a publicly accessible remote sensing satellite will be selected and used to train an ANN to improve the quality of the remote sensing DEM. In the training of ANN, high quality observed DEM is the key leading to a well-trained ANN. The trained ANN will then be ready to efficiently and effectively generate high quality DEM, at low cost, for places where DEM data is not available.
2. Rainfall data, resulting from a high spatial resolution Regional Climate Model, RCM, will first be extracted, analyzed, and compared with regard to accuracy with good quality observed rainfall data of gauged catchments. The comparisons are performed, among others, on their IDF curves which are the essential design curves for storm drainage. After checking its high accuracy, the rainfall data (from the RCM) for an ungauged catchment will be extracted and readily used as proxies to derive the IDF curves for that ungauged catchment.
3. Data from the aforementioned two steps are then used as input to a widely used numerical flood model to generate flood data and the much needed flood map of various return periods. Policy makers will then be well informed about anticipated flood prone areas and flood extents to come up with best flood preventive and mitigation measures.

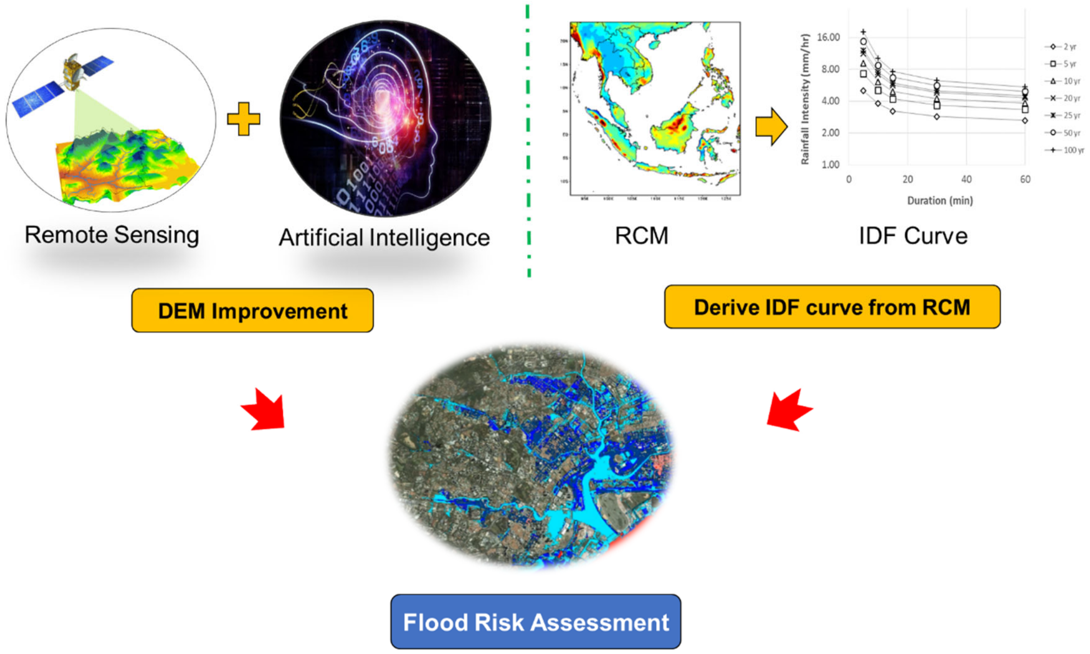


Figure 1.8 Schematic diagram of the research

1.5 Structure of Thesis

This thesis consists of six chapters. A brief description of each chapter is as follows:

- **Chapter 1** presents the background of the study, gaps, study areas, research motivations, and structure of thesis.
- **Chapter 2** presents a series of literature reviews relevant to (1) Digital Elevation Model (DEM): satellite data and multispectral imagery, and artificial neural networks with which cost effective DEM are derived; (2) Proxy for high temporal resolution and long rainfall data record: Regional Climate Model data (rainfall data in particular), and regional frequency analysis with which proxy rainfall data are then used to derive Intensity-Duration-Frequency curves; and (3) Numerical models: which assesse flood prone areas and flood extents.
- **Chapter 3** presents (1) the proposed methodology to improve publicly accessible DEM; and (2) the approach to select proxy rainfall data from downscaled climate model; suitability of data from various grid points will be checked using regional frequency analysis for derivation of the IDF curves which are relevant to storm drainage designs and flood analysis.
- **Chapter 4** demonstrates proof of concepts laid out in Chapter 3. It first shows the performance of the improved DEM through comparisons with their observed counterparts (in Nice, France, and in Singapore). It then presents the accuracy of derived high temporal and long rainfall proxy data using IDF curves comparison between model output and rainfall station data.
- **Chapter 5** presents the application of the aforementioned methodologies on Greater Jakarta (Indonesia) where both high resolution DEM and rainfall data are not available or not easily obtained. Flood map of two return periods (50- and 100-Years) are presented and compared with 2013 Jakarta flood footprint.
- **Chapter 6** summarizes and highlights the main results for each objective of the study; concludes the findings; and makes recommendations for future studies.

2 Literature Review

2.1 Introduction

This chapter reviews literatures relevant to remote sensing technology, artificial neural network, climate models and numerical flood modelling. To assess flood hazards, a series of data/information, such as digital elevation model and rainfall data, are required. In many countries, the developing ones in particular, these data are not readily available because (1) the high cost of the measurement and proper related tasks, or (2) no data record or data confidentiality. Hence publicly accessible remote sensing data, e.g. digital elevation model, are often the only option. However, the quality of remote sensing data, in some cases, requires further enhancement. This study considers Artificial Neural Network (ANN), a part of artificial intelligence technology or machine learning, to improve the accuracy of remote sensing data. Rainfall data in many developing countries, aside from data quality, are often of short record duration or not in existence. In areas of interest such as Greater Jakarta, Indonesia, this poses difficulty to engineers to arrive at appropriate design curves for drainages. As mentioned in Chapter 1, for Greater Jakarta, this study considers rainfall data, derived from a Regional Climate Model (RCM) driven by reanalysis data, as proxies. This information is parts of the essential input data for numerical flood model.

2.2 Remote Sensing Technologies

Remote sensing is the process of collecting information about an object area or phenomenon without physical contact [*Navalgund et al.*, 2007]. It has two aspects which are intimately linked with each other: the technology of obtaining the data through a device whose location is at a distance from the object, and analysis of the data for the interpretation of the physical objects [*Gupta*, 2018]. Going by the aforementioned definition various techniques of collecting the data where the object and sensor are not in contact with each other can be classified as remote sensing, for example photography, infrared, radiometers, radar (i.e. an object-detection system that uses radio waves to determine the range, angle, or velocity of objects) and Laser Imaging Detection and Ranging (LiDAR) (i.e. a surveying method that measures the illuminating target with pulsed laser light and the reflection of pulses with a sensor). Remote sensing can also be used as the technique of sensing the earth's surface from space by making use of the properties of electromagnetic wave emitted, reflected or diffracted by sensed objects, for the purpose of improving natural resource management, landuse and the protection of the environment.

Remote sensing has mainly evolved from the various methods of aerial photography and interpretation of the photos. It is a comparatively young discipline of science which has significantly grown over the past five decades. It has effectively improved human's ability to explore resources, map and monitor the Earth's environment globally and locally [*Lillesand et al.*, 2015; *Rencz et al.*, 1996; *Thenkabail*, 2015]. Earth observation is the main application of remote sensing. Various types of earth observations are listed below [*Entwistle et al.*, 2018; *Gupta*, 2018; *Kugler*, 2012].

- Weather forecasting
- Measuring land surface and mapping
- Tracking of biodiversity and wildlife trends
- Tracking of landuse changes (such as deforestation)
- Monitoring and responding to natural disasters, including forest/bush fires, floods, earthquakes and tsunamis
- Managing natural resources, such as energy, freshwater and agriculture
- Addressing emerging diseases and other health risks
- Predicting, adapting to and mitigating climate change.

Systematic and concise timelines of key developments in platforms and sensors for earth observations are well summarized in *Green and Jackson* [2009]. The first photo of the earth from the space was transmitted by Explorer-6 in 1959. It provided an orbital photography with the help of an unmanned camera. The Gemini mission of 1965 gave a number of vertical, stereo, oblique photographs of good quality. This demonstrated the potential of remote sensing techniques in the exploration of the resources on the earth. Later on, the experiments of the Apollo program included the coverage of the earth by multispectral 70-mm format photography and stereo vertical photography. The series of experiments on the photography led to the development of unmanned space orbital sensors.

2.2.1 Digital Elevation Models

Spaceborne radar or airborne laser scanning are widely applied to retrieve data on topography that is used to develop the Digital Elevation Model (DEM) [*Mirosław-Świątek et al.*, 2016; *Rodriguez et al.*, 2006; *Zhang and Montgomery*, 1994]. A DEM can be used to depict the terrain of the earth and is an organised array of the numbers which represent the elevations of spatial distributions above an arbitrary datum. The primary principle of a DEM is to describe the elevations of various points in a given area in digital format. The term DEM is usually applied to land surface topography, but it is a general term which is used to depict the spatial patterns of various surfaces e.g. surface water, ground surface, canopy, etc. Digital Surface Model (DSM) and Digital Terrain Model (DTM) are the two other terms which are frequently used for the ground terrain. DTM is referred as to the Earth terrain i.e. bare ground while DSM includes objects on ground like the buildings and trees, Figure 2.1.

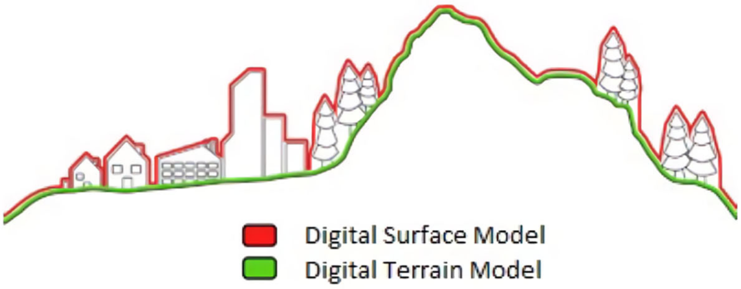


Figure 2.1 Digital surface model and digital terrain model [*Asharyanto et al.*, 2015]

A DEM can be obtained from various types of data sources. Traditionally, the ground survey data is most accurate but is also most expensive depending on the sampling density [*Bartosh*, 2012]. Recently the airborne laser scanning seems to be the most accurate method with the highest sampling density. It can record both object on surface and ground surface so that the elevation data is considered as the DSM [*Asharyanto et al.*, 2015; *Mirosław-Świątek et al.*, 2016]. Spaceborne interferometric radar system is a cost-effective technique to obtain the land cover and terrain data. A DEM can also be derived from radar satellite such as Shuttle Radar Topography Mission (SRTM DEM) [*Hensley et al.*, 2000; *Jacobs et al.*, 2001; *Kim et al.*, 2019; *Rosen et al.*, 2001], Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) [*Reuter et al.*, 2009; *Tachikawa et al.*, 2011], and TanDEM-X DEM [*Hajnsek et al.*, 2010].

SRTM DEM is an international joint project to collect three-dimensional digital mapping of over 80 % of the Earth's surface (between 60° N and 56° S) and it is available at no cost [*USGS*, 2000]. 3 arc-second resolution is available since 2005 and 1 arc-second resolution for globe is available after 2015. The performance requirements for the SRTM DEM data are such that the linear vertical absolute height error shall be less than 16 m and the relative height error shall be less than 10 m, for 90 % of the data [*Rodriguez et al.*, 2006]. It should be noted, however, that its accuracy is limited to Root Mean Square Error (RMSE) of approximately 14 m over Singapore's forest areas due to C-band wavelengths ($\lambda \approx 5.6$ cm) that does not adequately penetrate the vegetation canopy [*Wendi et al.*, 2016]. Thus, the elevation in vegetation area presents an intermediate height between top of canopy and the bare surface. Also, due to its coarse resolution (~ 30 m since 2015; ~ 92 m prior to 2015), it does not present precise urban characteristics.

The Global DEM (GDEM) is obtained from ASTER data, a product of Japan's Ministry of Economy, Trade, and Industry (METI) and USA's National Aeronautics and Space Administration (NASA) [*Tachikawa et al.*, 2011]. The ASTER GDEM produces a high resolution global digital elevation model, with 30 m spacing (1 arc second). Several literatures compared DEMs originated from SRTM DEM and ASTER DEM. *Guth* [2010] compared the data at 52 locations in Europe and North America and found that ASTER data was similar to SRTM DEM but about 20 % of ASTER data have anomalies that degrade its use for most applications. *Li et al.* [2012] conducted the evaluation of ASTER DEM, using GPS (Global Positioning System) as benchmarks, and SRTM DEM in China and concluded that ASTER DEM data requires further improvements as it appeared to overestimate the SRTM DEM data of the study area.

Graf et al. [2018] assessed the DTM for hydrogeomorphological modelling in small Mediterranean catchments, including SRTM DEM, ASTER DEM and LiDAR datasets. The RMSE results of the vertical accuracy show that SRTM DEM and ASTER DEM have differences of 6.98 m and 16.10 m respectively over the study areas due to systematic distortions and coarse horizontal resolution. The authors concluded that these limitations should be carefully considered when applying the data for numerical modelling.

Abily et al. [2015] developed a runoff model using very high resolution DSM. Two types of DSM data were used as topography data in the model and compared to LiDAR data only, and the combination of photogrammetric and LiDAR. Both data were able to capture the main buildings; but small buildings were not captured by LiDAR data. This resulted in significant differences in the flood map outputs. The authors recommended that fine-tuning topographic data is necessary for high resolution flood modelling.

The German Aerospace Center (DLR) has been operating Germany's first twin Synthetic Aperture Radar (SAR) satellites, TerraSAR-X and TanDEM-X DEM, to generate an updated global DEM shich has a spatial resolution of 0.4 arc-second (≈ 12 m) with 2 - 4 m in relative vertical accuracy [*Wessel et al.*, 2016]. *Gruber et al.* [2012] compared TanDEM-X DEM data against ground control points in Germany and US; they found that the absolute height errors are between 1 and 2 m. The elevation of TanDEM-X DEM is much more accurate than SRTM DEM; but TanDEM-X DEM data is not free (https://tandemx-science.dlr.de/). More detailed comparisons are given in Chapter 3.

In this study, SRTM DEM is selected to develop the DEM improvement scheme. Very high accuracy surveyed DEM is also used, in the DEM improvement scheme, for the Artificial Neural Network (ANN) to learn the patterns.

2.2.2 Multispectral Imagery

Multispectral imagery is produced by the sensors which measure the reflected energy within several specific bands/sections of the electromagnetic spectrum. It can be defined as "acquisition of images in hundreds of contiguous, registered, spectral bands such that for each pixel a radiance spectrum can be derived" [*Goetz et al.*, 1985]. The multispectral sensors have 3 to 10 different measurements of the band in each pixel of the images. Various earth observation satellites are being used to capture the images of the earth. Such satellites are called imaging satellites which are normally operated by commercial companies and governments around the world [*Nelli et al.*, 2018]. Over the years many countries have launched different satellites to acquire the images of the earth.

LANDSAT program is a network of the remote sensing satellites supported by NASA which provides repetitive acquisition of moderate-resolution multispectral data of the earth's surface on a global basis. The older satellites were gradually replaced by the more advanced and modern satellites. Presently, Landsat-7, launched in 1999, and Landsat-8, launched in 2013, are in operation. The information obtained from Landsat images meet the diverse needs of business, education, science, national security and the government. The data from the Landsat spacecraft represents the longest record of the earth's continental surfaces observed from space. It is a file unmatched in quality, detail, coverage, and value. Landsat is the only source of global, calibrated, moderate spatial resolution measurements of the earth's surface that are preserved in a national archive and freely available to the public [*Wulder et al.*, 2016]. The main objectives of Landsat 8 is to succeed the mission of Landsat 4, 5, 6 and 7 and to build, periodically refresh a global archive of sunlit, substantially cloud-free land images. Landsat 8 offers the features of data continuity, free standard data products, global survey mission, radiometric and geometric calibration, and responsive delivery.

Sentinel 2 is also an earth observation mission which was developed by the European Space Agency (ESA) as a part of Copernicus Programme to perform terrestrial observations in support of services such as forest monitoring, land cover changes detection, and natural disaster management [*Drusch et al.*, 2012]. It consists of twin polar orbiting satellites in the same orbit with a phase difference of 180 degrees with each other. The satellites were built by Airbus Defence Space, Sentinel-2A and Sentinel-2B, with two additional satellites being constructed by Thales Alenia Space. The Sentinel-2A multispectral instrument (MSI) obtains the reflective wavelength of the multispectral observations with directional effects caused because of the reflectance anisotropy of the surface [*Roy et al.*, 2017]. *Roy et al.* [2017] examined the magnitude of Sentinel-2A view zenith bidirectional reflectance distribution function (BRDF) effects observed for a large amount of data acquired over two 10-day periods across southern Africa acquired in the solar principal and orthogonal planes. An empirical c-factor approach was published that provides consistent Landsat view angle corrections to provide Nadir BRDF Adjusted Reflectance (NBAR) [*Roy et al.*, 2016]. Future Sentinel-2 and Landsat satellites may provide sufficient cloud-free observations to enable reliable local parameterization of the surface reflectance anisotropy over Sentinel-2 observation conditions.

The multispectral imagery can be used for land use classification, for seasonal monitoring, agricultural and environmental application [*Andres et al.*, 1994; *Ashish et al.*, 2009; *Moody et al.*, 2014; *Pande et al.*, 2018]. Using different reflectance values from different land use types, the area can be classified by clustering and machine leaning methods. *Kim et al.* [2018] analyzed the different reflectance of Sentinel 2 with different land uses. The reflectance of Short Wave Infrared (SWIR) bands (Bands 6-8) in forest areas is higher than that in urban areas; on the other hand, the reflectance of Near Infrared (NIR) bands (Bands 2-5) in urban areas is higher than that in forest areas. These different characteristics at each band help to classify land use in ANN as input nodes. These characteristics have been fully utilized for this study to generate the improved SRTM DEM using both multispectral imagery and ANN.

2.2.3 Artificial Intelligence

Artificial Intelligence (AI) is the recreation of human intelligence processes by machines, especially computer systems. These processes include learning (the acquisition of information and rules for using the information), reasoning (using the rules to reach approximate or definite conclusions), and self-correction [*Axelberg*, 2007].

In 1955, John MaCarthy, considered as the founder of AI, was the first person to introduce the term AI as to develop the machines that behave as though they were intelligent. Perception, learning, reasoning, problem-solving and language-understanding are the main components of AI [*Andresen*, 2002; *McCarthy*, 1956]. Some specialized areas of AI are game playing, expert systems, natural language processing, neural networks, and robotics etc. The advantages of AI, among others, are:

- It can take on stressful and complex work that humans may struggle/cannot do.
- It can complete a task faster than humans can.
- It can be used for discovering unexplored things.
- It yields less number of errors and, thus, less defects.
- It is more versatile when compared to humans.

ANN is one of the machine learning systems to achieve AI. ANNs apply mathematical learning algorithms which are simulated by properties of the biological neural networks. ANNs are loosely based on biological neural networks in such a way that they are implemented as a system of interconnected processing elements, sometimes called nodes, which are functionally analogous to biological neurons. The connections between distinct nodes have numerical values, called weights, and systematic altering of these values will give the ability to approximate the desired function [*Gurney*, 2014]. The characteristics of ANNs [*Kumar and Iyer*, 2010; *Sarve et al.*, 2015] are:

- It can map the input patterns to their associated output patterns.

- It can learn from the examples so that the ANN architectures can be trained with established examples of a problem before they are tested for their inference abilities for unknown instances of the problem. This helps in the identification of new objects which are not trained previously.
- It has generalization ability which helps predict the new outcomes based on the previous outcomes.
- The system can extract important features from incomplete, partial or noisy patterns.

The ANN is formed in three layers: input layer, hidden layer and output layer. The input layer has input neurons that transfer information via synapses to the hidden layer, and similarly the hidden layer transfers this information to the output layer via additional synapses. The synapses store values referred to as weights that help them to control the input and output to different layers. Figure 2.2 showed the schematic diagram of ANN.

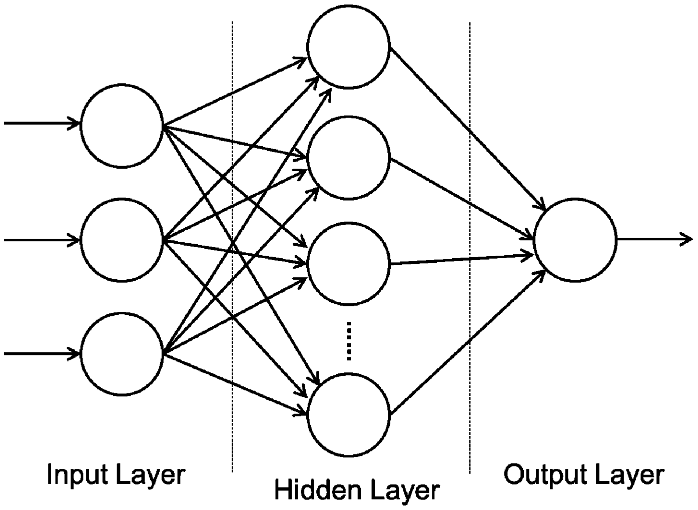


Figure 2.2 Schematic diagram of ANN layers [*Haykin*, 1994]

Each node within the network takes several inputs from alternative nodes and determines one output based mostly on the inputs and also the association weights. The network is able to converge to the optimal target function by the alteration in the weights systematically. Initially random values are assigned to weights and the network has to be trained to find the optimal weights. To achieve this the first output of the neural network has to be compared to the desired output, error is first determined; using this error the weights of the network are adjusted proportional to their contribution to the error in the output using back-propagation algorithm [*Rosenblatt*, 1961; *Widrow and Hoff*, 1962]

Together with the aforementioned characteristic of ANN, it has now been applied to system identification and control, quantum chemistry, game playing and decision making, pattern and sequence recognition, medical diagnosis and data mining. There are some applications of ANN in pattern recognition of remote sensing data.

The classification of images based on ANN uses a non-parametric path making it easy for the incorporation of the supplementary data while classifying so that the accuracy of the classification process is improved [*Abburu and Golla*, 2015]. In the training phase the ANN gains the information about the regularities which is present in the training data and then it will construct the rules which can be extended to the unknown data [*Foody*, 1999]. The main advantage of using ANN is that it can learn and generalize from inputs to produce a meaningful solution even when the input data contain errors or is incomplete. In the case of complex classification processes ANN algorithms are highly efficient [*Luk et al.*, 2000].

Kawabata and Bandibas [2009] utilized the ANN to generate the landslide susceptibility map using landslide data from an event of earthquake and a DEM derived from ASTER images. The ANN was trained using six geomorphic and geologic factors to produce the landslide hazard index map. The ANN was able to model the relationship between landslide occurrence and the factors.

Sun et al. [2016] applied the ANN to predict the ground water table in a freshwater swamp forest of Singapore. Unlike the physical modelling, the ANN-based approach did not require explicit characterization of the physical properties, or accurate representation of the physical parameters; it simply determined the system patterns based on the relationships between inputs and outputs mapped in the training process. The surrounding reservoir levels and rainfall information up to the immediate past 7 days were used as input in ANN. The forecast of the ground water table showed higher accuracy, as expected, for 1 lead-day than for 7 lead-days. The performance of longer lead-days results might be further improved if more variables, such as evapotranspiration, can be made available in the training of ANN process.

This study makes use of the strength of ANN in pattern recognition and classification to derive more accurate DEM. The ANN is able to classify the areas based on their reflectance values and identify the general error pattern, and reduce the errors between elevations of SRTM's DEM and reference DEM for different land uses from training process.

2.3 Downscaled Climate Model

Downscaling is the process of deriving the climate projections to scales which the decision makers require. The spatial scale, for dynamical downscaling particularly (discussed later in Section 2.4.1), depends very much on the computational resources made available as dynamical downscaling is computationally highly demanding [*Liu*, 2017]. There are various methods of downscaling with their own merits and demerits. International organizations or national governments currently provide no official guidance that assists researchers, practitioners, and decision makers in determining climate projection parameters, downscaling methods, and data sources that best meet their needs [*Daniels et al.*, 2012]. However, a large number of downscaling works have been carried out and shared in literatures which allow one to embark on the work easier now. Downscaling methods translate the large-scale coarse atmospheric fields (~100 - 300 km), used by GCMs (Global Climate Models), into regional- or local-scale information of climate variables (~5 - 50 km) required by climate change impact studies [*Feser et al.*, 2011]. Some of the applications benefits from downscaling of large scale information are:

- simulations of the spatial structure of near-surface temperature and precipitation over complex orographic terrain
- land use distributions
- regional and local atmospheric circulations that include jet cores, mesoscale convective systems

There are three fundamental approaches that exist for downscaling of large scale information to a regional or a local scale.

2.3.1 Dynamical Downscaling

The dynamical downscaling technique uses both physical and numerical models of the climate system by the mathematical formulation of the physical atmospheric processes, referred to as "parameterization schemes". Through this approach direct modelling of physical processes which characterize the climate of the region of interest. This method uses a Regional Climate Model (RCM) which is driven over a chosen limited area of the globe at high spatial resolution and hence also known as a limited area model [*Feser et al.*, 2011; *Giorgi*, 1990]. The main merit of this approach is that information provided by the RCM for the variables of the climate which are derived from the mesoscale atmospheric processes is of much higher spatial resolution compared to the GCM.

Soares et al. [2012] conducted a research on dynamically downscaled climatology of Portugal, and produced a high resolution (9 km) WRF (Weather Research and Forecast) simulation, driven by 20 years of ERA-Interim reanalysis (1989-2008). Model outputs were compared against all available stations including 32 daily temperature and 208 daily precipitation records. The results showed good representation of the annual cycles in each region. It should be noted that the model output was able to capture the occurrence of extreme, but rare, precipitation events with above 200 mm per day.

Liu [2017], using RCM WRF, downscaled present and future climate over Southeast Asia domain (80E – 125E; 15S – 26N) at a spatial resolution of 20 km; WRF was driven respectively by reanalysis data (ERA-Interim) and a series of GCMs. The present-day (1986 - 2005) model output was compared against stations data, gridded observations such as Climate Research Unit (CRU) and Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS). The comparisons served as a benchmark of model performance; and the comparison matched reasonably well. Figure 2.3, Figure 2.4 and Figure 2.5 showed the comparison, between WRF/ERAI (WRF driven by ERA-Interim) and some gridded observations, of winds, temperature and precipitation respectively. The well calibrated WRF model was then driven by the GCMs datasets of present climate and various future climate scenarios. The present-day WRF results, driven by GCMs, were first simulated and compared against gridded observations; the match was reasonably well. This gave a credible projection of the WRF results for the future climate data. The author analyzed the future climate for temperature and precipitation driven by three GCMs (Australian Community Climate and Earth-System Simulator (ACCESS1.3), European Centre HAMburg Model (ECHAM6) and Model for Interdisciplinary Research on Climate (MIROC5)) under two emission scenarios, Representative Concentration Pathway (RCP) 4.5 and RCP8.5 [*Allen et al.*, 2014]. It should be noted that this current study extracted only the output of WRF/ERAI of *Liu* [2017] to derive the IDF curves, of present days, over the study area.

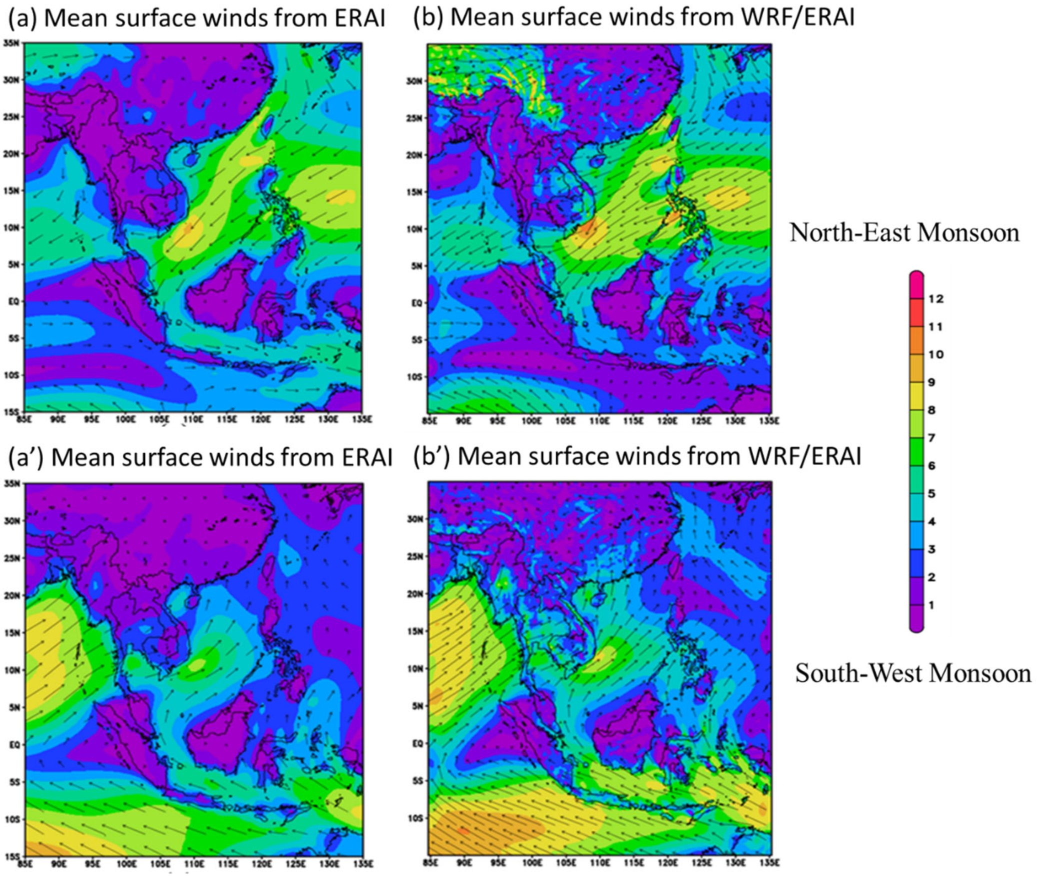


Figure 2.3 Mean seasonal surface winds (m/s) during Northeast Monsoon (top), and Southwest Monsoon (bottom), 1986-2005

(a), (a') ERAI, (b), (b') WRF/ERAI (Extracted from *Liu* [2017])

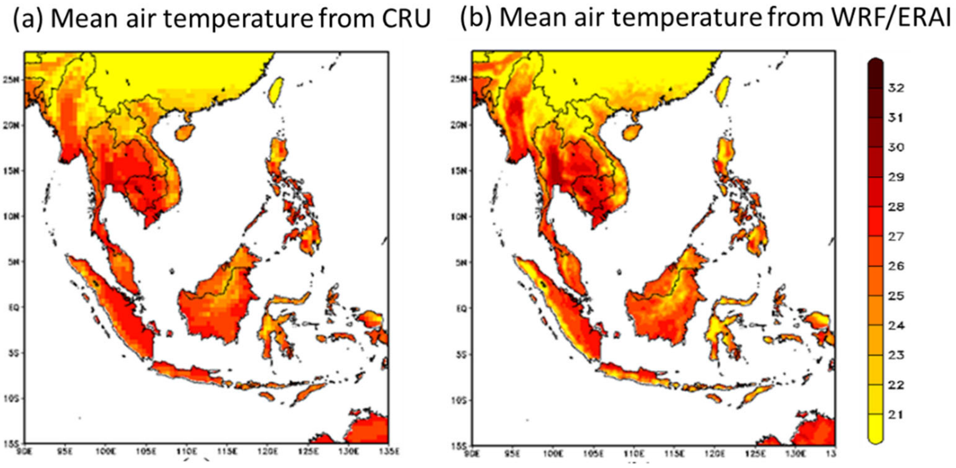
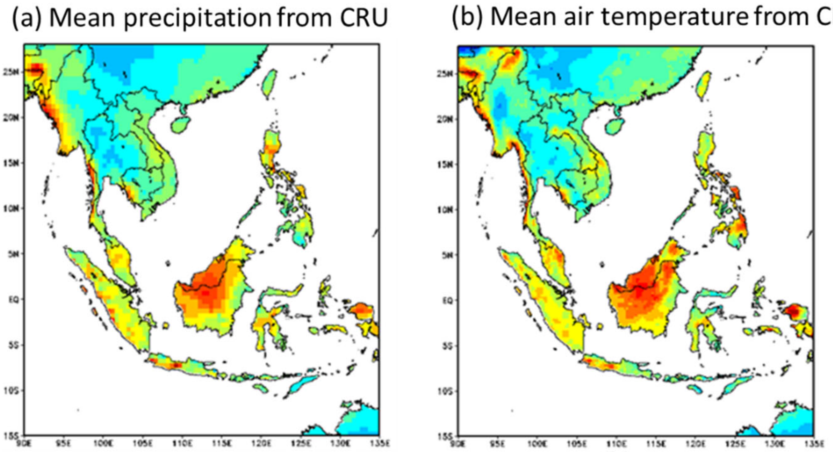


Figure 2.4 Climatological annual mean surface air temperature (°C), 1986-2005

(a) CRU (b) WRF/ERAI (Extracted from *Liu* [2017])


(a) Mean precipitation from CRU (b) Mean air temperature from CHIRPS

(c) Mean precipitation from WRF/ERAI

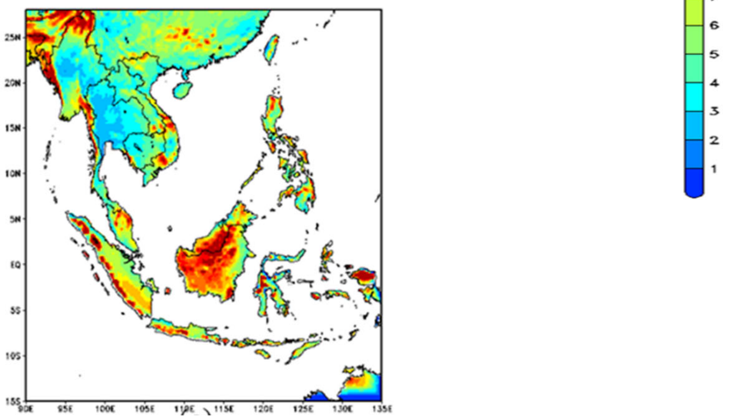


Figure 2.5 Climatological annual mean precipitation (mm/day), 1986-2005 (a) CRU (b) CHIRPS (c) WRF/ERAI (Extracted from *Liu* [2017])

2.3.2 Statistical Downscaling

This method establishes a statistical or empirical relationship between large scale and local scale atmospheric variables. This method usually requires three primary assumptions as listed below [*Benestad et al.*, 2008; *Giorgi*, 1990; *Hewitson and Crane*, 1996; *Robert et al.*, 2004]:

- High quality historical large scale atmospheric and local climate variables are available for a sufficiently long time period to establish robust relationships of the present climate.
- The statistical relationships derived under the present climate conditions remain time-invariant for different forcing conditions of possible future climates.
- The selected predictors are able to capture the climate change signal

Some difficulties of statistical downscaling techniques are: 1) Calibration requires sufficiently long and reliable observed data. 2) Poor reproduction of extremes, e.g. downscaling of heavy precipitation [*Haylock et al.*, 2006]. These drawbacks come from the fact that the calibration routines cannot address extreme data, and because the extremes are hardly well reproduced by the low-resolution GCM data [*Robert et al.*, 2004]. There are three statistical downscaling methods available, i.e. change factors, regression models and weather type approaches. Review and comparison of these statistical downscaling methods have been well described in various literatures (e.g., [*Fowler et al.*, 2007; *Hundecha et al.*, 2016; *Maraun et al.*, 2010; *Wilby and Dawson*, 2013])

2.3.3 Stochastic Downscaling

Stochastic downscaling is known as "Weather Generators" which generate synthetic time series of weather variables statistically similar to the ones of observed weather using statistical models. This method specifically links the variables of the local and large scale atmosphere with a statistical relation which is time invariant in current as well as future climates. The stochastic downscaling is better than statistical downscaling with respect to two main factors [*Burlando and Rosso*, 2002]. First, the technique of stochastic downscaling has the ability to generate efficient ensembles of synthetic time series through Monte Carlo simulations. Second, this method has a strong space time variability, which makes it useful tool for the evaluation of uncertainties. The main limitation of this technique is the requirement of a relatively long records and high quality data for calibrating the models. The simulated local scenarios are often sensitive to the parameter perturbation procedure, and thus a proper re-parameterization procedure of stochastic models is often very challenging and time-consuming. Moreover, some stochastic downscaling methods tend to underestimate inter-annual variability [*Maraun et al.*, 2010] and difficulties in reproducing low frequency climate variability [*Fatichi et al.*, 2011].

2.4 Rainfall Intensity Duration Frequency Curves and Regional Frequency Analysis

2.4.1 Rainfall Intensity Duration Frequency Curves

Intensity-Duration-Frequency (IDF) curves give a description of the relation among the intensity, duration of rainfall, and the return period (probability of exceedance). These curves are usually used for designing the hydrologic, hydraulic, and water resource systems. The IDF relationship was first established in 1930s [*Bernard*, 1932; *Sherman*, 1931] to capture the statistics of precipitation extremes, for the region of interest. The following three steps are followed for constructing an empirical IDF curve. Figure 2.6 illustrates derivation of IDF curve.

I. The Probability Density Function (PDF) or Cumulative Distribution Function (CDF) has to be fitted to each of the maximum annual rainfall data of a particular duration (e.g. 5, 10, 15, 60 minutes and 1 day).

II. For every recurrence interval or return period and duration, the intensity of the rainfall is calculated using the CDF selected in the previous step.

III. For every recurrence interval, the empirical IDF formula is considered to be a smooth function of the intensity of the rainfall with respect to the time duration obtained in the previous step. Least square method is used for the estimation of the IDF formula parameters.

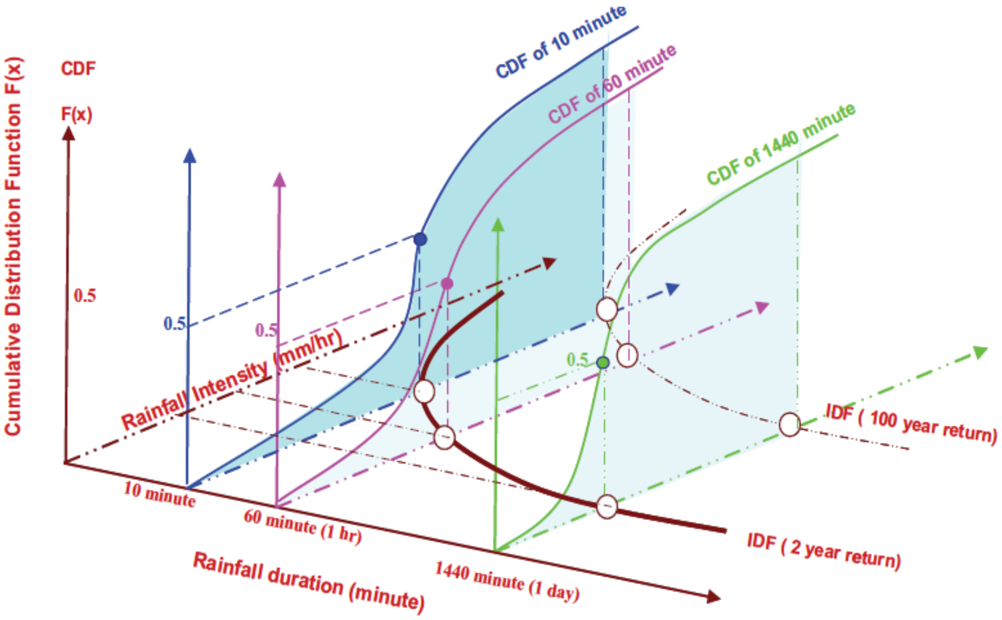


Figure 2.6 Illustration of derivation of IDF curve [*Nhat et al.*, 2006]

Koutsoyiannis et al. [1998] provided a detailed procedure and examples for constructing empirical IDF. Among the three steps of IDF curves construction, fitting the extreme values to the probability distribution function is very significant. The fitting procedure is referred as to “Extreme Value Theory” or “Frequency Analysis” [*Coles*, 2001; *Elsebaie*, 2012]. *Cunnane* [1978] reviewed a frequency analysis using a graphical method with different empirical plotting position formula. The ordered series of annual maximum rainfall intensities, ranking from the largest to smallest, is plotted against the return periods estimated by the empirical plotting position formulae (e.g. Gringorten plotting position formula) [*Gringorten*, 1963] on a specific probability distribution graph paper and a straight line is drawn to fit these points.

2.4.2 Regional Frequency Analysis

The Regional Frequency Analysis (RFA) utilises the data of several observation sites for estimating the distribution of the frequency for the data that is observed at each site. The IDF relationships primarily rely on the quality and length of the rainfall records. The two key steps related to RFA are regionalization (identifying the homogeneous region) and selecting the extreme value distribution of the region. Reliable frequency analysis of extreme rainfall events requires sufficiently long data records at that rain gauge station. For analysing the frequency of the single station, the length of the record for a single rain gauge needs to be more than 50 years for estimating the 100-year flood [*Das and Cunnane*, 2012]. But many regions around the world, primarily the developing countries, face the challenge of data scarcity and uncertainty. This is the main reason for the wide implementation of RFA. Some of the studies proved that RFA is more powerful when compared to the at-site frequency analysis, mainly when only short record lengths are available [*Ngongondo et al.*, 2011]. RFA leads to more accurate estimates than the traditional at-site analysis. *Weiss and Bernardara* [2013] compared four scaling factors, the at-site mean, the at-site median, an estimate of the location parameter of the parent distribution obtained by method of L-moments, and the trimmed mean. The at-site mean is the most accurate scaling factor in the presence of a slight regional heterogeneity or inter-site correlation; all scaling factors are equally good for a homogeneous region with correlated sites in terms of relative RMSE.

2.5 Numerical Modelling for Flood Analysis

Numerical modelling technique is a useful tool which is used to determine the dynamic behaviour, the causes and the effects of flooding. With the improvement in the computing technologies, many one dimensional (1D), two dimensional (2D), coupled 1D/2D and 3D numerical models have been developed. *Dimitriadis et al.* [2016] proposed the use of 1D and 2D models for assessing the uncertainty in the modelling of hydrological floodplain. *Bladé Castellet et al.* [2012] studied the conservation of mass and momentum by coupling 1D and 2D models for river channels and floodplain respectively. The use of mixed approach of 1D and 2D numerical models increases the quality of results and also saves time and computer memory which can be limiting factors for the application of 2D models [*Bladé Castellet et al.*, 2012]. Results of these models also are affected by the complexity and quality of topographic and input data.

Flood modelling can be divided into a number of approaches, characterised by their dimensionality or the way they combine approaches of different dimensionalities. Table 2.1 showed the classification of flood models (Adapted and modified from [*Neelz and Pender*, 2009; *Pender*, 2006]

Table 2.1 Classification of Flood models [*Neelz and Pender*, 2009; *Pender*, 2006]

Type	Description	Application	Typical computational time	Outputs	Example of models
1D	Solution of the one-dimensional St-Venant equations.	Design scale modelling which can be of the order of 10s to 100s of km depending on catchment size	Minutes	Water depth, Cross-section averaged velocity and discharge at each cross-section. Inundation extent if floodplains are part of 1D model, or through horizontal projection of water level	MIKE 11; Hec-Ras; ISIS; InfoWorks; RS
2D	Solution of the two-dimensional shallow water equations.	Design scale modelling of the order of 10s of km. May have the potential for use in broad scale modelling if applied with very coarse grids.	Hours or days	Inundation extent Water depths Depth-averaged velocities	TUFLOW; MIKE 21; TELEMAC; SOBEK; InfoWorks 2D

Neelz and Pender [2009] summarized the flood model classification in their research. One-dimensional models are based on some forms of the one-dimensional St-Venant or shallow water flow equations (Barré de St-Venant 1871), which can be derived by integrating the Navier-Stokes equations over the cross-sectional surface of the flow. The assumptions used in the derivation of the St-Venant equations limit their use to where the direction of water movement is aligned to the centre line of the river channel. Over the years their use has been extended to the modelling of flow in compound channels, that is, river channels with floodplains. In this case, floodplain flow is part of the one-dimensional channel flow and simulation of inundation is an integral part of the solution of the St-Venant equations. The technique has at least two disadvantages, namely that (1) floodplain flow is assumed to be in one direction parallel to the main channel, which is often not the case, and (2) the cross-sectional averaged velocity predicted by the St-Venant has a less tangible physical meaning in a situation where large variations in velocity magnitude exist across the floodplain. The approach has been enhanced in recent years for significant advances in parameterisation through the development of the conveyance estimation system [*McGahey and Samuels*, 2004].

Hydrodynamic models based on the two-dimensional shallow water equations are classified here as 2D approaches. The 2D shallow water equations (also referred to as 2D St-Venant equations, by extension to 2D of the use of this terminology, see [*Hervouet*, 2007]) can be derived by integrating the Reynolds-averaged Navier-Stokes equations over the flow depth. In this integration process, a hydrostatic pressure distribution is assumed (see [*Hervouet*, 2007]). A solution to these equations can be obtained from a variety of numerical methods (such as finite difference, finite element or finite volume) and use different numerical grids (such as Cartesian or boundary fitted, structured or unstructured) all of which have advantages and disadvantages in the context of floodplain modelling. The detailed information of Navier-Stokes equations can be found in Chapter 3.3.1.

The MIKE 21 Flow Model is a comprehensive modelling system for two-dimensional water modelling developed by DHI Water and Environment [*DHI*, 2017]. MIKE 21 Flow Model is applicable to the simulation of hydraulic and environmental phenomena in lakes, estuaries, bays, coastal areas and seas. Also modelling of tidal hydraulics, wind and wave generated currents, storm surges are some of the model applications.

Neal et al. [2010] addressed the issue of 2D hydraulic models that are time intensive to run due to their computational requirements. The authors recommended to run the model in parallel over multiple cores and it can reduce the computational time. *Mackay et al.* [2015] conducted 2D flood modelling using the approach of flexible mesh for large scale of catchment. The flexible mesh is more stable in the simulation than fixed grid as the terrain changes are smoothed. Also the user can define the different mesh size depending on the importance of the area. With CPU computation flexible mesh is slower than fixed grid; however, with GPU parallelization, the speed can be increased up to 8 times [*DHI*, 2014].

The effect of the input data's quality in the flood model is very significant [*Casas et al.*, 2006; *Duong and Gourbesville*, 2016; *Gourbesville et al.*, 2015]. *Prinadiastari and Bahri* [2018] assessed the flash flood events in North Sumatra, Indonesia, using 2D flow simulation. The flash flood occurred as a result of natural dam break which is triggered by heavy rainfall with long duration and change of land use on the upstream area. The model results were compared with field survey data for flood depth and there were some discrepancies. The author found that the discrepancies were mainly due to the rough resolution of DEM and lack of actual discharge data as these are the main sources of the data for computational simulations.

2.6 Summary

The literature reviews conducted focussed on remote sensing technology, artificial intelligence, downscaling of climate models, development of IDF curves, Regional Frequency Analysis (RFA), and numerical flood modelling. The following is a summary of the literature reviews:

- The assessment of freely available DEMs (from satellite remote sensing) showed that the original DEM dataset contains its abnormality, systematic errors and sensor limitations which result in low accuracy in their applications, for example, in flood modelling. These data have to be reprocessed. This study proposed a novel method to reprocess them using multispectral imagery and ANN to significantly improve their accuracy levels.
- Regional Climate Model driven by reanalysis data showed also its worthiness for its downscaled precipitation data used as proxies for ungauged sites and/or sites with short rainfall record. With these proxy data, the rainfall Intensity-Durational-Frequency (IDF) curves, essential for drainage designs, can be derived. For a large scale catchment, the study also review the procedure to identify the homogeneity of the rainfall zone; Regional Frequency Analysis is recommended to perform this homogeneity identification task.
- Different types of numerical model are available to address the flooding issues. This study adopts the 2D flood model with flexible mesh method to simulate the large scale catchment. In this study the GPU parallelization technique is adopted to speed up the computation time.

The following chapters aim to present the detailed data, methodologies and models used for the study and discuss results related to some of the key issues raised in this chapter.

3 Methodology and Data

3.1 Overview

To overcome data scarcities on flood hazard simulations and assessments, this study introduces the following:

- A methodology to derive improved DEM, with high accuracy, using publicly accessible multispectral imagery (SRTM DEM and Sentinel 2) and Artificial Neural Networks technique (ANN).
- Selecting proxies for rainfall data from high resolution regional climate model (Weather Research and Forecast, WRF, 20 x 20 km) driven by perfect climate model reanalysis data (ERA-Interim), WRF/ERAI. Several key parameters resulting from WRF/ERAI are validated against observation data. After checking its accuracy level, these proxies are significantly valuable particularly for ungauged catchments or gauged catchments with short rainfall record. These data are then used to derive the drainage design curves, the Intensity-Duration-Frequency (IDF) curves, for the catchment of interest.
- A numerical model, MIKE21FM, is used to generate flood data of Greater Jakarta, Indonesia, using the aforementioned DEM data from improved DEM, and design storms originating from the IDF curves.

3.2 Derivation of High-Accuracy DEM

As mentioned in Chapter 2, the freely accessible DEM data often contains its abnormality, systematic errors and sensor limitation, which cause the DEM data to be less accurate. Therefore, in the areas where high-accuracy surveyed DEM is not available, additional studies are required to obtain reliable topography and to overcome the uncertainty introduced by the measurement protocol. As shown in Figure 3.1 SRTM DEM has two main limitations: (1) as sensors do not penetrate the vegetation area, the top of the canopy level represents the elevation in forest area; (2) with its coarse resolution, it does not allow to present the precise urban characteristics, i.e. averaged elevation between (low lying) road and high rise buildings. The impacts of these limitations are quite serious particularly for flood assessment as road levels become unrealistically high.

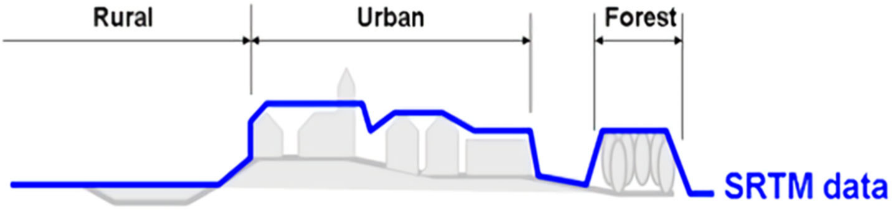


Figure 3.1 Limitations of SRTM DEM on the scanning of surface [*Radiomobile*, 2018]

Different DEMs from different sources have been compared in an urban area of Nice, France, as shown in Figure 3.2 and Figure 3.3. Figure 3.2 (a) is the satellite image of the area showing the land characteristics. Figure 3.2 (b) is a high accuracy surveyed DEM with 1 m resolution from Nice Côte d'Azur Metropolis (France), an urban area. It can be seen that the 1 x 1 m resolution DEM gives much clearer views of land and building shapes than their counterparts from TanDEM-X DEM (12 m resolution; Figure 3.2 (c)) or SRTM DEM(30 m resolution; Figure 3.2 (d)). This 1 x 1 m resolution DEM will be used in this study as the reference DEM. Note that TanDEM-X DEM can better capture the roads and buildings than SRTM DEM where the land shapes are smudged due to its coarse resolution and the limitation of the sensor. Figure 3.3 showed the performance of SRTM DEM and TanDEM-X DEM compared to surveyed 1 x 1 m resolution DEM. The Root Mean Square Error (RMSE), used to evaluate the performance of SRTM DEM and TanDEM-X DEM, showed 8.36 m and 7.24 m respectively.

Figure 3.4 and Figure 3.5 showed similar comparative study but for forest (vegetated) area. Land shapes are similar to each other but SRTM DEM showed less clear shape due to its coarse resolution. In term RMSE, SRTM DEM and TanDEM-X DEM showed 14.37 m and 2.05 m respectively. Thus, the performance accuracy of TanDEM-X DEM is much higher than SRTM DEM as it uses a unique Interferometric Synthetic Aperture Radar (InSAR) [Krieger et al., 2011; Martone et al., 2018].

Although SRTM DEM showed less clear land shape, it can still be the only option for some places where high-accuracy or surveyed DEM is not available.

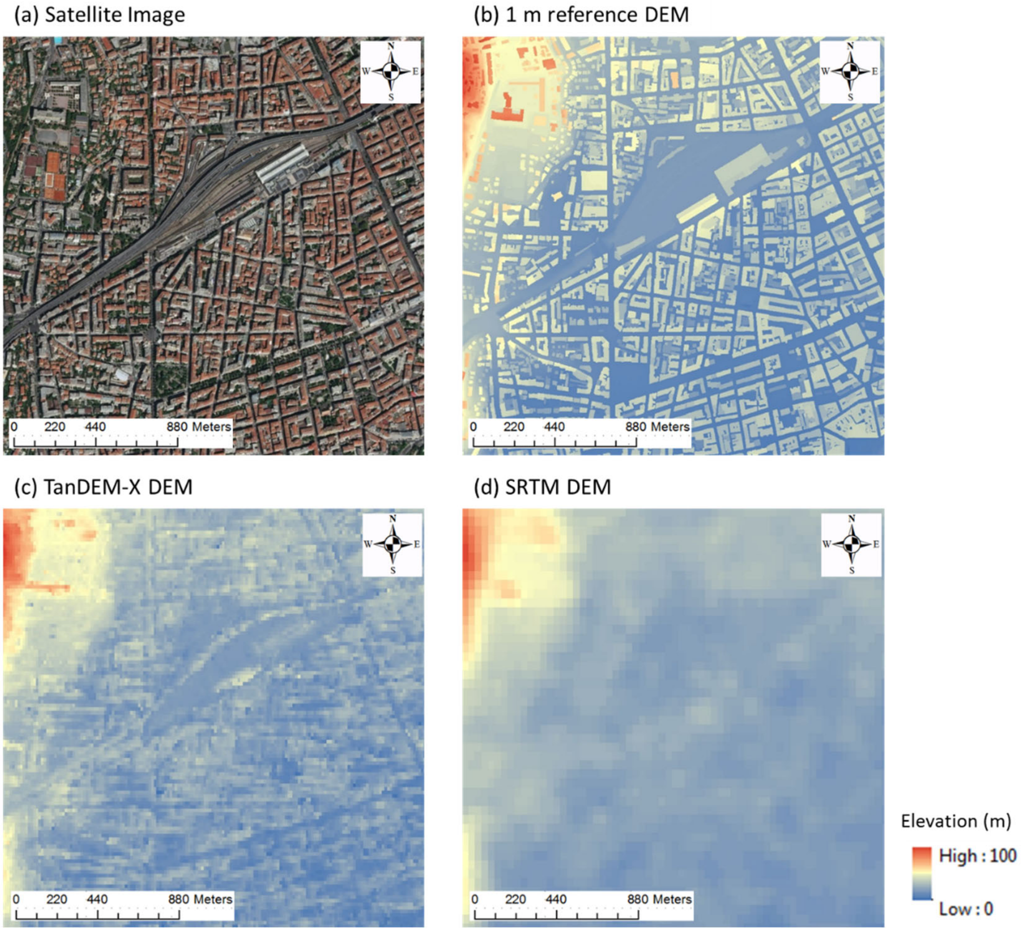


Figure 3.2 Different DEMs from different sources; urban area in Nice, France; (a) satellite imagery, (b) surveyed DEM (1 m resolution), (c) TanDEM-X DEM (12 m resolution), (d) SRTM DEM (30 m resolution; publicly accessible satellite data)

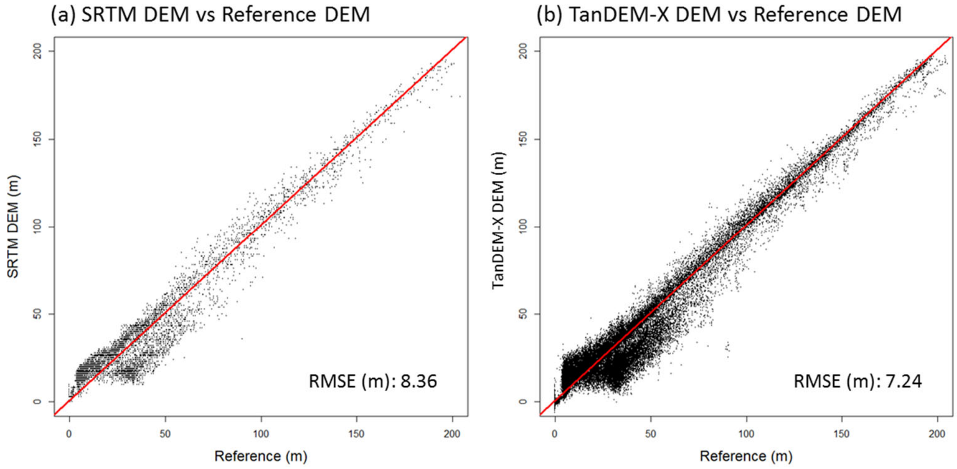
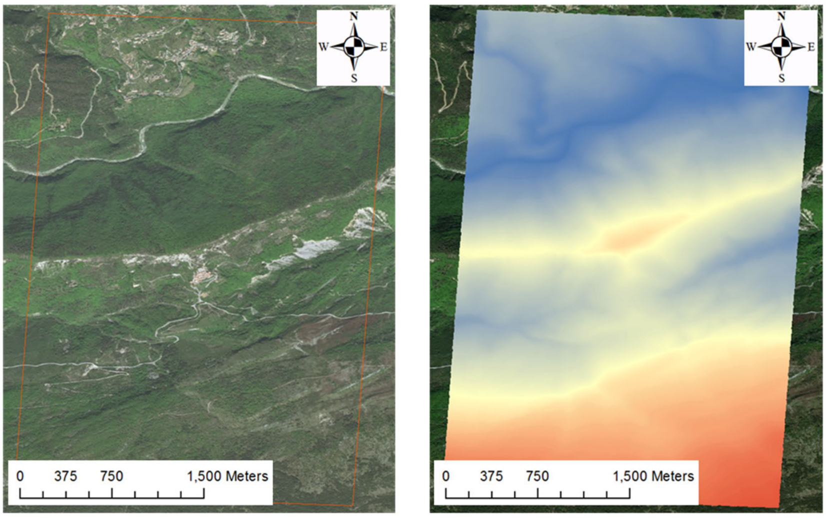
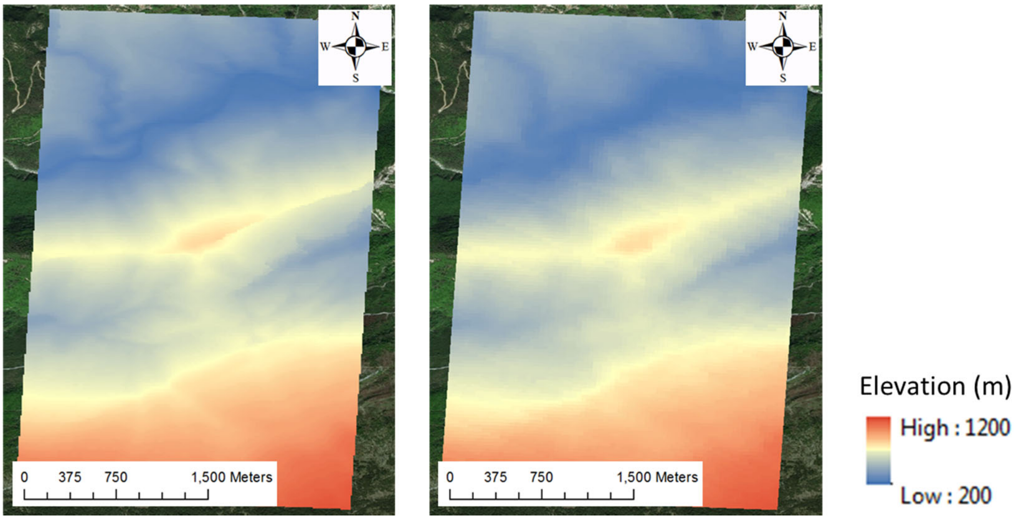


Figure 3.3 Performance of (a) SRTM DEM and (b) TanDEM-X DEM compared to surveyed/reference DEM over an urban area in Nice, France



(a) Satellite Image

(b) 1 m reference DEM



(c) TanDEM-X DEM

(d) SRTM DEM

Figure 3.4 Different DEMs from different sources; forested area in Nice, France; (a) satellite imagery, (b) surveyed DEM (1 m resolution), (c) TanDEM-X DEM (12 m resolution), (d) SRTM DEM (30 m resolution; publicly accessible satellite data)

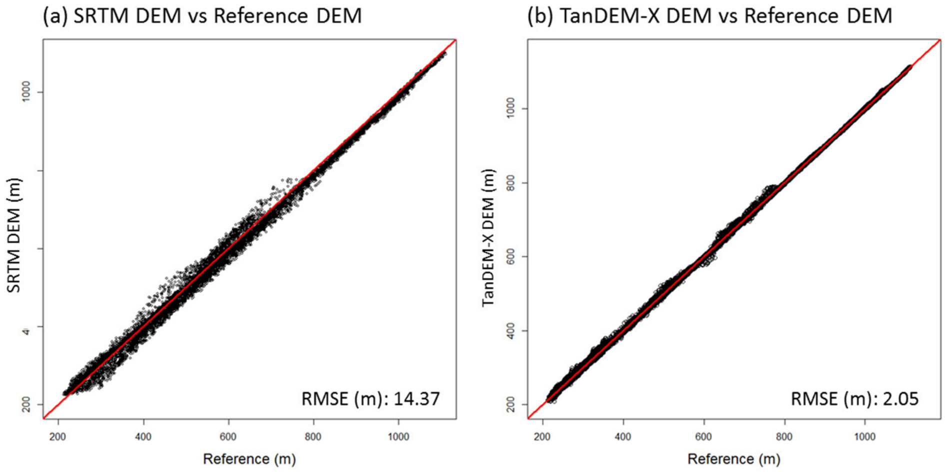


Figure 3.5 Performance of (a) SRTM DEM and (b) TanDEM-X DEM compared to surveyed/reference DEM over forested area in Nice, France

The improved DEM scheme is developed using remote sensing data and ANN technique, as mentioned in Chapter 2. Figure 3.6 demonstrates the schematic diagram of DEM improvement methodology. Generally it requires 3 types of data; multispectral imagery, the DEM to be improved (SRTM DEM in this study), and a reference DEM (high accuracy elevation). These data are input for the ANN for training, and later for validation. Once the performance of the trained ANN is acceptable, it can be applied to areas where their SRTM DEMs are to be improved.

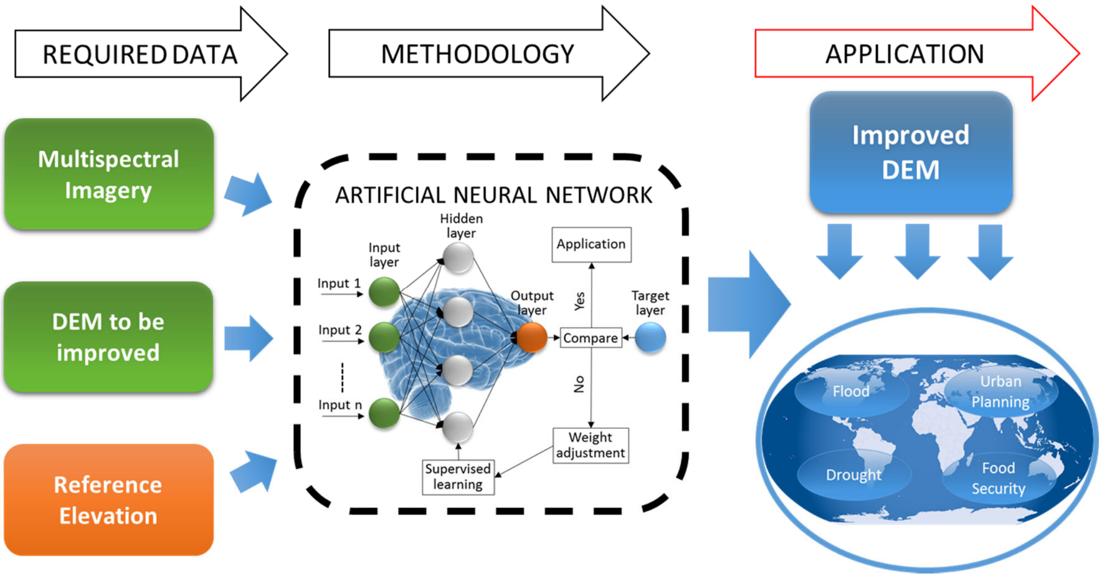


Figure 3.6 Schematic diagram of DEM improvement methodology

3.2.1 Data Pre-Processing

Since all of the remote sensing data have different resolutions (i.e. SRTM DEM 30 m; Sentinel 2 10 - 60 m; surveyed DEM 1 m), all input layers need to be standardized to a common resolution through resampling method as shown in Figure 3.7. In this study, 10 m resolution is chosen for the assessment of the developed methodology.

Figure 3.8 showed the required data and study areas for the development of DEM improvement scheme. High resolution and high accuracy DEM (surveyed) is available in Nice, France. The data has been provided by Geographic Information Service, Metropolitan of Nice Côte d'Azur Metropolis. The 1 m reference DEM is used as a reference DEM to train and validate the ANN. The reference DEM of Singapore has been used for both ANN training and validation. SRTM DEM and Sentinel 2 data are used as common input for all areas. Further details of the data are discussed here.

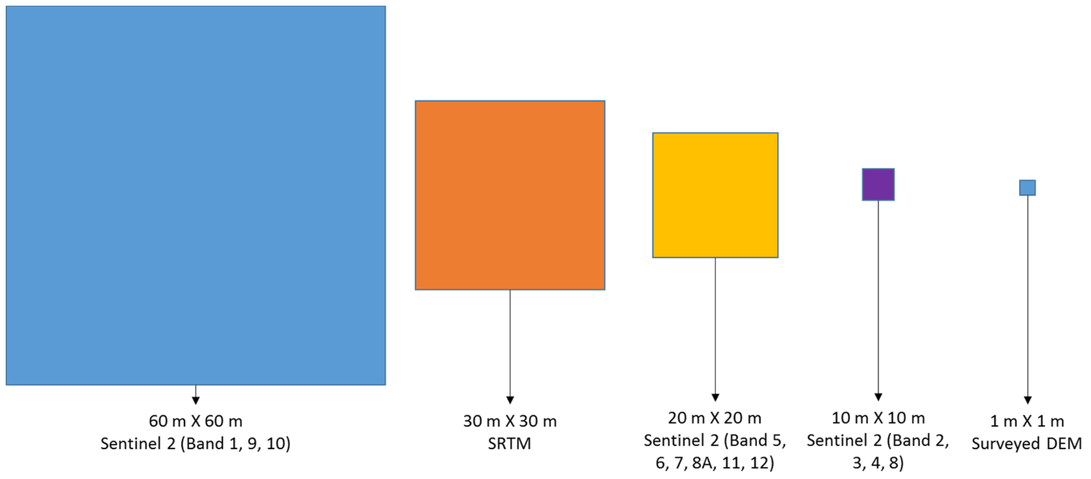


Figure 3.7 Different resolutions of different remote sensing data from Sentinel 2, SRTM DEM, TanDEM-X DEM and Surveyed DEM

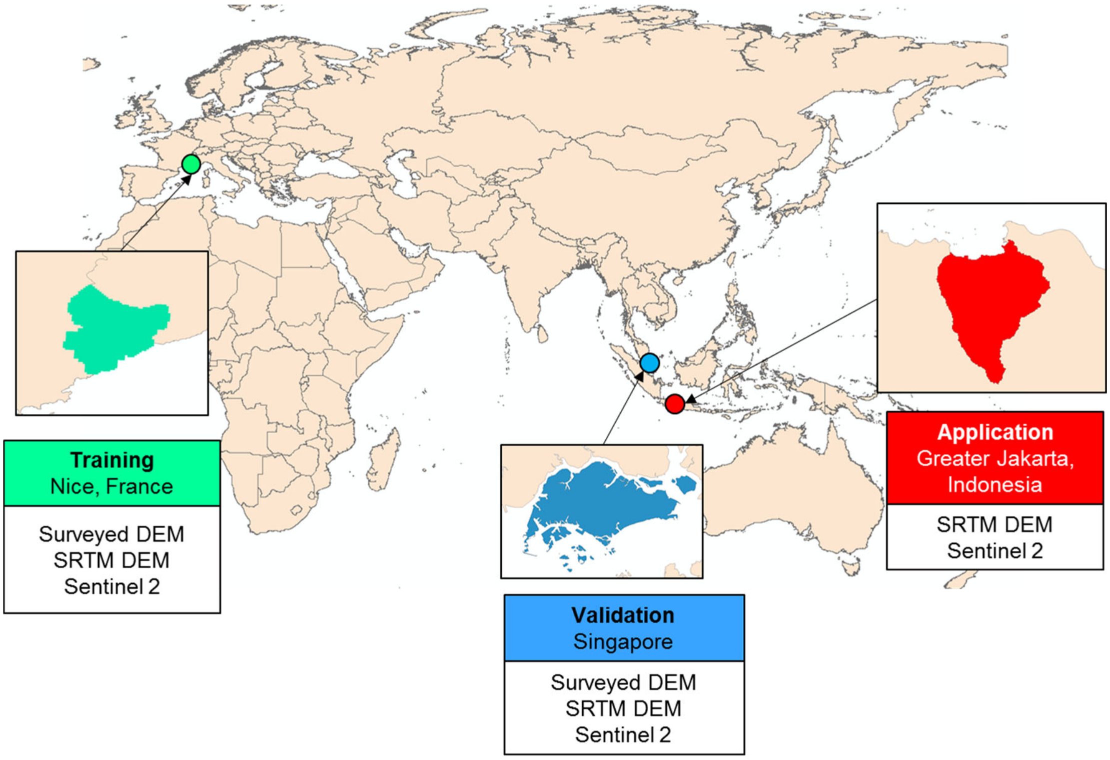


Figure 3.8 Areas used for ANN's training (Nice, France), validation (Singapore) and application (Greater Jakarta, Indonesia) of DEM improvement scheme

3.2.1.1 SRTM DEM Data Pre-Processing

As mentioned in Chapter 2, SRTM DEM is widely used for mapping and geography application as it is publicly accessible. The data is available from USGS Earthexplorer website (https://earthexplorer.usgs.gov/). Figure 3.9 showed the global data coverage of SRTM DEM.

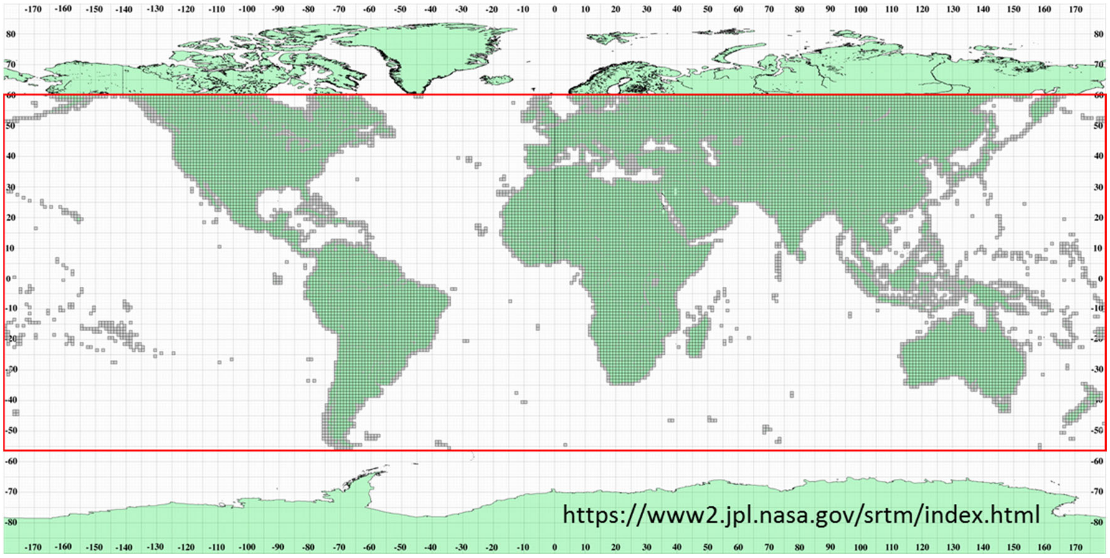
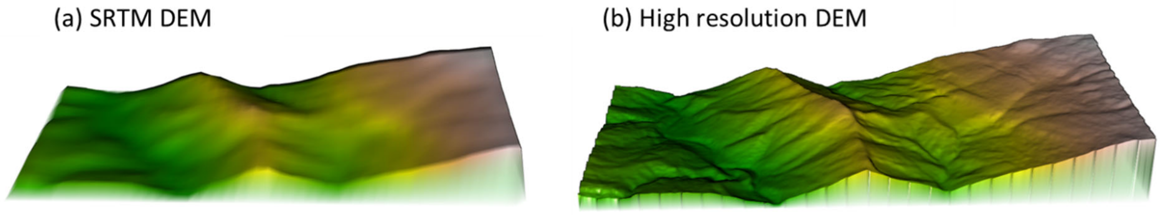


Figure 3.9 Data coverage of SRTM DEM (https://www2.jpl.nasa.gov/srtm/index.html)

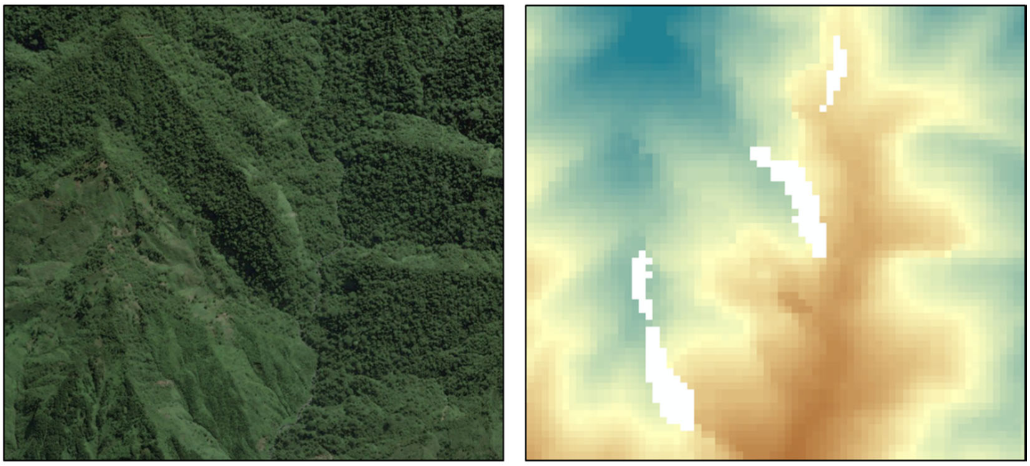
One of the significant limitations of SRTM DEM is the original data; they may not represent the actual topography due to its systematic distortions and coarse horizontal resolution [*Graf et al.*, 2018; *Rodriguez et al.*, 2006; *Wendi et al.*, 2016]. Figure 3.10 showed the example of differences between SRTM DEM and high resolution DEM. The land shapes are much clearer in the high resolution DEM (i.e. streams, valleys and mountains) so it is highly recommended to reprocess the original SRTM DEM data before its usage.


(a) SRTM DEM

(b) High resolution DEM

Figure 3.10 Sharpness comparison between (a) SRTM DEM and (b) high-resolution DEM

Additionally, SRTM DEM has its own issues when it is used in topographical applications. Although USGS updated the data with filled void (missing data), it still can be captured in some areas due to geometric artifacts, specular reflection of water, phase unwrapping artifacts and voids due to complex dielectric constant [*François*, 2001; *Reuter et al.*, 2007]. This is the reason why the SRTM DEM data requires significant levels of pre-processing to fill the missing or no data. In this study, the neighbouring interpolation method was used. The neighbouring interpolation is a simple method of interpolation and it provides a smoother approximation on the unknown values using neighbouring values [Sibson, 1981]. As shown in Figure 3.11 (a) the area is classified as the mountain with steep slopes and no data areas are captured (Figure 3.11 (b)). Using the aforementioned method the void areas were calculated and filled (Figure 3.11 (c)).



(a) Satellite imagery

(b) SRTM DEM with void hole

(c) SRTM DEM without void hole

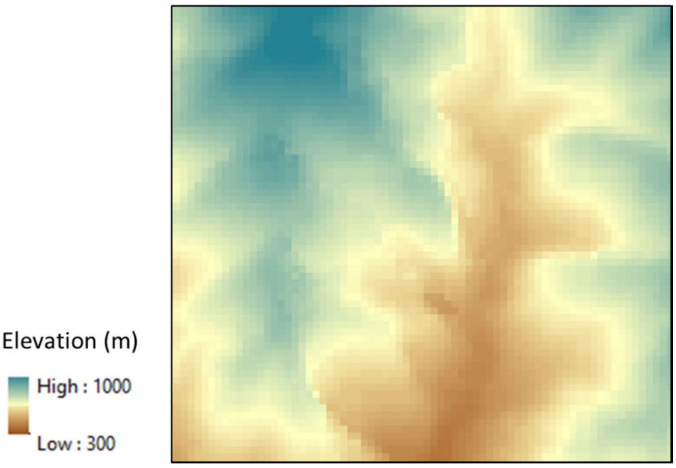


Figure 3.11 SRTM DEM void filling with interpolation method; (a) satellite imagery, (b) void hole in SRTM DEM data, (c) void hole filled after interpolation using data from neighboring cells

3.2.1.2 TanDEM-X DEM Data Pre-Processing

The German Aerospace Center (DLR) has been operating Germany's first two formations flying Synthetic Aperture Radar (SAR) satellites, TerraSAR-X and TanDEM-X DEM, with the objective to generate an updated global DEM. The DEM has a spatial resolution of 0.4 arc-second (≈ 12 m) with 2-4 m in relative vertical accuracy [*Wessel et al.*, 2016]. The data are also available with a larger spacing of 1 arc-second (≈ 30 m) and 3 arc-second (≈ 90 m). The vertical datum of TanDEM-X DEM is WGS84-G1150 ellipsoidal heights and it has changed to geoid system to standardize the datum system with SRTM DEM. Table 3.1 showed the specification of the TanDEM-X DEM product.

Table 3.1 TanDEM-X DEM product overview [*Wessel et al.*, 2016]

Independent Pixel Spacing of TanDEM-X DEM	Absolute Horizontal Accuracy	Absolute Vertical Accuracy	Relative Vertical Accuracy	Coverage
~ 12 m (0.4 arc-second at equator)	< 10 m	< 10 m	2 m (slope ≤ 20 %) 4 m (slope > 20 %)	Global
~ 30 m (1 arc-second at equator)	< 10 m	< 10 m	Not specified	Global
~ 90 m (3 arc-second at equator)	< 10 m	< 10 m	Not specified	Global

- Absolute horizontal accuracy is defined as the uncertainty in the horizontal position of a pixel with respect to a reference datum, caused by random and uncorrected systematic errors. The value is expressed as a circular error at 90 % confidence level.
- Absolute vertical accuracy is the uncertainty in the height of a pixel with respect to a reference height caused by random and uncorrected systematic errors. The value is expressed as a linear error at 90 % confidence level.
- Relative vertical accuracy is specified in terms of the uncertainty in height between two points (DEM pixels) caused by random errors. The corresponding values are expressed as linear errors at 90 % confidence level. The reference area for two height estimates is a 1° x 1° area, corresponding to approximately 111 km x 111 km at the equator.

The data can be obtained from https://tandemx-science.dlr.de/. Figure 3.12 showed the global data coverage of TanDEM-X DEM. Note that the 90 m resolution of TanDEM-X DEM is publicly accessible since September 2018.

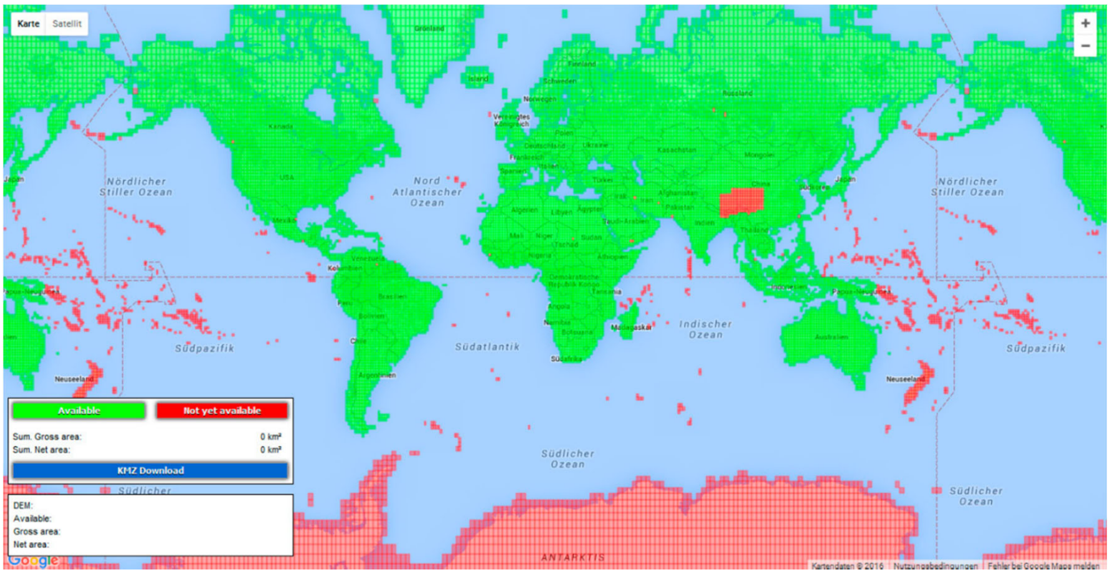


Figure 3.12 Data coverage of TanDEM-X DEM (status as of August 2016)

Additional pre-processing of TanDEM-X DEM is required prior its usage. The very noisy appearance of water bodies in the DEM can be observed due to the temporal de-correlation and low backscattering. Consequently, the corresponding elevation values derived from the interferogram are random and produce meaningless values [*Wendleder et al.*, 2013]. TanDEM-X DEM provides the water indication mask product to detect the disturbed surface of water bodies.

In this study the water body areas are replaced by SRTM DEM values. SRTM DEM water body was portrayed as a series of 'stepped down' elevations to maintain proper water to land relationship [*USGS*, 2003]. Figure 3.13 showed a comparison of the values on the water body in TanDEM-X DEM, SRTM DEM and surveyed high accuracy DEM.

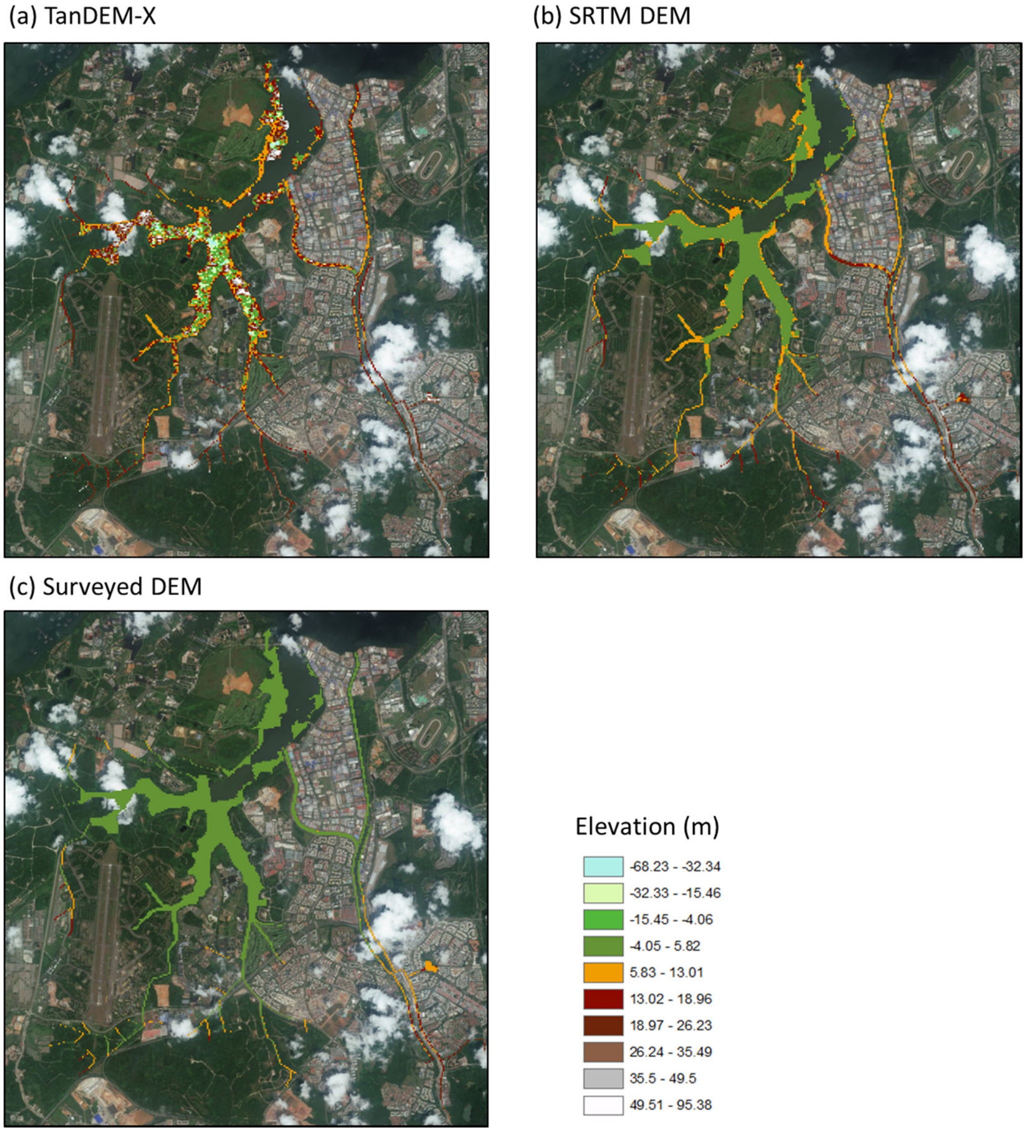


Figure 3.13 Limitation of TanDEM-X DEM on water body; (a) TanDEM-X DEM, (b) SRTM DEM, (c) surveyed DEM

3.2.1.3 **Sentinel 2 Multispectral Imagery Pre-Processing**

Sentinel 2 Mission is an European earth polar-orbiting satellite (Sentinel-2A and 2B) designed to feed the Global Monitoring for Environment and Security (GMES) system with continuous and operational high-resolution imagery for the global sustained monitoring of earth surface and coastal areas [*Gatti and Bertolini*, 2018]. The Sentinel 2 system is based on the concurrent operations of two identical satellites flying on a single orbit plane but phased at 180º, each hosting a Multi-Spectral Instrument (MSI) covering from the visible to the shortwave infrared spectral range and delivering high spatial resolution imagery at a global scale and with a high revisit frequency. The MSI aims at measuring the earth reflected radiance through the atmosphere in 13 spectral bands spanning from the Visible and Near Infra-Red (VNIR) to the Short Wave Infra-Red (SWIR):

(1) 4 bands at 10 m: blue (490 nm (nano meter)), green (560 nm), red (665 nm) and near infrared (842 nm).

(2) 6 bands at 20 m: 4 narrow bands for vegetation characterisation (705 nm, 740 nm, 783 nm and 865 nm) and 2 larger SWIR bands (1610 nm and 2190 nm) for applications such as snow/ice/cloud detection or vegetation moisture stress assessment.

(3) 3 bands at 60 m mainly for cloud screening and atmospheric corrections (443 nm for aerosols, 945 for water vapour and 1375 nm for cirrus detection).

It provides a 5-day revisit frequency of a dual spacecraft operations. Figure 3.14 and Table 3.2 showed the MSI spectral bands and their wavelength, bandwidth and resolution.

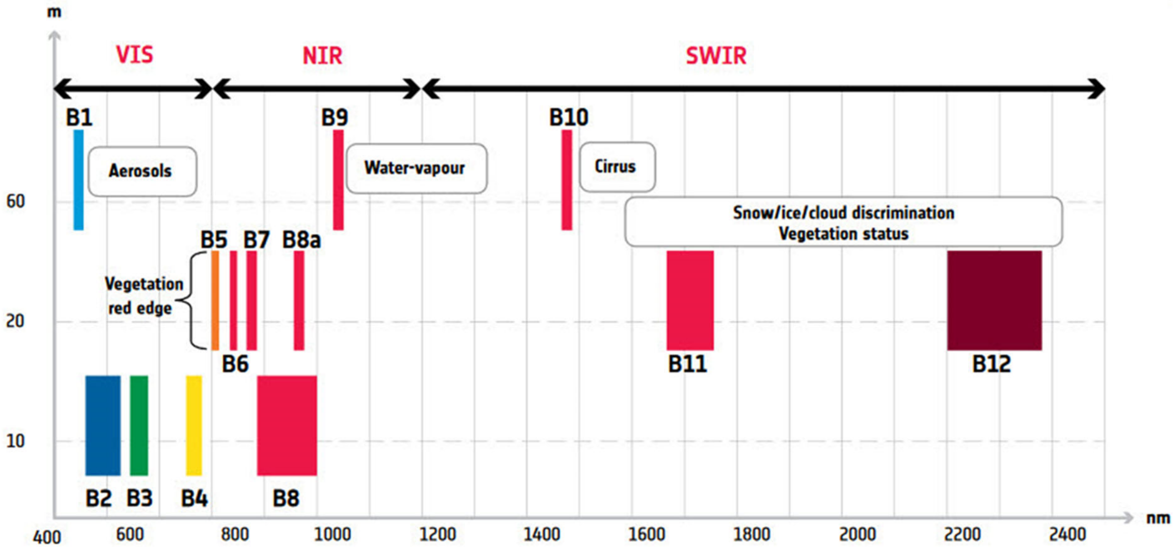


Figure 3.14 MSI spectral bands versus spatial resolution [*Gatti and Bertolini*, 2018]

Table 3.2. Sentinel 2 spectral bands [*Gatti and Bertolini*, 2018]

Bands	**Central wavelength (nm)**	**Band width (nm)**	**Pixel resolution (m)**
Band 1 – Coastal aerosol	443	20	60

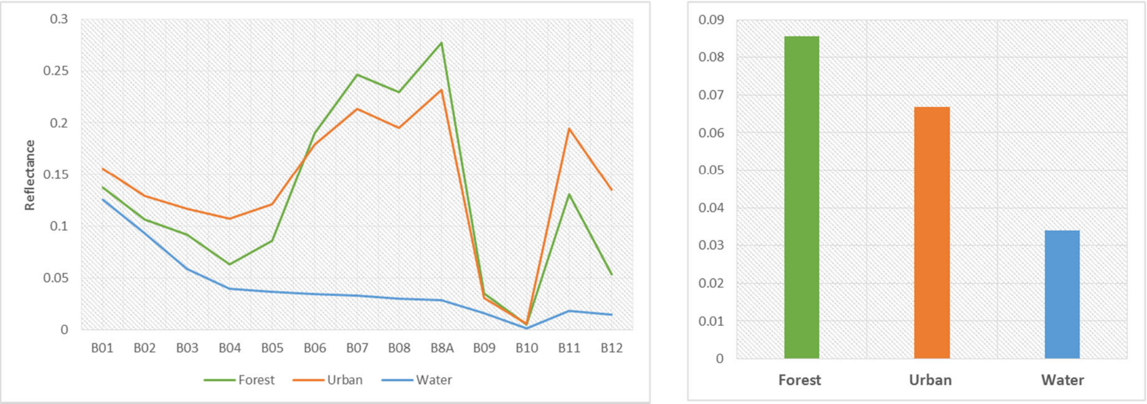
Band 2 – Blue	490	65	10
Band 3 – Green	560	35	10
Band 4 – Red	665	30	10
Band 5 – Vegetation Red Edge	705	15	20
Band 6 – Vegetation Red Edge	740	15	20
Band 7 – Vegetation Red Edge	783	20	20
Band 8 – NIR	842	115	10
Band 8A – Narrow NIR	865	20	20
Band 9 – Water vapour	945	20	60
Band 10 – SWIR – Cirrus	1380	20	60
Band 11 – SWIR	1610	90	20
Band 12 – SWIR	2190	180	20

The data is available from Copernicus Open Access Hub (https://scihub.copernicus.eu/) or USGS Earthexplorer website (https://earthexplorer.usgs.gov/). The products are available for users with different processing levels as shown in Table 3.3. When users download the data, they can choose several criteria for each data set (i.e. cloud coverage, orbit direction, platforms). In this research, Top-Of-Atmosphere (TOA) from Level-1C products were used with less than 10 % of cloud coverage.

Table 3.3 Sentinel 2 product processing levels

Processing Level	Description	Remark
Level-0	Compressed raw data and contains all the information required to generate upper levels	Not available to users
Level-1A	Uncompressed raw data with spectral bands coarsely coregistered and ancillary data appended	Not available to users
Level-1B	The physical geometric model is refined using available ground control points and appended to the product, but not applied	Not available to users
Level-1C	Provides orthorectified Top-Of-Atmosphere (TOA) reflectance, with sub-pixel multispectral registration. Cloud and land/water masks are included in the product	Online available
Level-2A	Provides orthorectified Bottom-Of-Atmosphere (BOA) reflectance, with sub-pixel multispectral registration. A Scene Classification map (cloud, cloud shadows, vegetation, soils/deserts, water, snow, etc.) is included in the product	Available in Euro-Mediterranean region since March 2018 Globally available since December 2018

The bands 2, 3, 4 and 8 are used for land cover classification and the bands 5, 6, 7 and 8A are mainly dedicated to vegetation area. Bands 1, 9, 10, 11 and 12 are used for atmospheric correction and cloud screening. In this study, 8 bands (2, 3, 4, 5, 6, 7, 8 and 8A) are used for the ANN nodes with standardized resolution. Figure 3.15 (a) showed the different reflectance values based on various land use over Nice, France. The reflectance in forest area is higher than urban area at SWIR bands (Bands 6-8) while urban area is higher than forest area at NIR bands (Bands 2-5). Generally the forest area has higher values in standard deviation than the urban and the water body areas (Figure 3.15 (b)). These different characteristics at each band help to classify land use in ANN as input nodes. Table 3.4 showed the average of reflectance values from each band with different landuses.


(a) Averaged reflectance from each band for different landuses

(b) Standard deviation of different landuses

Figure 3.15 Different reflectance of Sentinel 2 in different landuses; (a)-average values of reflectance from each band for different landuses, (b) standard deviation of different landuses

Table 3.4. Average of reflectance values with different landuses

Landuse	B01	B02	B03	B04	B05	B06	B07	B08	B8A	B09	B10	B11	B12
Forest	0.137	0.106	0.092	0.063	0.086	0.190	0.246	0.230	0.277	0.035	0.005	0.130	0.053
Urban	0.156	0.129	0.116	0.107	0.121	0.179	0.214	0.195	0.232	0.031	0.006	0.194	0.135
Water	0.125	0.093	0.059	0.040	0.037	0.035	0.033	0.030	0.029	0.016	0.001	0.018	0.015

3.2.1.4 **Preparation for ANN Input, Target Layers**

As mentioned in Session 3.2.1, all of the remote sensing data with different resolutions are standardized to common resolution through resampling method as shown in Figure 3.16. In this study, 10 m standardized resolution was used for performance evaluation of the methodology and 30 m resolution was used for assessing Greater Jakarta flood maps.

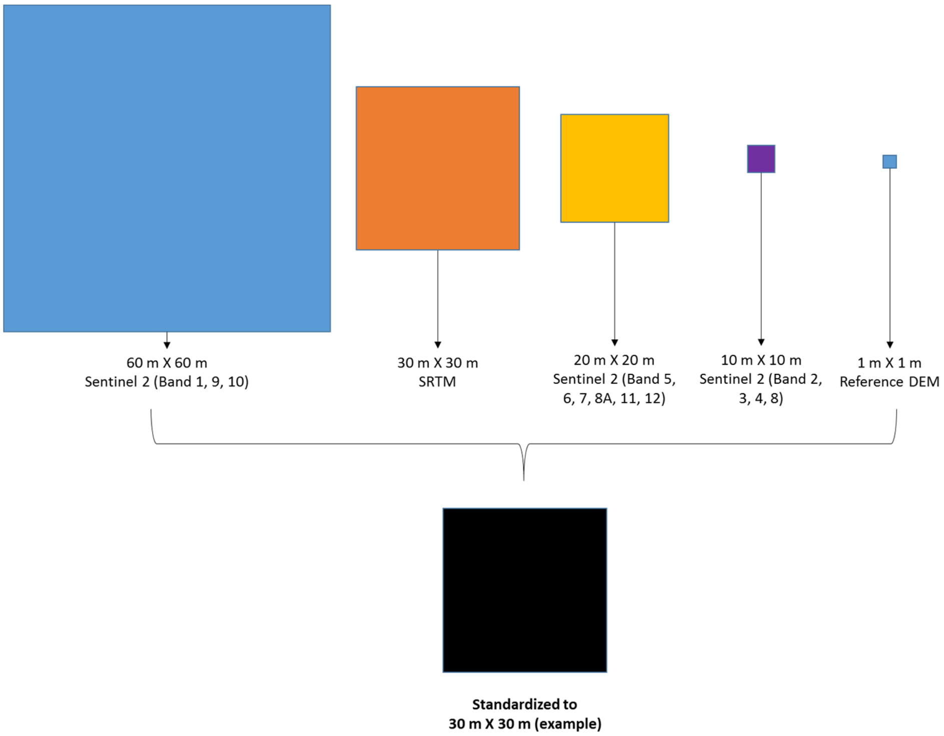


Figure 3.16 Standardization of different resolutions from different sources

All remote sensing data were processed using ArcGIS desktop software developed by Environmental System Research Institute (ESRI). ArcGIS is a Geographic Information System (GIS) for working with maps and geographic information. It is used for compiling geographic data, analysing mapped information, and managing geographic information in a database [*ESRI*, 2018].

All raster layers from its original data set were standardised into common resolution and matched cell alignment (extent and origin of each cell). The raster format of data was extracted by point type of shapefile and the attribute table of the shapefile can generate the table which is numeric inputs to the ANN. The extracted information from the table is then used to ANN as input and target layers as shown in Table 3.5. The detailed ANN setup is discussed in the following session.

Table 3.5 Artificial Neural Network layers

Input Layer	Target Layer	Output Layer
Reflectance values of Sentinel 2, multispectral imagery SRTM DEM elevation	Surveyed DEM elevation	Improved (Rectified) elevation

3.2.2 Artificial Neural Network Setup

Matlab Neural Network Toolbox was used for developing DEM improvement scheme in this study. It provides a neural network to generalize nonlinear relationships between inputs and outputs using feedforward backpropagation networks. The feedforward neural network is the first and simplest type of ANN devised [*Schmidhuber*, 2015]. It contains multiple neurons (nodes) arranged in layers and all these have the connections. It consists of four types of nodes:

(1) Input Layer: The input layer of a neural network consists of artificial input neurons, and carries the initial data into the system for further processing by subsequent layers of artificial neurons. No computation is performed in any of input nodes; it just passes the information to the hidden nodes.

(2) Target Layer: Target is the desired output for the given input.

(3) Hidden Layer: A hidden layer is the layer between input and output layers, where artificial neurons take in a set of weighted inputs and produce an output through an activation function.

(4) Output Layer: The output layer in an artificial neural network is the last layer of neurons that produce given outputs for the program. They are responsible for computations and transferring information from the network to the outside.

The ANN algorithm is divided into 3 steps:

(1) Training/Learning: The network processes the input and compares its resulting outputs against the target layer. The errors are then propagated back through the system to adjust the weights.

(2) Validation: This is used to measure the performance of the network generalization, and to halt training when generalization stops improving.

(3) Testing: This has no effect on training and thus provides an independent measure of network performance after the training.

The network is trained with Levenberg-Marquardt (LM) backpropagation algorithm [*Levenberg*, 1944; *Marquardt*, 1963]. This method is a standard technique for solving nonlinear least squares problems to fit a curve by minimizing the sum of the square of the errors between input and output nodes. The training is continued until the validation error ceased to decrease; the trained ANN is then applied to test data set. In this study the data set was randomly divided into 70 % for training, 15 % to validate the network to stop training before the overfitting and 15 % for independent testing of network generalization. Table 3.6 showed the example of input, target and output layers in ANN.

Table 3.6 Input, Target and Output Layers in Artificial Neural Network (example)

Input Layer									Target Layer	Output Layer
B02	**B03**	**B04**	**B05**	**B06**	**B07**	**B08**	**B8A**	**SRTM DEM (m)**	**Reference DEM (m)**	**Improved DEM (m)**
0.0898	0.0884	0.0604	0.0922	0.2329	0.3059	0.2901	0.3296	26	23.95	To be calculated
0.089	0.0865	0.0608	0.0928	0.2408	0.3137	0.2787	0.3416	27	24.32	
0.0835	0.0749	0.0454	0.0866	0.2321	0.2976	0.2733	0.325	27	30.24	
0.0933	0.0927	0.0715	0.0892	0.2442	0.3112	0.2762	0.3503	27	25.55	
0.0879	0.0797	0.0517	0.0676	0.2178	0.272	0.2666	0.308	26	25.08	
0.0856	0.0761	0.0496	0.0834	0.2172	0.2736	0.2367	0.3096	25	29.17	
0.0944	0.0901	0.0695	0.0976	0.2332	0.2924	0.2684	0.3247	26	23.39	
0.0884	0.0856	0.0515	0.0915	0.2488	0.3174	0.2898	0.3507	28	22.93	

3.3 Rainfall Data from Regional Climate Model (RCM) used as Rainfall Proxies

This section discusses the selection of rainfall proxies, of the present climate, for the development of Intensity-Duration-Frequency (IDF) curves for sites with either short or no rainfall record at all. For present climate, dynamical downscaling of reanalysis data must first be performed to obtain high-resolution climate outputs from Regional Climate Model (RCM), Weather Research and Forecasting (WRF). Reanalysis data used in here to drive WRF was ERA-Interim (denoted henceforth as WRF/ERAI) over Southeast Asia Domain (Figure 3.17). The data was validated with gridded observation data. Upon satisfactory validation, the simulated rainfall data for Java Island, Indonesia, are then extracted from WRF/ERAI and used to construct the IDF curves for Java.

Figure 3.17 Domain of Regional Climate Model, WRF (20 km resolution): 90°E - 135°E, 15°S - 28°N

3.3.1 Climate Data from Downscaled Model

As mentioned in Section 2.4.1 (Limitation of GCMs), downscaling method translates the coarse spatial resolution atmospheric fields into regional or local scale information of climate variables to provide detailed information with finer spatial and temporal resolutions. To do this, RCMs need to be driven by large scale fields such as reanalysis or GCMs. Here, the study of *Liu* [2017], who conducted the dynamical downscaling method using WRF/ERAI, is chosen. ERAI is one of the latest global atmospheric reanalysis, produced by the European Centre for Medium-Range Weather Forecasts (ECMWF), and widely used by the scientific community. ERAI has two strengths compared to its counterpart: (1) a more advanced data assimilation model in the representation of hydrological cycle, the quality of the stratospheric circulation, and the consistency in time of reanalysed geophysical fields; (2) more robust technical aspects of reanalysis such as data selection, quality control, bias correction and performance monitoring [*Dee et al.*, 2011]. The spatial resolution of ERA-Interim is approximately 80 km on 60 vertical levels from the surface up to 0.1 hPa. The dataset is available at 6-hourly interval since 1979 and continuously updated in real time. Further information of the data set can be found at http://apps.ecmwf.int/datasets/ and documented by *Dee et al.* [2011].

WRF/ERAI provides high spatial resolution information (20 km) of climate variables which are much finer than the original data in ERAI. Table 3.7 showed the comparison of the different data sets of ERAI and WRF/ERAI.

Table 3.7 Basic characteristic of ERAI and WRF/ERAI data

Source	Scale	Resolution (km)	Frequency (Hourly)
ERAI	Global	~ 80	6
WRF/ERAI (Southeast Asia)	Regional	20	6

3.3.2 Validation of Rainfall Proxies from WRF/ERAI

The downscaled precipitation data from the WRF/ERAI, i.e. WRF driven by ERAI, was first extracted from *Liu* [2017]. Its performance was compared with one of their counterparts from the widely used gridded observation precipitation data such as CHIRPS (Climate Hazards Group InfraRed Precipitation with Station; [*Funk et al.*, 2015]).

CHIRPS data is more than a 30 year quasi-global rainfall dataset, developed by Earth Resources Observation and Science (EROS) Center at USGS, Climate Hazards Group at University of California at Santa Barbara and Earth Systems Research Laboratory at National Oceanic and Atmospheric Administration (NOAA). This daily precipitation product covers 50°S - 50°N and is available from 1981 to near-present. CHIRPS incorporates 0.05° resolution satellite imagery with in-situ station data to create gridded rainfall time series for trend analysis and seasonal drought monitoring [*Funk et al.*, 2015]. As of February 12th, 2015, version 2.0 of CHIRPS is complete and freely accessible to the public. There are two spatial resolutions available, 0.25° and 0.05°. In Liu's study (2017), the product from 0.25° was used. Detailed information on CHIRPS can be found at http://chg.geog.ucsb.edu/data/chirps/ and *Funk et al.* [2015]). The dataset was used to evaluate WRF/ERAI performances on representation of precipitation climatology and extremes.

The WRF/ERAI model simulated 6-hourly precipitation data from 1986-2005. Their daily mean precipitation, Simple Day Intensity Index (SDII, reflecting the mean precipitation for the wet day), and the 95th percentile of precipitation (P95p) were compared with their counterparts from CHIRPS as shown in Figure 3.33 to Figure 3.35. Both SDII and P95p represent the precipitation extremes. Detailed information can be found in *Liu* [2017].

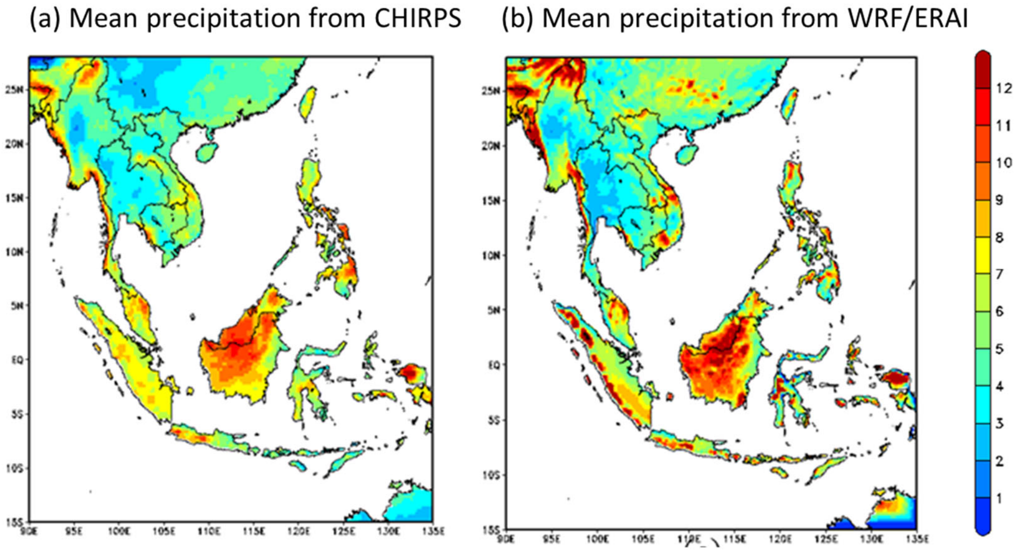

(a) Mean precipitation from CHIRPS (b) Mean precipitation from WRF/ERAI

Figure 3.18 Comparison of climatological daily mean precipitation (mm/day; 1986-2005) (a) CHIRPS, (b) WRF/ERAI (extracted from *Liu* [2017])

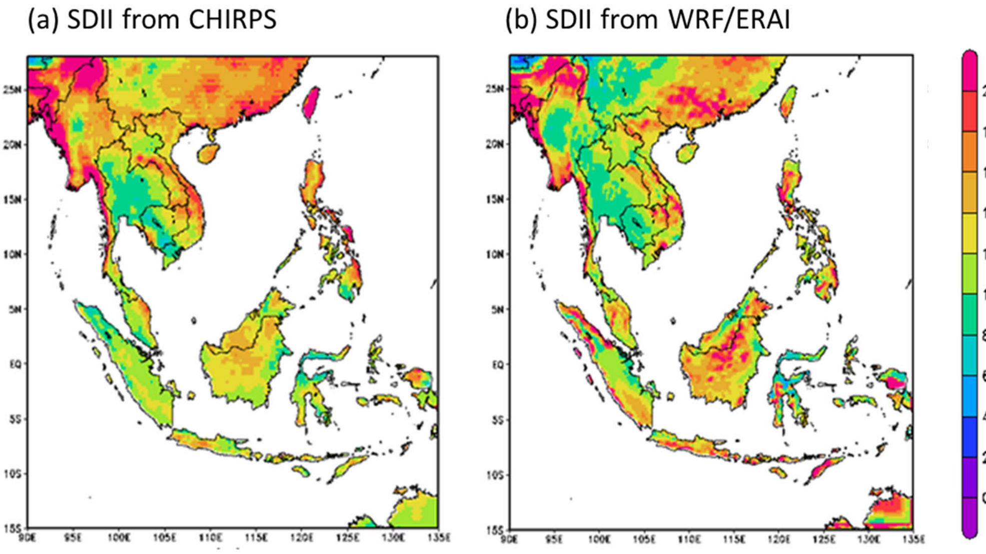


Figure 3.19 Comparison of climatological daily Simple Day Intensity Index of precipitation (SDII; mm/day; 1986-2005) (a) CHIRPS, (b) WRF/ERAI (extracted from *Liu* [2017])

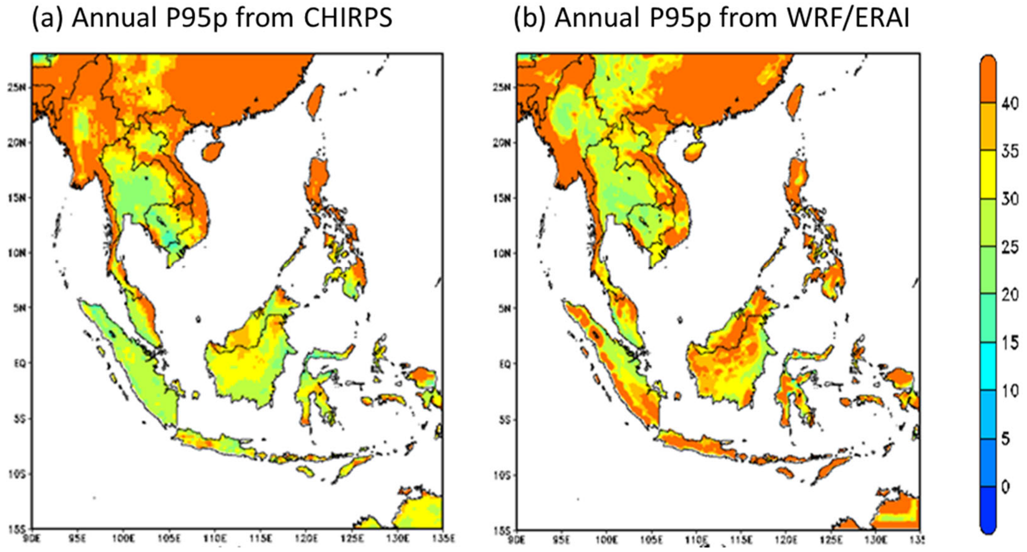


Figure 3.20 Comparison of climatological annual 95th percentile of precipitation (P95p; mm/day; 1986-2005) (a) CHIRPS, (b) WRF/ERAI (extracted from *Liu* [2017])

3.3.3 Regional Frequency Analysis (RFA)

Hosking and Wallis [1997] developed a complete algorithm for the regional frequency analysis method based on the approach of L-moments by pooling the sites with similar statistical characteristics in a homogeneous region instead of a single site in the at-site frequency analysis.

In conducting the regional frequency analysis, the heterogeneity measure is the primary indicator for accepting or rejecting a proposed region (grouping of sites). The discordancy measures for the various sites provide a secondary indicator to consider whether a discordant site should be moved to another region [*Nunez et al.*, 2011].

To generate the flood maps of study area, Greater Jakarta, with the rainfall proxies extracted from WRF/ERAI, RFA method was again first applied and validated in Singapore with observation data. In this method, Annual Maximum Rainfall (AMR) of 6-hourly rainfall for 36 years (1981 – 2016) was considered.

3.3.3.1 Initial Data Screening by Discordancy Measure

A discordancy measure is useful to identify sites with gross errors in their data or those that are grossly discordant with the region as a whole. In practice, discordancy measure suggested by *Hosking and Wallis* [1993] is widely used by hydrologists. To estimate discordancy values for sites in a region, the sites are considered as points in three-dimensional space of sample L-moment ratios (L-CV (L-Coefficient of Variation), L-Skewness and L-Kurtosis). Centroid of the region is regarded as a point depicting average of sample L-moment ratios of the sites in the region. Any point that is far from the centroid of the region is flagged as discordant.

The main aim of discordancy measure filters out the outlier sites from the pooling area. The discordancy measure at site i, D_i among N sites, introduced by *Hosking and Wallis* [1993], is expressed by

$$\boldsymbol{D_i = \frac{1}{3}N(u_i - \bar{u})^T A^{-1}(u_i - \bar{u})} \qquad \textbf{(3-1)}$$

Where u_i is a vector containing sample L-CV, L-skewness and L-kurtosis, denoted by $u_i = \begin{bmatrix} t^{(i)} & t_3^{(i)} & t_4^{(i)} \end{bmatrix}$, the procedure to calculate L-CV, L-skewness and L-kurtosis; vector $\bar{u}$ is the average of u_i, $\bar{u} = (\sum_{i=1}^{N} u_i)/N$; matrix A is sum of square of cross-product of error, $A = \sum_{i=1}^{N}(u_i - \bar{u})(u_i - \bar{u})^T$.

Hosking and Wallis [1993] suggested matrix A as the critical value for the discordancy statistic for regions containing any number of sites. Later it was found that the critical value of D_i for a region depends on its size.

Hosking and Wallis [1993] provides critical values of D_i for regions of various sizes, which are presented in the Table 3.8. In many instances the site discordancy may arise out of sampling variability. Therefore, the data at all sites with large values D_i should be carefully scrutinized before deciding whether the sites are discordant.

Table 3.8 Critical values for the discordancy statistic D_i [*Hosking and Wallis*, 1997]

N_R	Critical value of D_i
5	1.333
6	1.648
7	1.917
8	2.140
9	2.329
10	2.491
11	2.632
12	2.757
13	2.869
14	2.971
≥ 15	3.000

3.3.3.2 **Identification of Homogeneous Region by Heterogeneity Measure**

The heterogeneity measure tests whether the pool can be considered as a homogeneous region, which is calculated based on only L-CV for the whole region. The heterogeneity measure introduced by *Hosking and Wallis* [1993] is expressed by

$$\mathbf{H} = \frac{V - u_v}{\sigma_v} \qquad \textbf{(3-2)}$$

where the dispersion V is the weighted standard deviation of at-site sample L-CVs, denoted by

$$\mathbf{V} = \left\{ \frac{\sum_{i=1}^{N} n_i (t^{(i)} - t^R)^2}{\sum_{i=1}^{N} n_i} \right\}^{1/2} \qquad \textbf{(3-3)}$$

And $t^R = \sum_{i=1}^{N} n_i t^{(i)} / \sum_{i=1}^{N} n_i$, is the regional average L-CV; u_v and σ_v are the mean and standard deviation of V, which are obtained from a large number (N_{sim} = 500 or 1000) of Monte Carlo simulations. The procedure for each Monte Carlo simulation was described in *Liu* [2017]. *Hosking and Wallis* [1997] suggested a region to be considered as “acceptably homogeneous” if $H < 1$; “possibly heterogeneous” if $1 \leq H < 2$; “definitely heterogeneous” if $H \geq 2$.

3.3.3.3 **Probability Distribution Fitting by Method of L-moment**

The Generalized Extreme Value (GEV) distribution function is recommended by *Hosking and Wallis* [1997] in RFA method as it is a three-parameter distribution function that has the flexibility to fit to the pool. The Cumulative Distributed Function (CDF) of GEV distribution is expressed by

$$\mathbf{F(x)} = \begin{cases} \exp\left(-\left(1 - k \cdot \frac{x-\mu}{\sigma}\right)^{1/k}\right) & k \neq 0 \\ \exp\left(-exp\left(-\frac{x-\mu}{\sigma}\right)\right) & k = 0 \end{cases} \qquad \textbf{(3-4)}$$

where μ is the location parameter, σ is the scale parameter and k is the shape parameter.

The three parameters of GEV distribution were estimated by the method of L-moments. The estimators of the three parameters are expressed by

$$\hat{k} = 7.8590c + 2.9554c^2, c = \frac{2}{3+\tau_3} - \frac{log2}{log3} \qquad \textbf{(3-5)}$$

$$\hat{\sigma} = \frac{\lambda_2 \hat{k}}{\Gamma(1+\hat{k})(1-2^{-\hat{k}})} \qquad \textbf{(3-6)}$$

$$\hat{\boldsymbol{u}} = \boldsymbol{\lambda_1} + \hat{\boldsymbol{\sigma}}\{\Gamma(\mathbf{1} + \hat{\boldsymbol{k}}) - \mathbf{1}\}/\hat{\boldsymbol{k}} \qquad \textbf{(3-7)}$$

where τ_3 is the L-skewness; Γ denotes the gamma function $\Gamma(\mathrm{x}) = \int_0^\infty t^{x-1}e^{-t}dt$; λ_1 and λ_2 are the first and second L-moments.

3.3.3.4 Goodness of Fit Measure

To find the best distribution candidate or test performance of fitted distribution, the goodness-of-fit of the distribution candidate is measured by the standardized difference of L-kurtosis of the distribution candidate and the regional average L-kurtosis of observed data with the bias of measurements, i.e.

$$\boldsymbol{Z^{DIST}} = (\boldsymbol{\tau_4^{DIST}} - \boldsymbol{t_4^R} + \boldsymbol{B_4})/\boldsymbol{\sigma_4} \qquad \textbf{(3-8)}$$

where t_4^R is the regional average L-kurtosis value based on observed data; B_4 is the bias of L-kurtosis value t_4^R, denoted by $B_4 = N_{sim}^{-1}\sum_{m=1}^{N_{sim}}(t_4^{[m]} - t_4^R)$, $t_4^{[m]}$ is the regional average L-kurtosis from the mth Monte Carlo simulation with kappa distribution; N_{sim} is number of simulations. τ_4^{DIST} is the L-kurtosis value computed from a fitted distribution candidate, τ_4^{DIST} of GEV is denoted by $\tau_4^{GEV} = \{5(1-4^{-k}) - 10(1-3^{-k}) + 6(1-2^{-k})\}/(1-2^{-k})$;

$\sigma_4 = \left[(N_{sim}-1)^{-1}\left\{\sum_{m=1}^{N_{sim}}\left(t_4^{[m]} - t_4^R\right)^2 - N_{sim}B_4^2\right\}\right]^{1/2}$, is standard deviation of L-kurtosis values t_4^R derived from N_{sim} Monte Carlo simulations with kappa distribution to the regional average L-moment ratios 1, t^R, t_3^R and t_4^R. The procedure of Monte Carlo simulation of the goodness-of-fit measure is the same as heterogeneity measure.

The distribution is accepted if it satisfies the criteria of $|Z^{DIST}| \leq 1.64$, which corresponds to the acceptance of the hypothesized distribution at the confidence level of 90 %.

3.3.4 Derivation of Intensity-Duration-Frequency (IDF) Curves and Design Storms

3.3.4.1 Derivation of IDF Curves

The IDF curves represent the relationships among rainfall intensity, duration and return period of the event by empirical mathematical equations. The IDF curves are expressed by a generalized from a basic empirical equation as follow [*Koutsoyiannis et al.*, 1998]:

$$\mathbf{I} = \frac{\boldsymbol{a}}{(\boldsymbol{t_d}+\boldsymbol{b})^{\boldsymbol{c}}} \qquad \textbf{(3-9)}$$

where I is the rain intensity in mm/hour, t_d is the duration in minute, and a, b, and c are coefficient obtained by fitting to an IDF curve for a given return period (T) using least squares method. Talbot equation with c = 1 is applied in the study. More information on constructing IDF curves can be found in *Koutsoyiannis et al.* [1998].

3.3.4.2 Chicago Design Storm

Keifer and Chu [1957] developed the Chicago method which has since been extensively applied and used in the hydrology. Initially, its intended application was the sizing of sewers for design storm durations of three hours, although the method does not limit its applications to different purposes.

The hyetograph equation of Chicago design storm for specific return period (T) is expressed by

$$\mathbf{I(t)} = \begin{cases} \boldsymbol{a} \cdot \dfrac{\boldsymbol{b}+(\mathbf{1}-\boldsymbol{c})\frac{\boldsymbol{t_p}-\boldsymbol{t}}{\boldsymbol{r}}}{\left(\boldsymbol{b}+\frac{\boldsymbol{t_p}-\boldsymbol{t}}{\boldsymbol{r}}\right)^{\boldsymbol{c}+\mathbf{1}}} & \boldsymbol{for\ t \leq t_p} \\ \boldsymbol{a} \cdot \dfrac{\boldsymbol{b}+(\mathbf{1}-\boldsymbol{c})\frac{\boldsymbol{t_p}-\boldsymbol{t}}{\mathbf{1}-\boldsymbol{r}}}{\left(\boldsymbol{b}+\frac{\boldsymbol{t_p}-\boldsymbol{t}}{\boldsymbol{r}}\right)^{\boldsymbol{c}+\mathbf{1}}} & \boldsymbol{for\ t \geq t_p} \end{cases} \qquad \textbf{(3-10)}$$

where t_p is the time at which maximum rainfall intensity $I(t_p)$ occurs; r $(0 < r < 1)$ is a ratio of time to storm peak (t_p) to storm duration (t_d), which describes the asymmetry of hyetograph. In this study, a value of $r = 0.5$ is assumed, relating to an equal distribution of rainfall on either side of the peak.

The derived design storms using the above equations with the assumption of four hour storm durations and $r = 0.5$ (equal distribution on either side of peak rainfall) is presented in Figure 3.21

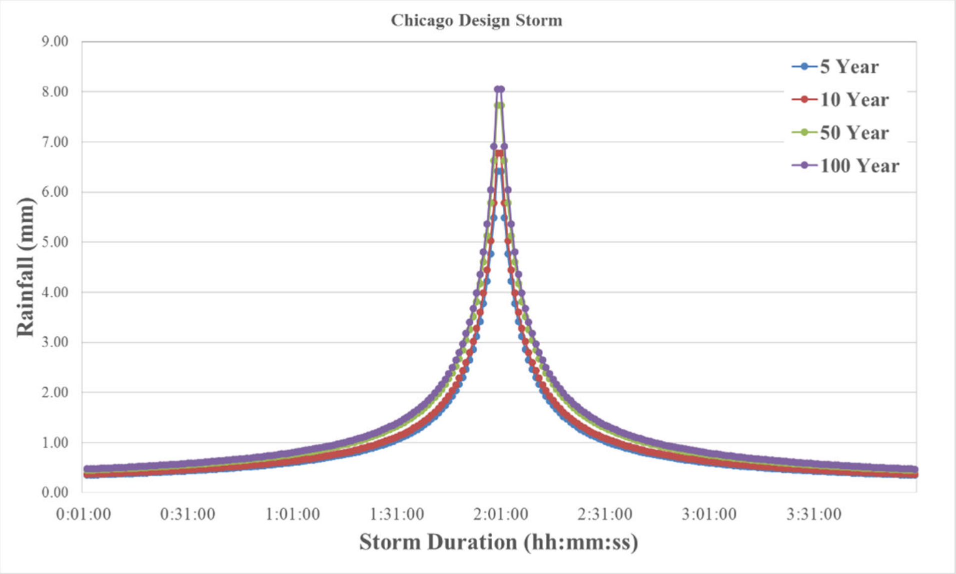


Figure 3.21 A sample of design storm time series for 5, 10, 50 and 100 year return periods and a total duration of 4 hours

3.3.4.3 Derivation of IDF Curves and Design Storms over Study Area

This study investigates the frequency and magnitudes of extreme rainfall events from WRF/ERAI for Greater Jakarta, Indonesia. The study area is located in the west side of Java Island. Figure 3.22 showed the catchment area with grid points from WRF/ERAI. The derived IDF curves and rainfall intensity can be found in Chapter 5.3.

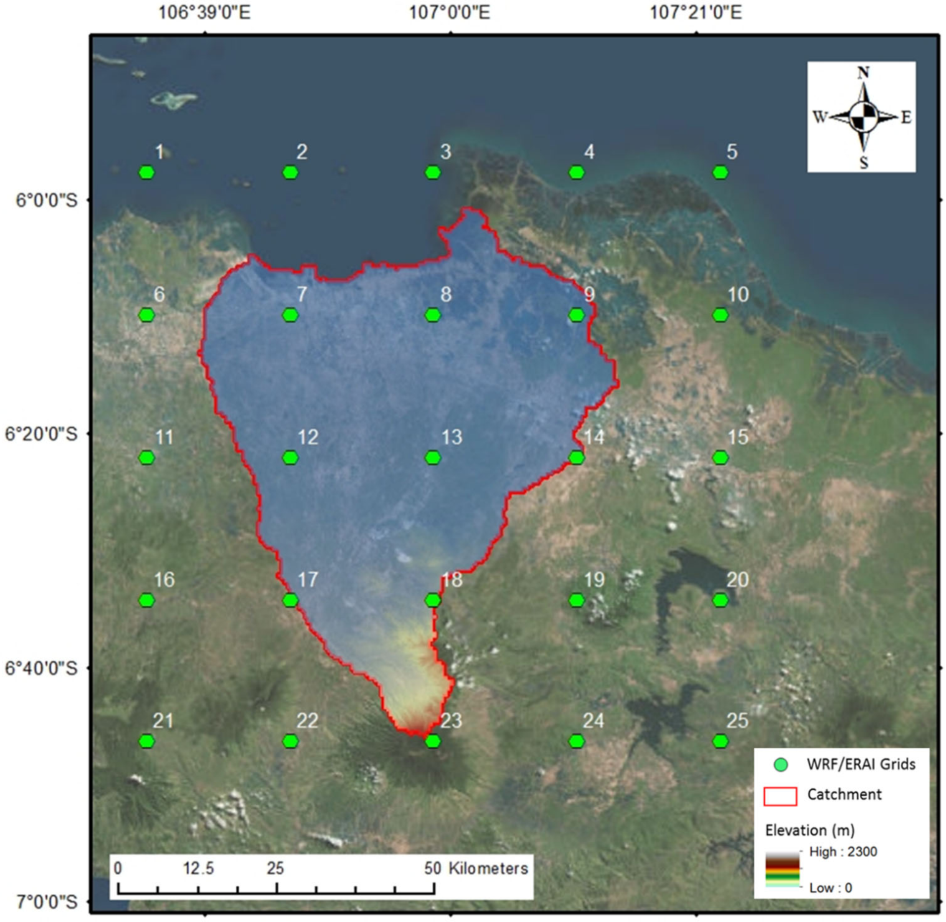


Figure 3.22 Map showing twenty-five (25) grid points of WRF/ERAI model; the color represents DEM

3.4 Numerical Flood Modelling

3.4.1 MIKE 21 Flow Model (MIKE 21FM)

MIKE 21 Flow Model FM (MIKE21FM) is a two-dimensional flow modelling system developed by DHI Water & Environment [*DHI*, 2017]. The 'FM' refers to the type of model grid with Flexible Mesh. The model has been developed for oceanographic environments and inland surface water simulations, e.g. overland flooding and reservoirs.

MIKE21FM is based on the numerical solution of the two-dimensional incompressible Reynolds averaged Navier-Stokes equation for shallow water flow condition. The equation adopts the following assumptions:

1. (1) Water is incompressible (hydrostatic pressure) and homogeneous
2. (2) Horizontal length scale is much greater than vertical length scale (thus the vertical velocity is considered negligible small compared to the horizontal velocity)

The two-dimensional shallow water equations derived from Navier-Stokes equation can be expressed as follows:

$$\frac{\partial \vec{u}}{\partial t} + \frac{\partial \vec{f}}{\partial x} + \frac{\partial \vec{g}}{\partial y} = \vec{h} \qquad \textbf{(3-11)}$$

where, x and y represent the two spatial dimensions and the four vectors $\vec{u}$, $\vec{f}$, $\vec{g}$, $\vec{h}$ are defined as follows:

$$\vec{u} = \begin{pmatrix} h \\ hu \\ hv \end{pmatrix}, \vec{f} = \begin{pmatrix} hu \\ g\frac{h^2}{2} + hu^2 \\ huv \end{pmatrix}, \vec{g} = \begin{pmatrix} hv \\ huv \\ g\frac{h^2}{2} + hv^2 \end{pmatrix}, \vec{h} = \begin{pmatrix} 0 \\ gh(S_{0x} - S_{fx}) \\ gh(S_{0y} - S_{fy}) \end{pmatrix} \qquad \textbf{(3-12)}$$

The depth-averaged velocities (m/s) in the x- and y- directions are marked by the u and v, respectively. S_{0x} and S_{0y} represent the bed slopes in the x- and y- directions and g is the acceleration due to gravity (m/s^2). S_f is the friction slope, which can be expressed in the x and y directions as follows, where h is the depth (m) and n is the Manning coefficient (s/m$^{1/3}$):

$$S_{fx} = -\frac{n^2 u\sqrt{u^2+v^2}}{h^{\frac{4}{3}}}, S_{fy} = -\frac{n^2 v\sqrt{u^2+v^2}}{h^{\frac{4}{3}}} \qquad \textbf{(3-13)}$$

The equation (3-11) is a simplification of the full 2D shallow water equations, which also include viscosity, Coriolis, wind shear stress, inflow volume and momentum.

In MIKE21FM the spatial discretization of the equation is carried out through the use of a cell-centred finite volume method. In the two-dimensional case, the elements can be arbitrarily shaped polygons like triangles and quadrilateral elements but here in this study only triangles are considered. The primitive variables representing the total water depth and the velocity components (h, u, and v) are recorded in the cell centres. The volume fluxes are then calculated perpendicular to the three faces of the elements as shown in Figure 3.23.

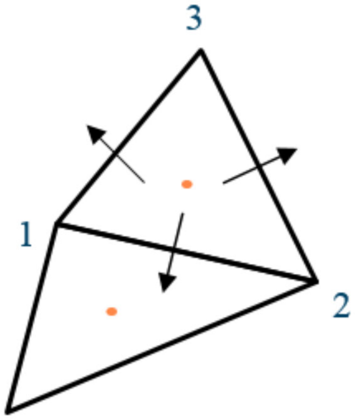


Figure 3.23 Volume fluxes perpendicular to element faces [*DHI*, 2017]

The time integration of the shallow water equations are performed using an explicit upwinding scheme. This scheme limits the time step to satisfy a specified Courant-Friedrich-Lewy (CFL) number less than 1; this is to avoid the model instability and miscalculations. The CFL number is defined as:

$$\mathbf{CFL} = \left(\sqrt{gh} + |u|\right)\frac{\Delta t}{\Delta x} + \left(\sqrt{gh} + |v|\right)\frac{\Delta t}{\Delta y} \qquad \textbf{(3-14)}$$

where g is the gravitational acceleration, h is the total water depth, u and v are the velocity components in the x- and y-directions respectively, Δt is the time step interval, and Δx and Δy are a characteristic length scale in the x- and y-directions. Δx and Δy are approximated by the minimum edge length for each element, i.e. the shortest element face. h, u and v are evaluated, as mentioned before, at the centre of the element.

The computation time is subject to the spatial factors included in the CFL number definition. In order to minimize the computation time, it is recommended to avoid too small elements and angles in the elements as this will generate short lengths on the edges. For example, when the lengths are short, the time step must be decreased significantly so that the required CFL number can be satisfied; this, at the same time, may significantly increase the computation time. Hence, it is important to find the balance of the cell size and the purpose of the flood modelling. DHI has developed the computational engine of MIKE21FM using Graphical Processing Units (GPU) to speed up the simulation time [DHI, 2014]. This study used dynamic time step up to 1 second and it limits the critical CFL number up to 0.8.

3.4.2 MIKE21FM Model Inputs

Various input data are required for MIKE21FM to produce the flood information. The input data can be divided into hydraulic and hydrological categories. The hydraulic parameters are, for examples, bathymetry, initial water surface level, and water level or discharge for boundary conditions which are related to land surface information. The hydrology parameters are, for examples, precipitation and evapotranspiration which are related to the climate conditions.

3.4.2.1 Domain and Bathymetry

In this study, the Improved SRTM DEM (henceforth referred as iSRTM DEM) is used as the topography of Java Island in the model as it shows much clearer land shape than the original SRTM DEM. The catchments were divided based on hydrological catchment delineation (USGS watershed boundary). 30 m resolution of grid was imported as scatter points in MIKE Zero Mesh Generator. From the scatter points, the mesh structures interpolated values to the final bathymetry meshes by applying a natural neighbour interpolation method. Different maximum mesh sizes are applied based on the landuses. For the city and industrial areas, a mesh size of 900 m^2, the maximum size, was applied which is relevant to the grid resolution of 30 m. For the other areas such as mountain and plain areas, a mesh size of 8,100 m^2 was used. A minimum angle of 26° was applied for this study. Figure 3.24 describes the different mesh sizes in the model.

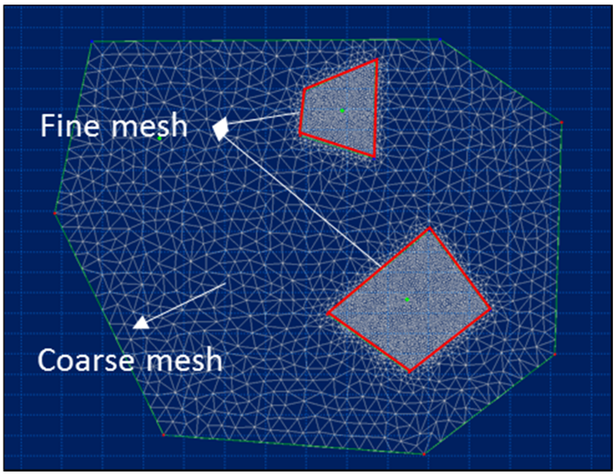


Figure 3.24 Sample of different mesh sizes in MIKE Zero Mesh Generator

3.4.2.2 Precipitation

As the precipitation is the main driving force in flood model, the rainfall data used is very important. As mentioned in the Section 3.2, the Intensity-Duration-Frequency (IDF) curves, derived from the rainfall data resulting from WRF/ERAI, are extracted to calculate the design storms using Chicago method. The Time of Concentration (TC) was calculated for each subcatchment to measure the response of a watershed to a given rain event. This is done so to give a conservative flood risk assessment. Passini method was applied to calculate the TC of large catchment. The Passini equation is defined as follows:

$$\mathbf{TC = 6.48(AL)^{0.333}} \boldsymbol{S^{-0.5}} \qquad \textbf{(3-15)}$$

where A is the basin area (km^2), L is a length of main channel (km), and S is the average slope of the basin (m/m).

Figure 3.21 showed the sample of the design storm using Chicago method with return periods of 5, 10, 50 and 100-year return periods and a total rainfall duration of 4 hours. The calculation of TC for the study area is discussed in Chapter 5.4.

3.4.2.3 Tide Level for Boundary Condition

As the study catchment ends at the coasts, tide level data are required and used as the downstream boundary condition. The tide level was generated and obtained from the Global Tide Model (GTM) which is developed by DTU Space (The National Space Institute (NSI) at the Technical University of Denmark). The GTM is available on a 0.125° x 0.125° resolution grid for the globe. The model is utilising the latest 17 years' multi-mission measurements from satellite altimetry for sea level residuals analysis [*Andersen et al.*, 1995]. MIKE Zero tool box can extract the time series for any period and any position on the globe [*DHI*, 2017].

This study extracted the one point of tide level. This was then applied to MIKE21FM; the respective highest tide level, for the year 2015, was set at 75 % of total rainfall duration as shown in Fig. 3.45. Note the aforementioned set-up between the peaks of rainfall and the tide was done so for most conservative scenario.

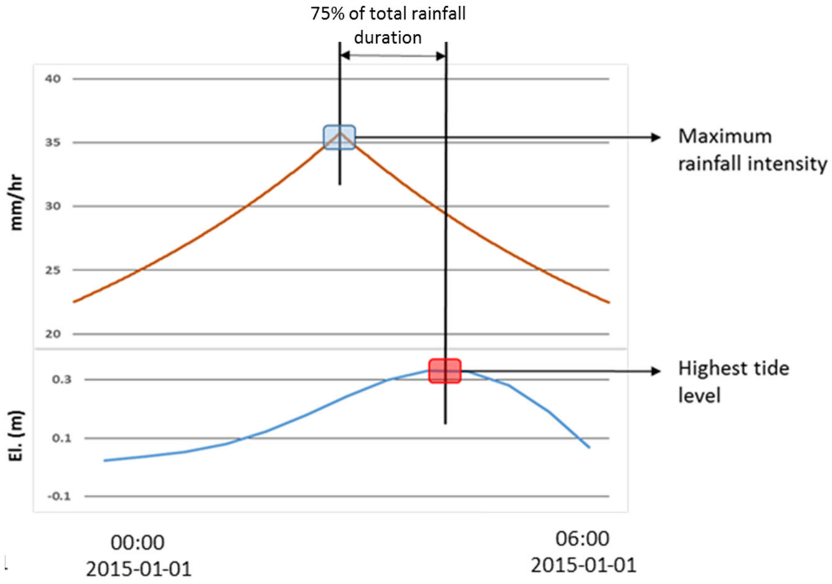


Figure 3.25 Tide level setup with rainfall intensity time series

3.5 Summary

This chapter presented the methodology and data for derivation of improved DEM, construction of IDF curves from downscaled climate model and setup of numerical flood model. The following is a summary:

- A methodology to derive improved DEM, with high accuracy, using publicly accessible multispectral imagery (SRTM DEM and Sentinel 2) and Artificial Neural Networks technique (ANN) was presented.
- Proxies for rainfall data, resulting from downscaled WRF/ERAI, was extracted, analyzed and compared with observed rainfall data. The performance of ERAI/WRF is reasonably well and used as rainfall proxies to derive IDF curves for ungauged catchment, Greater Jakarta.
- Derived improved SRTM DEM and IDF curves will then be used in MIKE21FM Flexible Mesh, in Chapters 4 and 5, to generate flood data and flood maps of various return periods.

4 Proof of Concepts

4.1 Overview

This chapter serves as the platform for proof of concept of the methodologies and approaches presented in Chapter 3. The concepts under consideration are:

- The improved DEMs: They are validated with high-resolution and high accuracy reference DEMs. A series of scenarios (urban/forested areas) are taken into consideration. A convincing/reasonable match between the improved DEM and the reference DEM will provide confidence for the usage of the developed DEM improvement scheme in other catchments where no observation DEM is available.

- The derived Intensity-Duration-Frequency curves: IDF curves derived from rainfall proxies, originated from the WRF/ERAI, are validated with IDF curves resulting from high quality observed data (in Singapore). This will later give a credibility of the usage of IDF curves, derived in the same manner, at other catchments where rainfall data are not available and/or rainfall records are not long enough to construct credible IDF curves. Note that high quality IDF curves are essential for flood analysis and flood mitigation measures.

4.2 Assessment of Derived DEM

This section evaluates the performance of derived DEMs using the method developed in this study as described in Section 3.2 of Chapter 3. DEMs in Nice (France) and Singapore are taken into consideration. Various scenarios of test cases are introduced in urban and forested areas: (1) ANN model trained and validated in Nice, France; (2) ANN model trained and validated in Singapore; (3) ANN model trained in Nice and validated in Singapore. The third case is essential as we need to ascertain the applicability of ANN model, trained in Nice, at other places where no high quality DEM, except satellite data, is available.

The improved DEMs will be compared with the high-resolution (1 m) and high accuracy (40 cm) DEM from Nice Côte d'Azur Metropolis (France) and Building and Construction Authority (BCA, Singapore) to evaluate their performances. Also, the drainage networks from different DEMs are compared.

Note that the resolutions of all remote sensing products are standardized at 10 m.

4.2.1 Scenario 1: ANN Model Trained and Validated in Nice (France)

Scenario 1 considers 2 separate cases: forested areas and urban areas.

For the case of forested areas, the ANN model is both trained and validated in two separate forested areas in Nice, France. The forested area used for ANN training has an area of 16 km^2 while the test area 8 km^2. Figure 4.1 showed the satellite image of the training (box with blue comb pattern) and test (box with red comb pattern) areas. The area mainly consists of vegetated mountains with valleys, steep slopes with elevations ranging from 200 to 1,200 m.

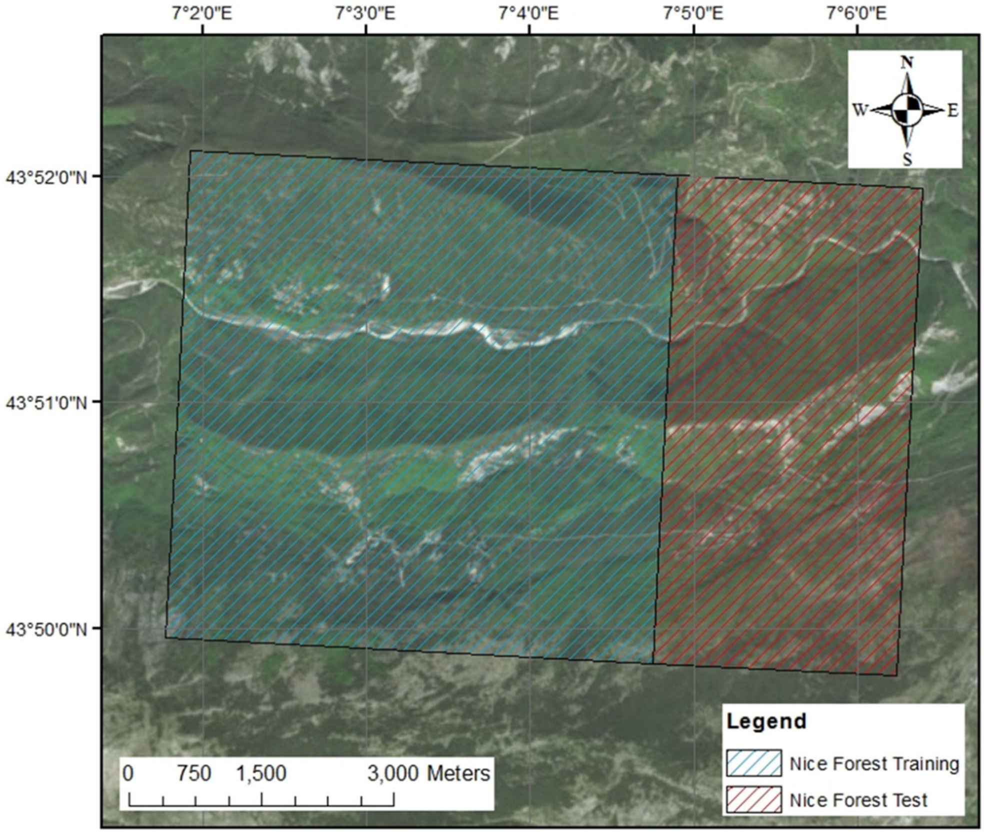
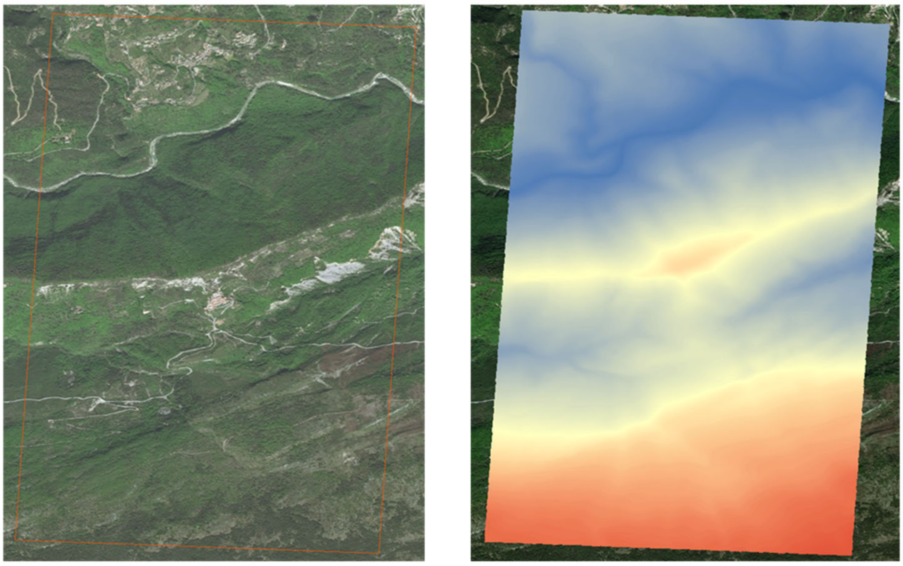


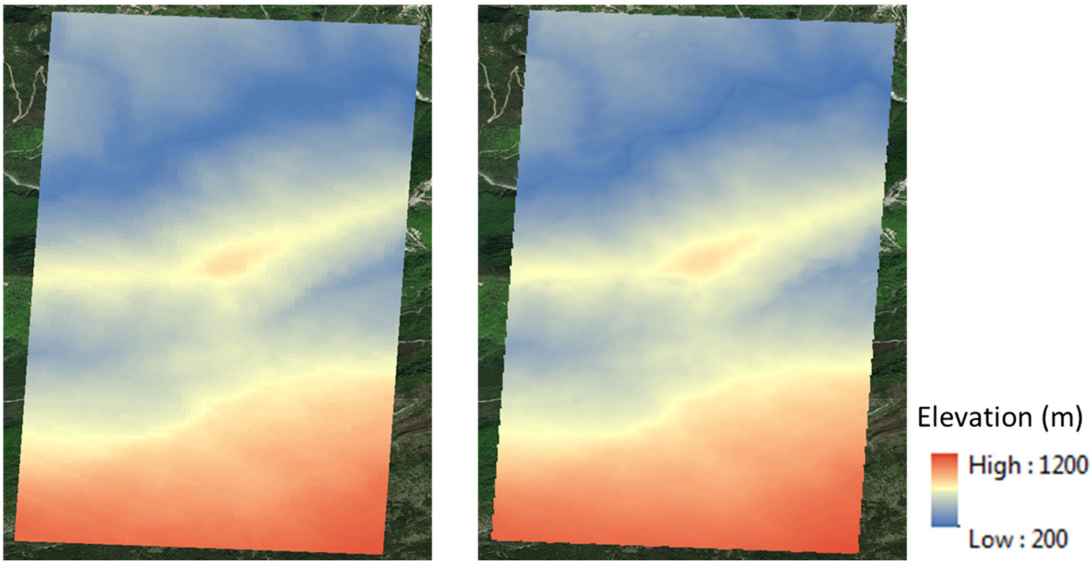
Figure 4.1 Training and test areas in Nice, France: forested areas

The ANN is trained in an area with the 1 m reference DEM data used in the target layer of ANN; the trained ANN is then applied to the test area. The performance is evaluated in the test area with 1 m reference DEM data as well. Figure 4.2 showed the comparison of elevation maps of various DEMs. Figure 4.2 (a) is a satellite image of the test area to depict the land shapes; Figure 4.2 (b) is the 1 m reference DEM; Figure 4.2 (c) is the DEM from the original SRTM DEM with 30 m resolution; and Figure 4.2 (d) is the DEM from improved SRTM DEM (iSRTM DEM) with 10 m resolution.



(a) Satellite Image

(b) 1 m reference DEM

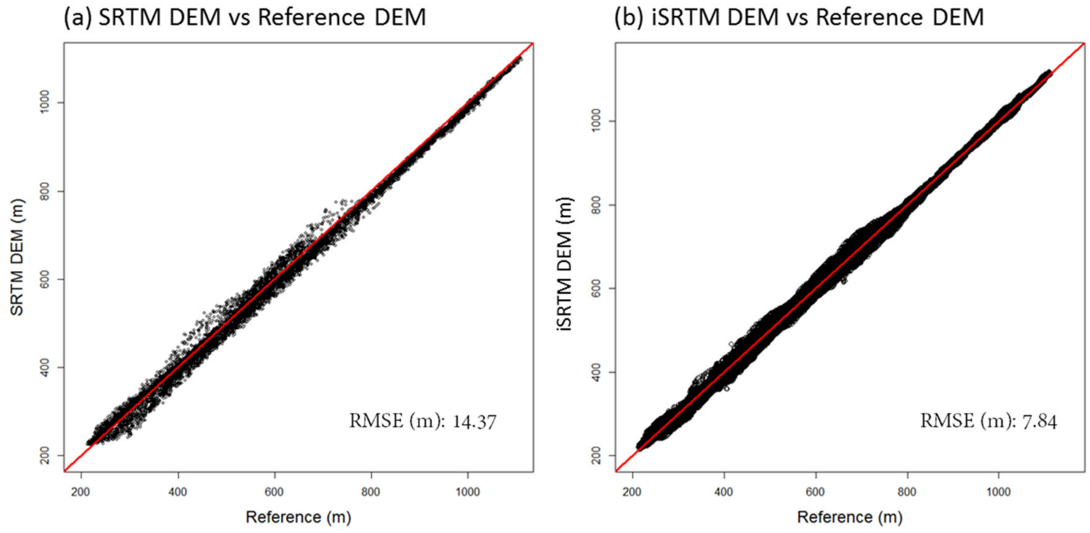


(c) SRTM DEM

(d) iSRTM DEM

Figure 4.2 Comparisons of (a) satellite image, (b) 1 m reference DEM, (c) SRTM DEM and (d) iSRTM DEM in forested area in Nice, France

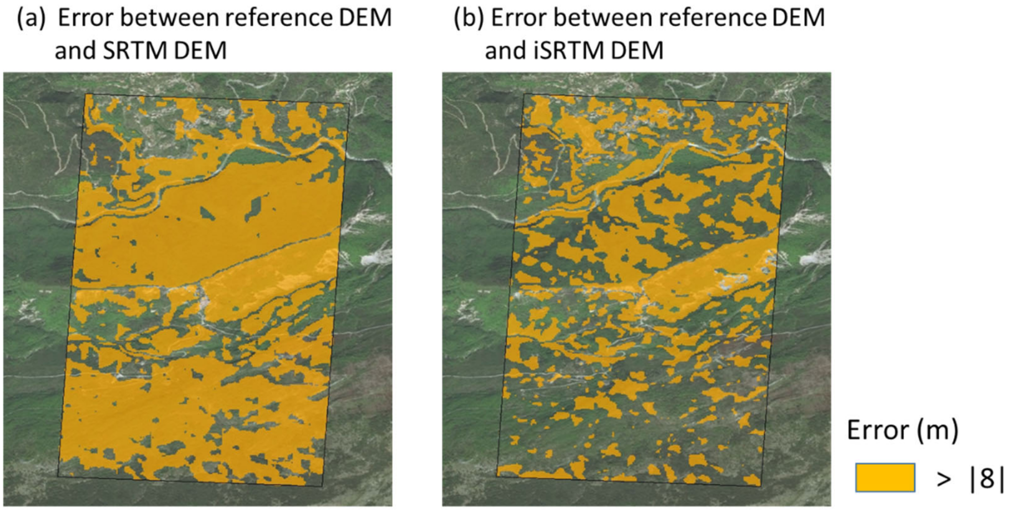
The 1m reference DEM showed most clear land shapes (i.e. particularly in the valley areas); iSRTM DEM showed improved clarity in the valley areas than its counterpart from the original SRTM DEM. The significant improvements can be captured in the scatter plots and their Root Mean Squared Error (RMSE) shown in Figure 4.3. The RMSE of iSRTM DEM is reduced to 7.84 m from SRTM DEM's RMSE 14.37 m, a 45.4 % reduction. Figure 4.4 showed the comparison of errors between reference DEM with SRTM DEM, and between the reference DEM and iSRTM DEM. The errors greater than the absolute error value of 8 m are reduced in iSRTM DEM.



(a) SRTM DEM vs Reference DEM

(b) iSRTM DEM vs Reference DEM

Figure 4.3 Scatter plots and RMSE comparisons between (a) SRTM DEM and (b) iSRTM DEM in forested area in Nice, France



(a) Error between reference DEM and SRTM DEM

(b) Error between reference DEM and iSRTM DEM

Figure 4.4 Absolute errors (a) between reference DEM and SRTM DEM; and (b) between reference DEM and iSRTM DEM in forested area in Nice, France

For the case of urban areas: the ANN model is both trained and validated in dense urban areas in Nice (France). Two versions of ANN training for dense urban areas are presented: (1) One single ANN training for the entire urban area (Version 1; henceforth referred as iSRTM_v1 DEM), and (2) Two separate ANN trainings, one for buildings only while the other for entire area without building heights (Version 2; henceforth referred as iSRTM_v2 DEM). Buildings are classified with building footprints from Open Street Map (OSM).

The training area has an area of 12.0 km^2 while the test area 5.2 km^2. Figure 4.5 showed the satellite image of the training (box with blue comb pattern) and test (box with red comb pattern) areas. The areas are mainly urbanized with buildings, mild slopes with elevation ranging from 0 m to 200 m.

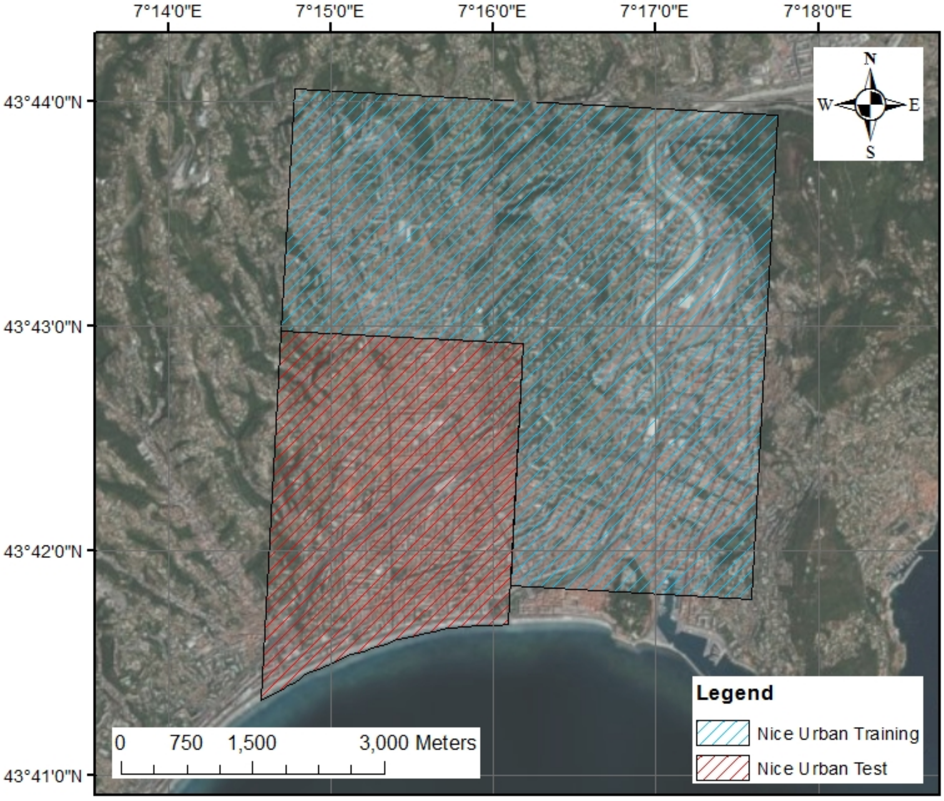


Figure 4.5 Training and test areas in Nice, France: dense urban areas

The ANN is trained in the training area with 1 m reference DEM data used in the target layer for iSRTM_v1 DEM. The iSRTM_v2 DEM is obtained from two ANN trainings, one with and one without building heights. The trained ANNs are then applied to the test area and the performances are evaluated using the reference DEM. Figure 4.6 showed the comparison of elevation maps of various DEMs.

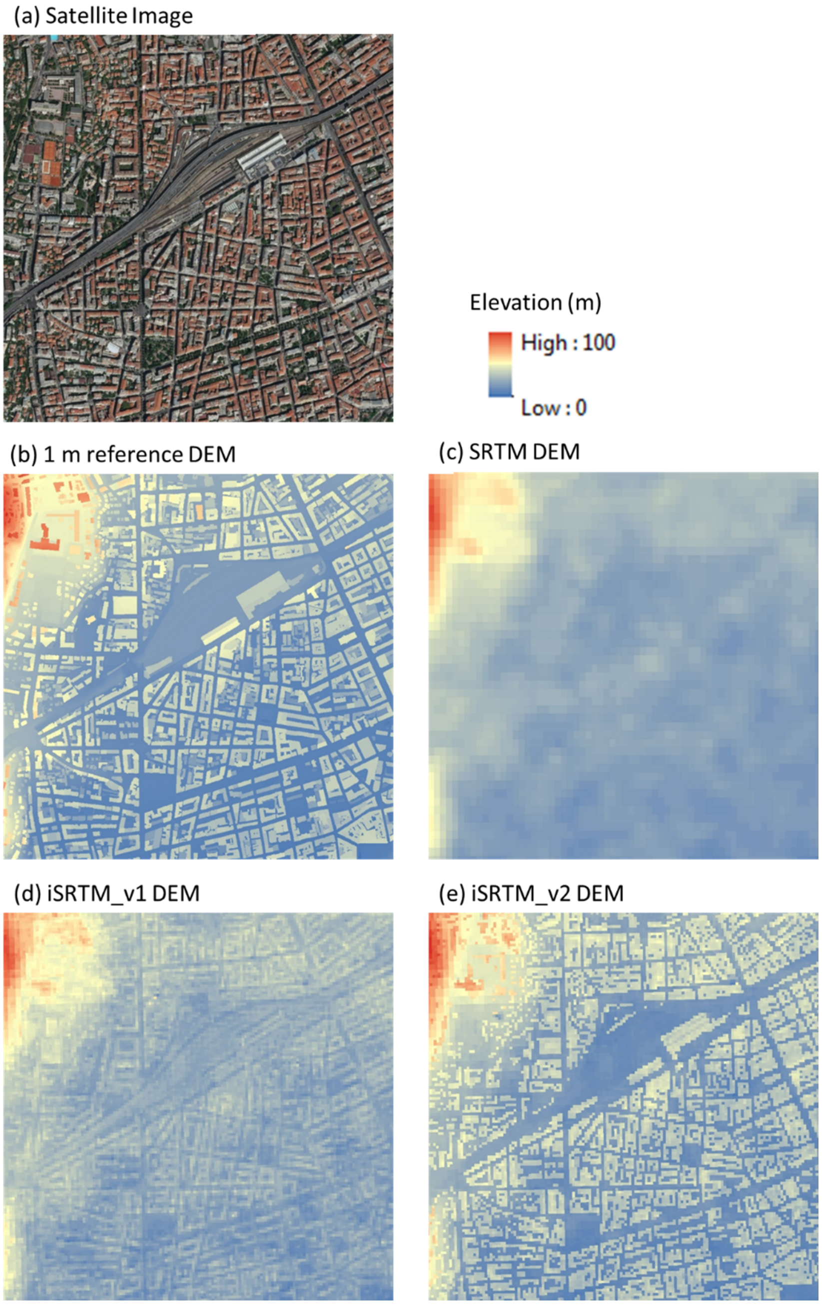


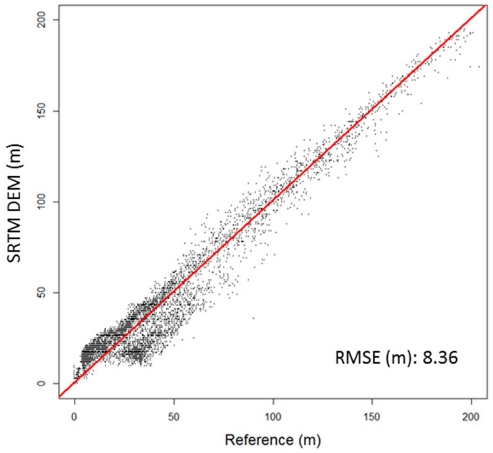
Figure 4.6 Comparisons of (a) satellite image, (b) 1 m reference DEM, (c) SRTM DEM, (d) iSRTM_v1 DEM, and iSRTM_v2 DEM in dense urban areas in Nice, France

Figure 4.6 (a) is a satellite image of test area depicting the land shapes; Figure 4.6 (b) is the area from 1 m reference DEM; Figure 4.6 (c) is the area from the original SRTM DEM with 30 m resolution; Figure 4.6 (d) is the area resulting from iSRTM_v1 DEM; Figure 4.6 I is the area resulting from iSRTM_v2 DEM. Both iSRTM DEM's resolutions are 10 m.

The reference DEM showed most clear land shapes (i.e. buildings and roads); iSRTM_v1 DEM and iSRTM_v2 DEM showed clearer land shapes in the developed areas than the original SRTM DEM; iSRTM_v2 DEM matches the reference DEM categorically much more.

Significant improvements are reflected in the scatter plots and the RMSE (Figure 4.7) as well. The RMSE of iSRTM_v1 DEM reduces to 7.82 m from 8.36 m (6.5 % reduction) while RMSE of iSRTM_v2 DEM reduces to 5.18 m (38.0 % reduction). Figure 4.8 showed the comparison of errors between reference DEM and SRTM DEM, between the reference DEM and iSRTM_v1 DEM, and between the reference DEM and iSRTM_v2 DEM. The errors greater than the absolute value of 5 m are reduced in both iSRTM DEMs. As the iSRTM_v2 DEM showed much better performance, this approach is selected for further validation and application. Table 4.1 showed the summary of the performance of iSRTM DEM in Nice for forested and urban areas.

(a) SRTM DEM vs Reference DEM



(b) iSRTM_v1 DEM vs Reference DEM

(c) iSRTM_v2 DEM vs Reference DEM

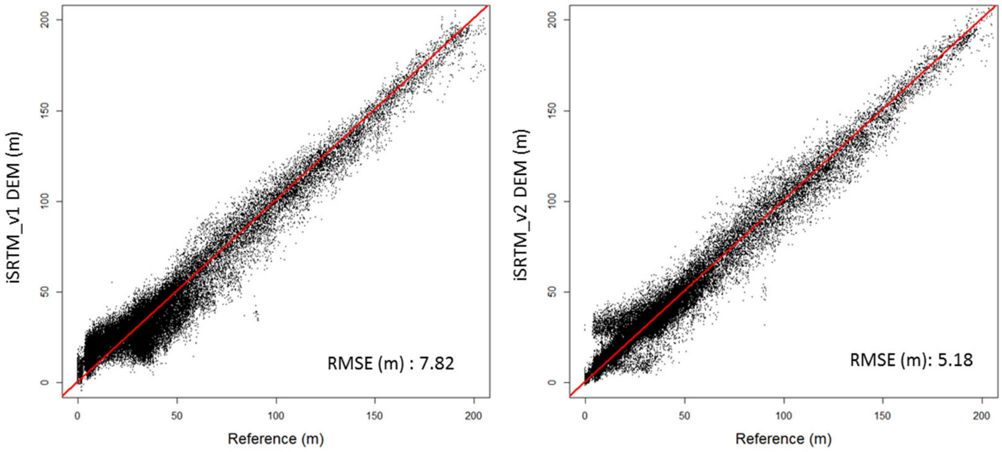


Figure 4.7 Scatter plots and RMSE comparisons between (a) SRTM DEM, (b) iSRTM_v1 DEM and iSRTM_v2 DEM in dense urban area in Nice, France

(a) Error between reference DEM and SRTM DEM

(b) Error between reference DEM and iSRTM_v1 DEM

(c) Error between reference DEM and iSRTM_v2 DEM

Figure 4.8 Absolute errors between (a) reference DEM and SRTM DEM; (b) reference DEM and iSRTM_v1 DEM; and (c) reference DEM and iSRTM_v2 DEM in dense urban area in Nice, France

Table 4.1 Summary table of the performances of iSRTM DEM in Nice, France

Description		**Training areas (km^2)**	**Test areas (km^2)**	**RMSE (m)**	
				SRTM DEM	**iSRTM DEM**
Forested area in Nice, France		16	8	14.37	7.84 (45.4 % ↓)
Dense urban area in Nice, France	Version 1	12	5.2	8.36	7.82 (6.5 % ↓)
	Version 2	12	5.2	8.36	5.18 (38.0 % ↓)

4.2.2 Scenario 2: ANN Model Trained and Validated in Singapore

Scenario 2 also considers 2 separate cases: forested areas and dense urban areas.

For the case of forested areas, the ANN model is both trained and validated in two separate forested areas in Singapore. The training area is an area of 1.84 km^2 while the test area 1.54 km^2. Figure 4.9 showed the satellite image of the training (box with blue comb pattern) and test (box with red comb pattern) areas. The area mainly consists of forest, mild slopes with elevations ranging from 0 to 80 m.

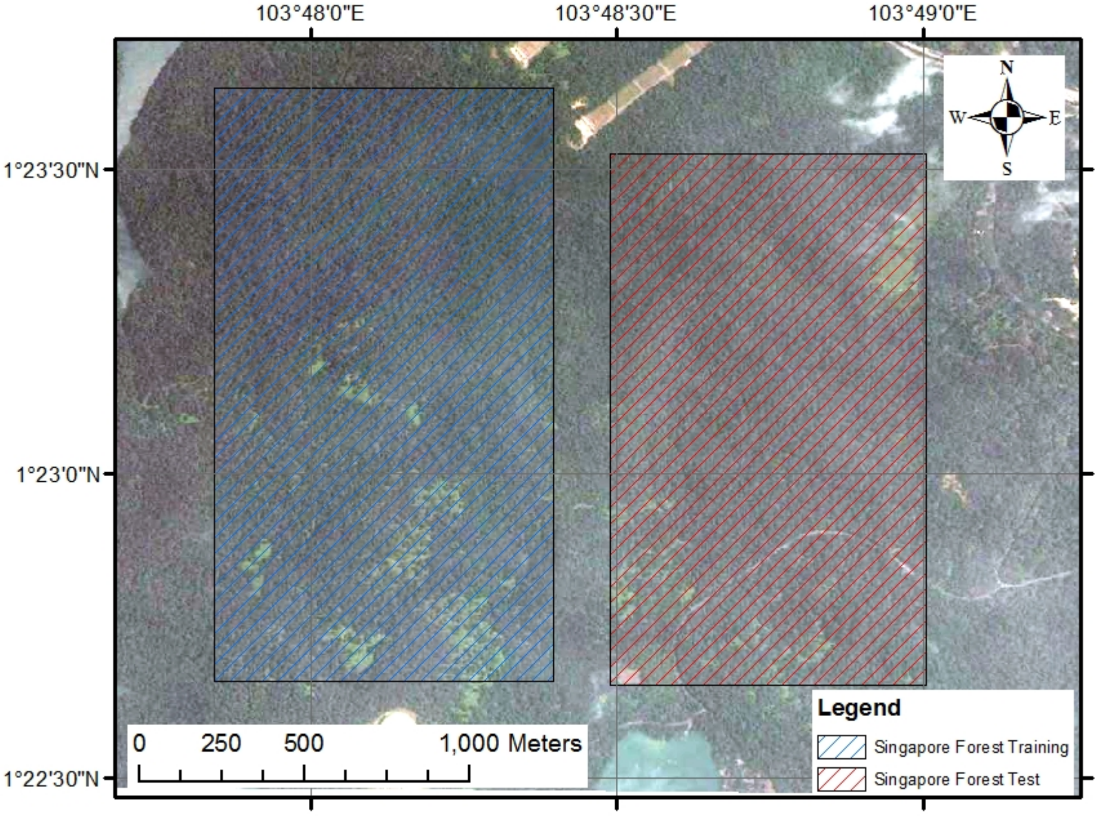


Figure 4.9 Training and test areas in Singapore: forested areas

The ANN is trained in an area with a reference DEM data of 1 m resolution used in the target layer of ANN. The trained ANN is then applied to the test area. The performance is evaluated in the test area with, again, a reference DEM data of 1 m resolution. Figure 4.10 showed the comparison of elevation maps of various DEMs. Figure 4.10 (a) is a satellite image of the test area depicting the land shapes; Figure 4.10 (b) is the DEM from 1 m reference DEM; Figure 4.10 (c) is the DEM from the original SRTM DEM with 30 m resolution; and Figure 4.10 (d) is the DEM from improved SRTM DEM (iSRTM DEM) with 10 m resolution. As the area is densely covered by trees, the land shape cannot be identified clearly on the satellite image. iSRTM DEM showed more similarity to reference DEM than the original SRTM DEM. Significant improvements are reflected in the plots of correlation and RMSE (Figure 4.11). The RMSE of iSRTM DEM reduces to 6.01 m from 14.13 m (57.4 % reduction). Figure 4.12 showed the comparison of absolute errors between the reference DEM and SRTM DEM, and between the reference DEM and iSRTM DEM. The errors greater than absolute value of 6 m are reduced in iSRTM DEM.

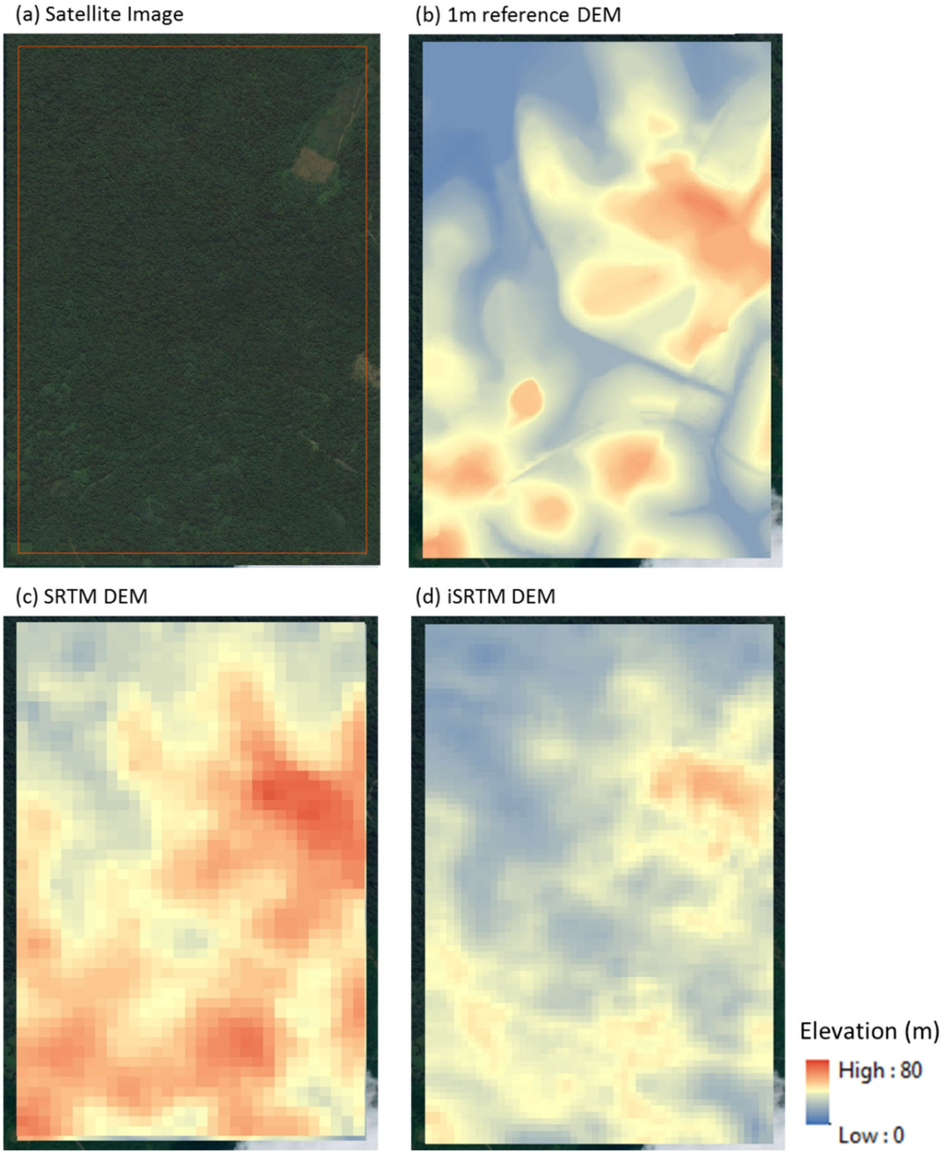


Figure 4.10 Comparisons of (a) satellite image, (b) 1 m reference DEM, (c) SRTM DEM and (d) iSRTM DEM in forested area in Singapore

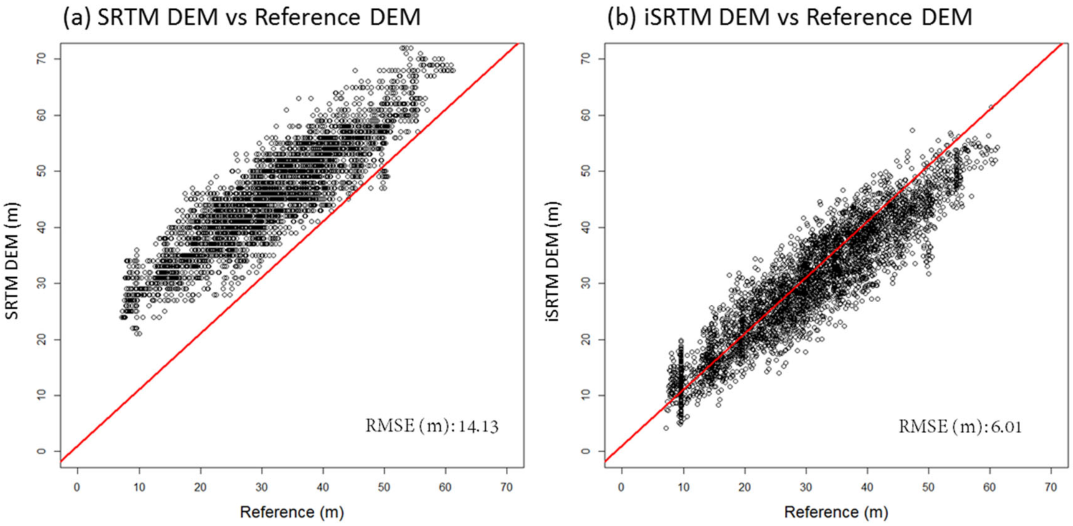


Figure 4.11 Scatter plots and RMSE comparisons between (a) SRTM DEM and (b) iSRTM DEM in forested area in Singapore

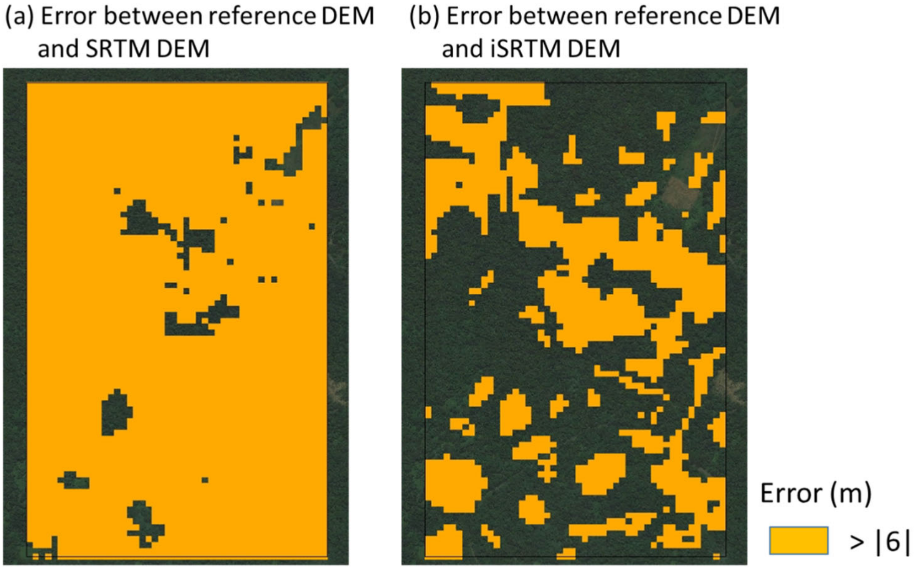


Figure 4.12 Absolute errors between (a) reference DEM and SRTM DEM, and (b) reference DEM and iSRTM DEM in forested area in Singapore

The performance of iSRTM DEM is now evaluated in a dense urban area in Singapore. The area of training area is 2.08 km^2 and test area is 1.62 km^2. Figure 4.13 showed the satellite image of the training (box with blue comb pattern) and test (box with red comb pattern) areas. The areas are very dense urban areas with high buildings (Bukit Timah area, Singapore). The elevation ranges from 0 to 100 m.

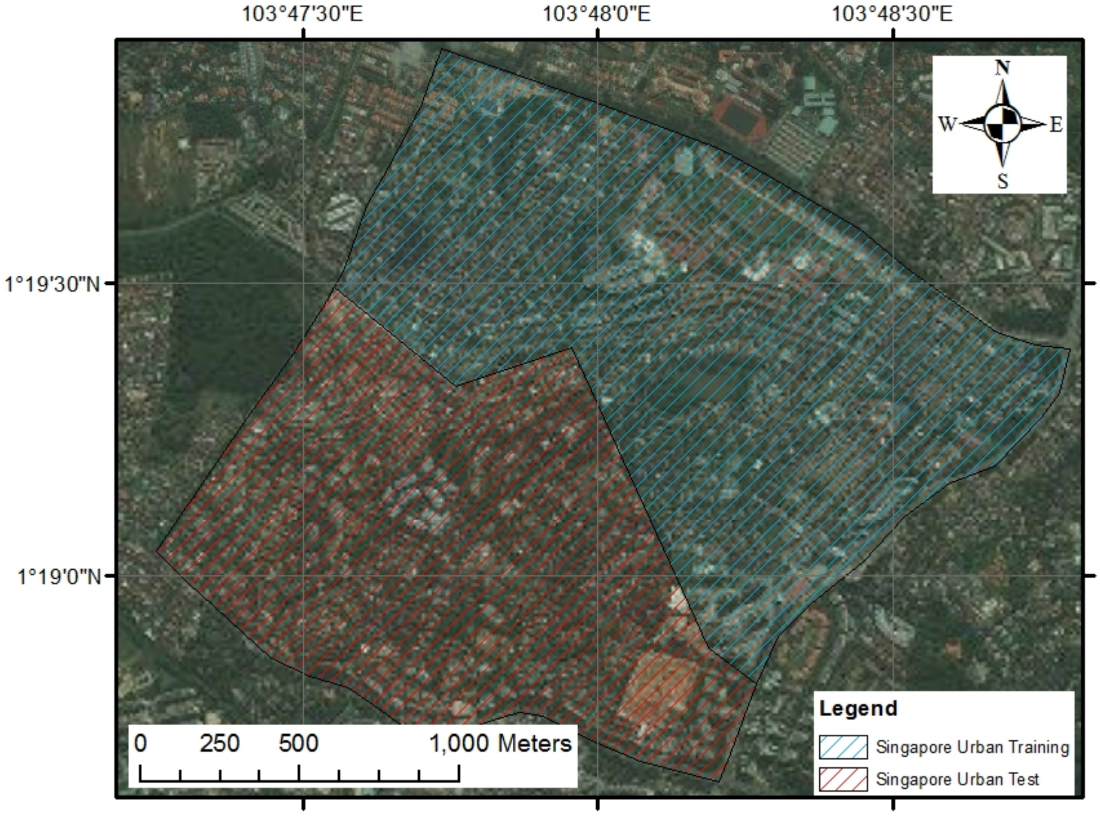


Figure 4.13 Training and test areas in Singapore: dense urban areas

The ANN is trained in the urban area with 1 m reference DEM data used in the target layer of ANN. The performance of the trained ANN is then validated in the test area (1 m reference DEM). Figure 4.14 showed the comparison of elevation maps of various DEMs. Figure 4.14 (a) is a satellite image of test area depicting the land shapes; Figure 4.14 (b) showed 1 m reference DEM; Figure 4.14 (c) showed the original SRTM DEM with 30 m resolution; and Figure 4.14 (d) showed the iSRTM DEM with 10 m resolution. The reference DEM showed most clear land shapes (i.e. buildings and roads); iSRTM DEM captures the characteristic of the buildings much better than the original SRTM DEM. Significant improvements are reflected in the scatter plots and RMSE (Figure 4.15) as well. The RMSE is reduced in iSRTM DEM from 10.70 to 6.60 m (a 38.3 % reduction). Figure 4.16 showed the comparison of absolute errors between reference DEM with SRTM DEM and between reference DEM and iSRTM DEM. The absolute errors greater than 7 m are reduced in iSRTM DEM. Table 4.2 showed the summary of the performance of iSRTM DEM in Singapore for forested and urban areas.

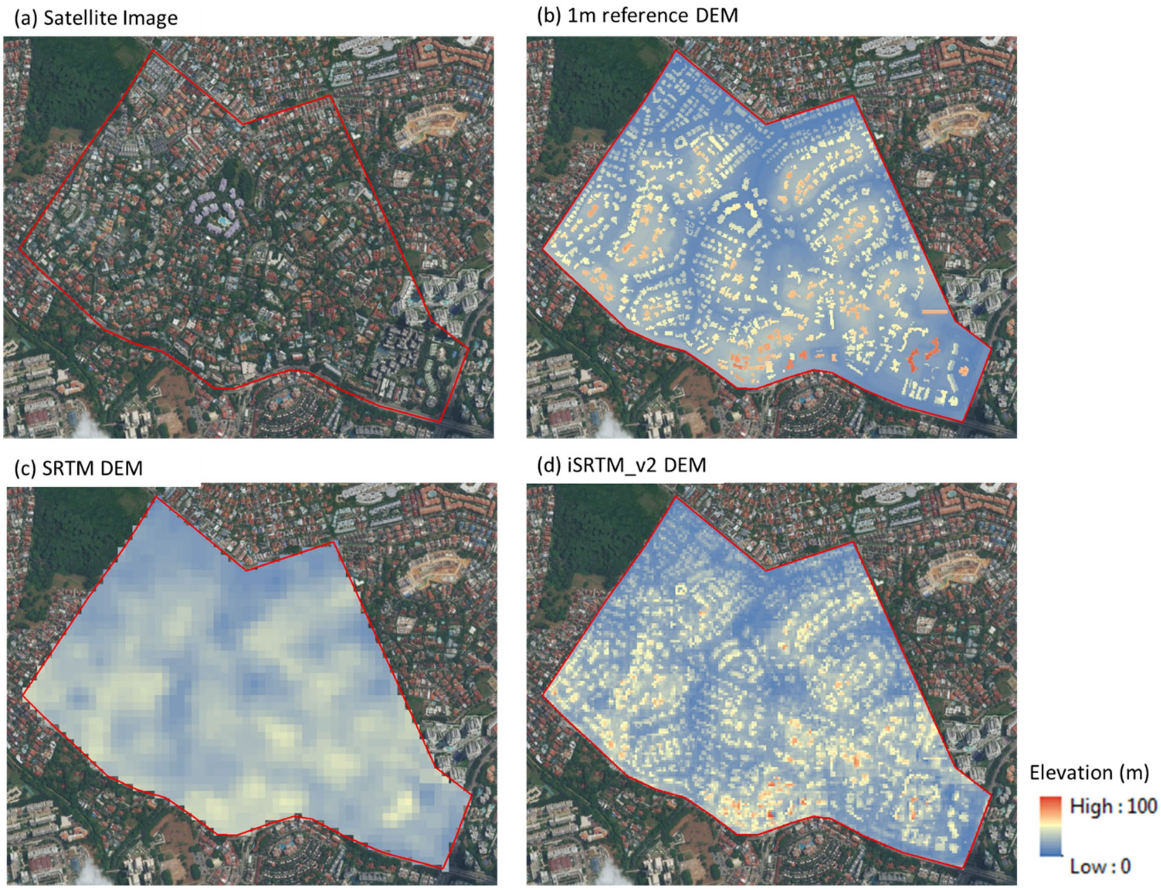


Figure 4.14 Comparisons of (a) satellite image, (b) 1 m reference DEM, (c) SRTM DEM and (d) improved SRTM_v2 DEM in dense urban area in Bukit Timah, Singapore

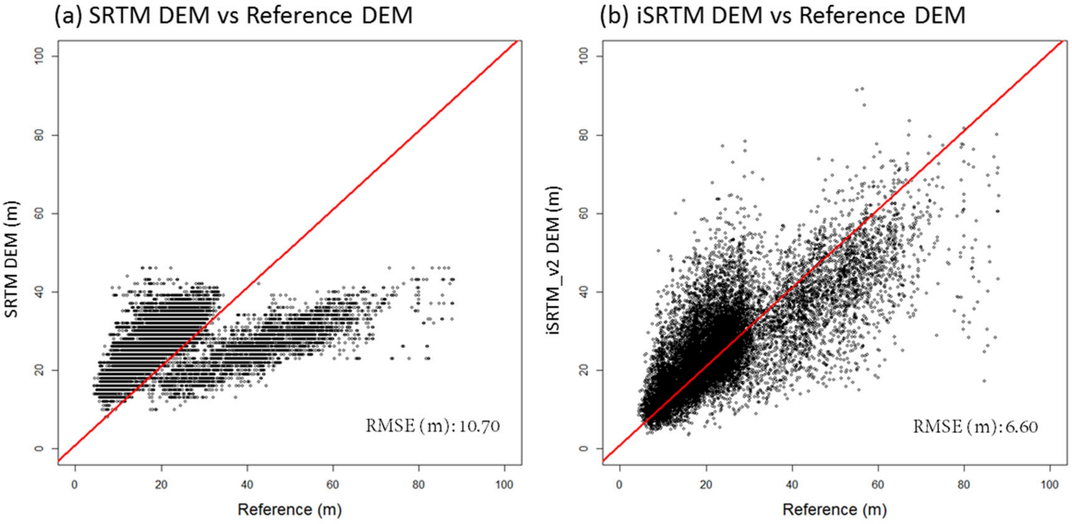


Figure 4.15 Scatter plots and RMSE comparisons between SRTM DEM and iSRTM_v2 DEM in Urban in Bukit Timah, Singapore

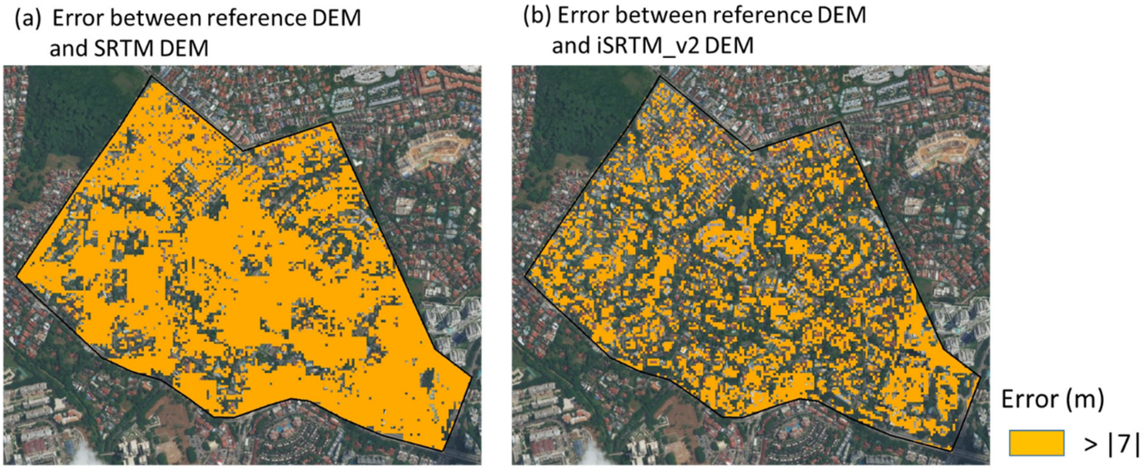


Figure 4.16 Absolute errors between (a) reference DEM and SRTM DEM, and (b) reference DEM and iSRTM_v2 DEM in dense urban area in Bukit Timah, Singapore

Table 4.2 Summary table of the performances of iSRTM DEM in Singapore

Description	**Training areas (km^2)**	**Test areas (km^2)**	**RMSE (m)**	
			SRTM DEM	**iSRTM_v2 DEM**
Forested area in Singapore	1.84	1.54	14.13	6.01 (57.5 % ↓)
Dense urban area in Singapore	2.08	1.62	10.70	6.60 (38.3 % ↓)

4.2.3 Scenario 3: ANN Model Trained in Nice (France) and Validated in Singapore

In Scenario 3 the interest is to investigate the quality of DEM generated by an ANN, trained in an area with mixed/variety of patterns (e.g. flat and open surfaces, dense urban areas with low and high rise buildings, mountains, and forests), when it is applied in a faraway mixed areas within the aforementioned patterns learnt. This is important to know any one has no other choice but to totally rely on SRTM DEM. Nice city in France is chosen for this experiment with training data sets consisting of dense urban and forested areas as presented in Section 4.2.1 (Scenario 1); the iSRTM_v2 DEM approach is used here. The trained ANNs are then validated in Singapore. The area of training is 42.2 km^2 while the test area of forested is 1.54 km^2 and test area of dense urban area is 1.62 km^2. These are same test areas presented in Section 4.2.2 (Scenario 2).

Figure 4.17 showed the comparison of elevation maps of various DEMs. Figure 4.17 (a) showed the 1 m reference DEM; Figure 4.17 (b) showed the original SRTM DEM with 30 m resolution; Figure 4.17 (c) showed the iSRTM DEM trained in Singapore with 10 m resolution; and Figure 4.17 (d) showed the iSRTM DEM trained in Nice with 10 m resolution. Although the ANN was trained in Nice and applied to Singapore's forested area, the iSRTM DEM matches closer the 1 m reference DEM than the original SRTM DEM. The improvements are reflected in the scatter plots and RMSE (Figure 4.18) as well. The RMSE of iSRTM DEM trained in Nice is reduced from 14.13 to 9.24 m (34.6 % reduction). The finding from this experiment also showed that iSRTM DEM trained in Singapore performs better than the iSRTM DEM trained in Nice as the pattern of terrain shapes and tree heights learnt in Singapore are obviously more similar to the test forested areas. Lesson in this case is the SRTM DEM can still be significantly improved with ANN trained in a faraway forested area where high quality ground truth data are available.

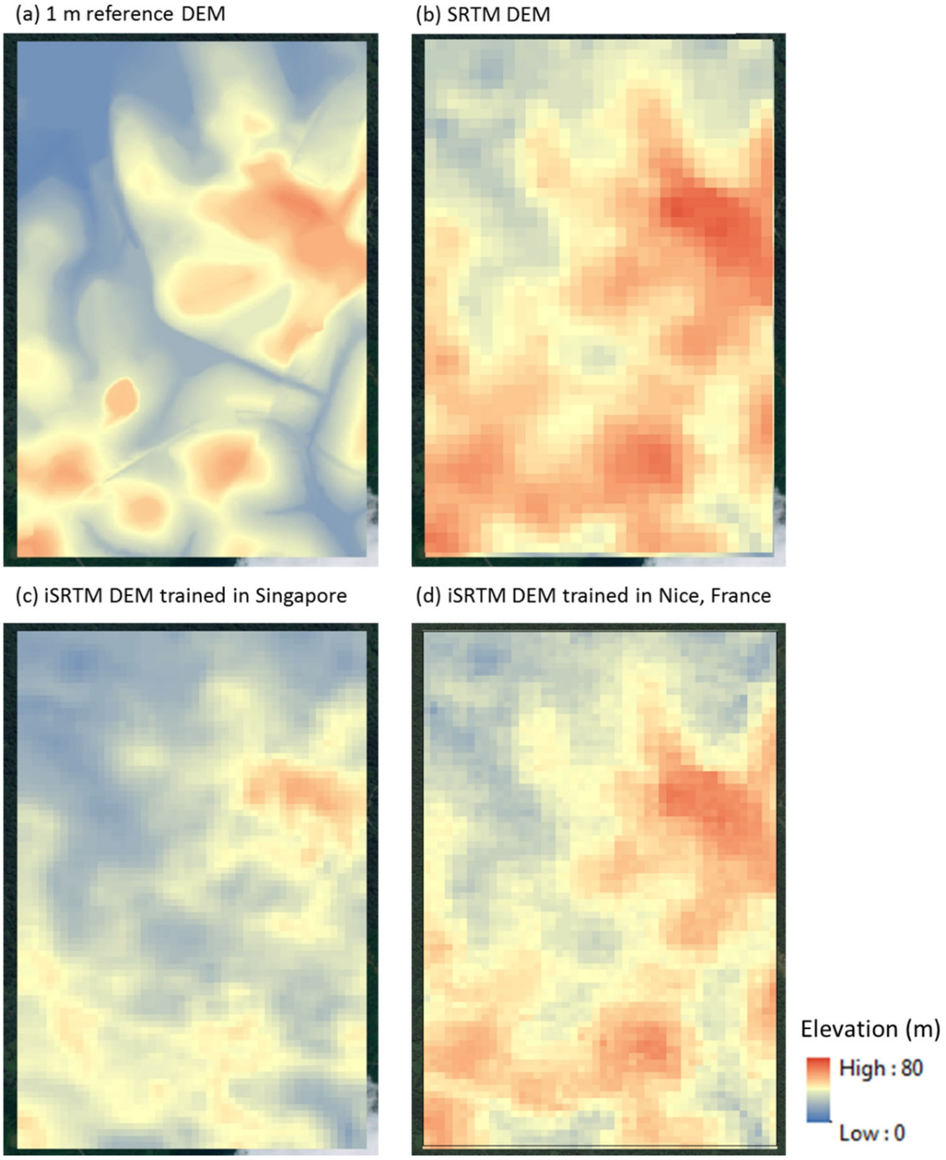
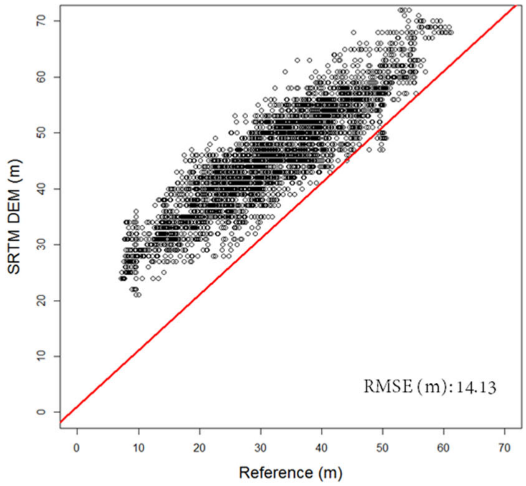


Figure 4.17 Performance comparisons of (a) 1 m reference DEM, (b) SRTM DEM, (c) iSRTM DEM trained in Singapore and (d) iSRTM DEM trained in Nice, France: forested area

(a) SRTM DEM vs Reference DEM



(b) iSRTM DEM trained in Singapore vs Reference DEM

(c) iSRTM DEM trained in Nice vs Reference DEM

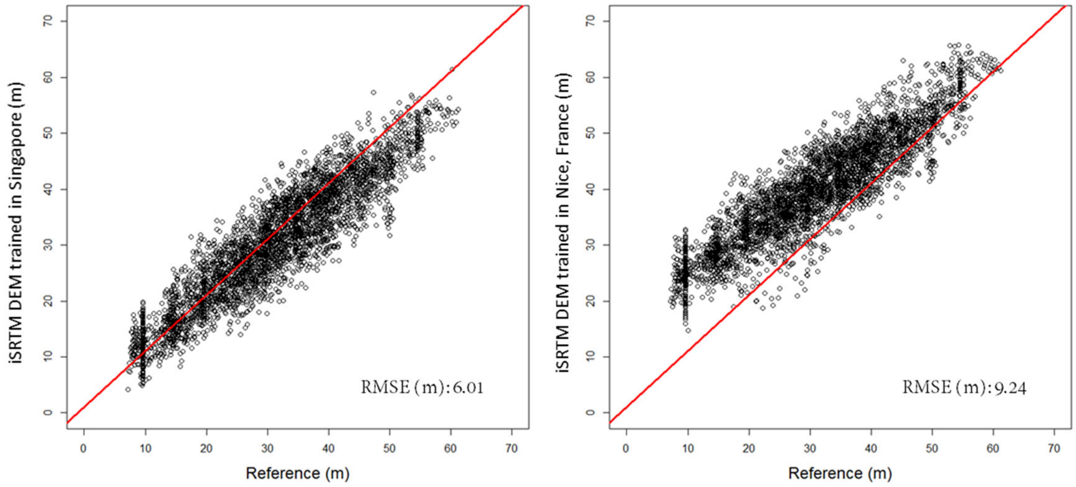


Figure 4.18 Scatter plots and RMSE comparison between (a) SRTM DEM, (b) iSRTM_v2 DEM trained in Singapore and (c) iSRTM_v2 DEM trained in Nice, France: forested area

A similar experiment is conducted for dense urban area. It uses the same ANN in Nice and test it in Singapore.

Figure 4.19 showed the comparison of elevation maps of various DEMs. Figure 4.19 (a) showed the 1 m reference DEM; Figure 4.19 (b) showed the original SRTM DEM with 30 m resolution; Figure 4.19 (c) showed the iSRTM DEM trained in Singapore with 10 m resolution; and Figure 4.19 (d) showed the iSRTM DEM trained in Nice with 10 m resolution. Although the ANN trained in Nice and applied to Singapore's dense urban area, the iSRTM DEM matches the 1 m reference DEM closer than the original SRTM DEM. The improvements are reflected in the scatter plots and RMSE (Figure 4.20) as well. The RMSE of iSRTM DEM trained in Nice is reduced from

10.70 to 6.93 m (35.2 % reduction). Similar to the forested area, for dense urban cities the performance of iSRTM DEM trained in Singapore performs better than iSRTM DEM trained in Nice. Again, lesson learnt is: the SRTM DEM can still be significantly improved with ANN trained in a faraway dense urban area where high quality ground truth data are available. Table 4.3 showed the summary of the performance of iSRTM DEM trained in Nice and validated in Singapore for forested and urban areas.

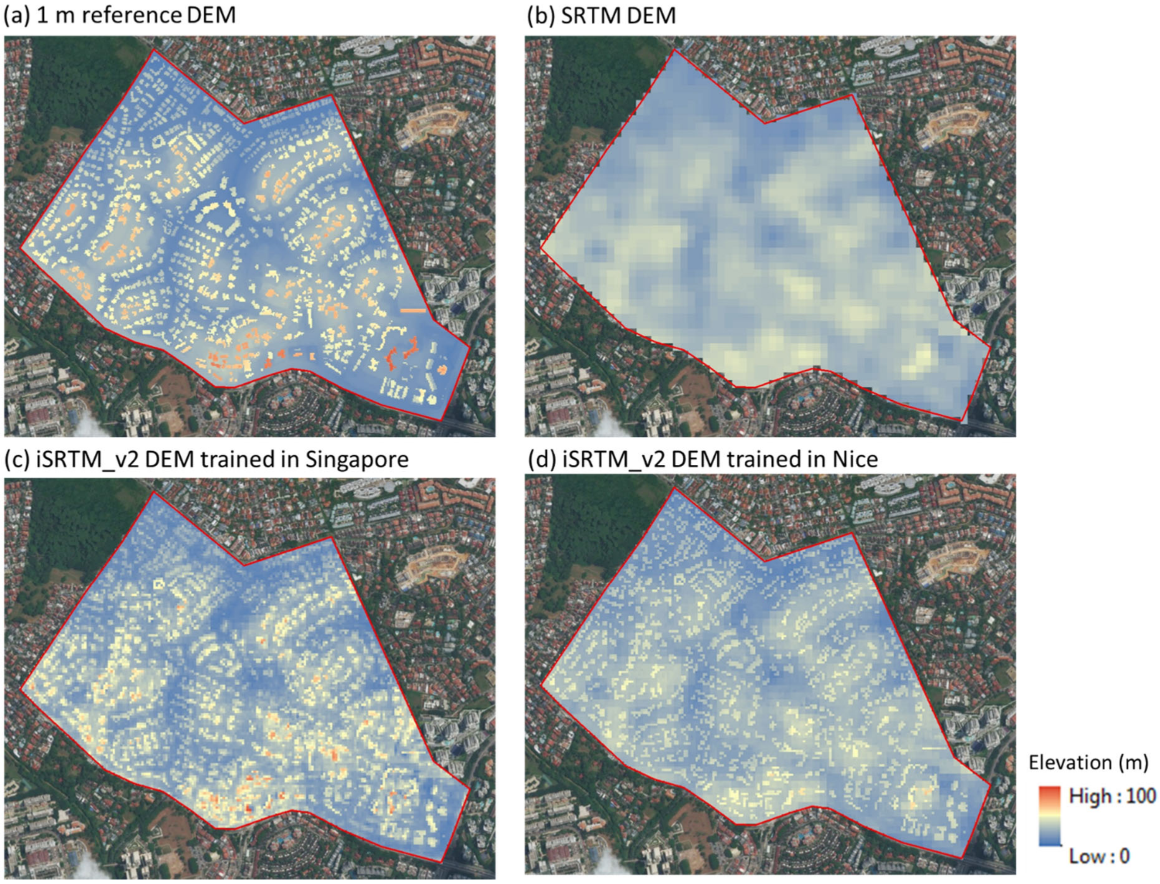
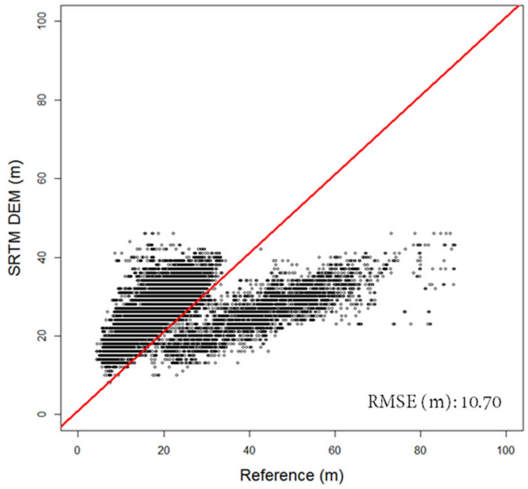
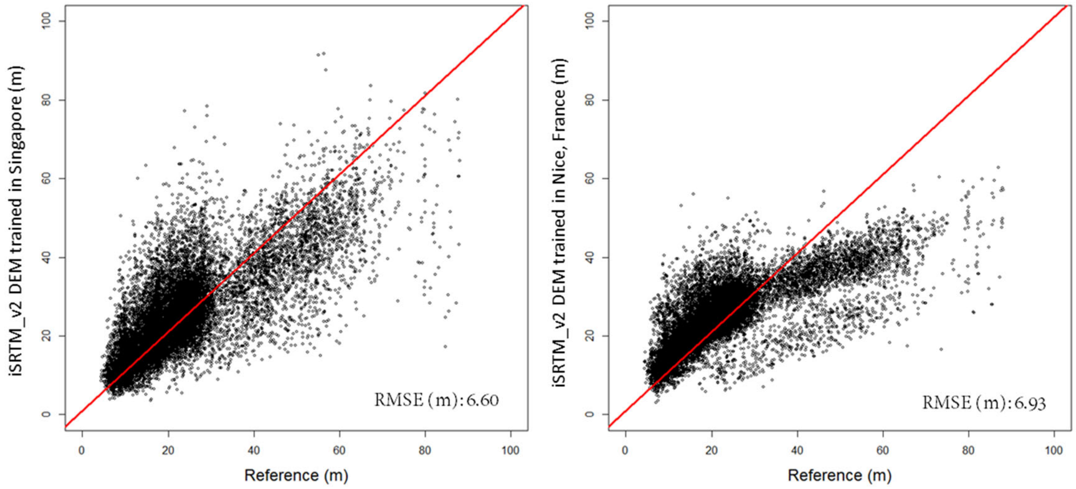


Figure 4.19 Performance comparisons of (a) 1 m reference DEM, (b) SRTM DEM, (c) iSRTM_v2 DEM trained in Singapore and (d) iSRTM_v2 DEM trained in Nice, France: dense urban area


(a) SRTM DEM vs Reference DEM


(b) iSRTM_v2 DEM trained in Singapore vs Reference DEM

(c) iSRTM_v2 DEM trained in Nice vs Reference DEM

Figure 4.20 Scatter plots and RMSE comparison between SRTM DEM, iSRTM DEM trained from Singapore and iSRTM DEM trained from Nice, France: dense urban area

Table 4.3 Summary table of performances of iSRTM DEMs trained in Nice, France and trained in Singapore and applied to test areas in Singapore

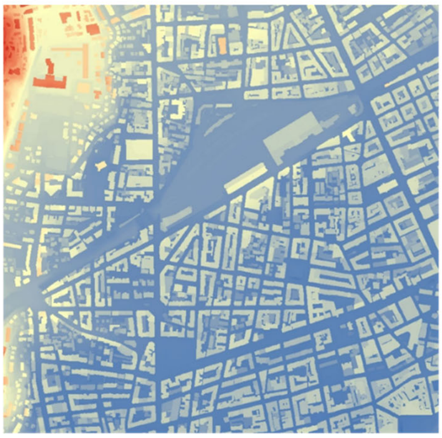
Description	RMSE (m)		
	SRTM DEM	iSRTM DEM	
		Trained in Singapore	Trained in Nice
Forested area in Singapore	14.13	6.01 (57.5 % ↓)	9.24 (34.6 % ↓)
Dense urban area in Singapore	10.70	6.60 (38.3 % ↓)	6.93 (35.2 % ↓)

With lessons learnt from Scenario 3, in Chapter 5 SRTM DEM of Greater Jakarta, Indonesia, where we do not have high quality resolution DEM, the iSRTM_v2 DEM trained in Nice is applied

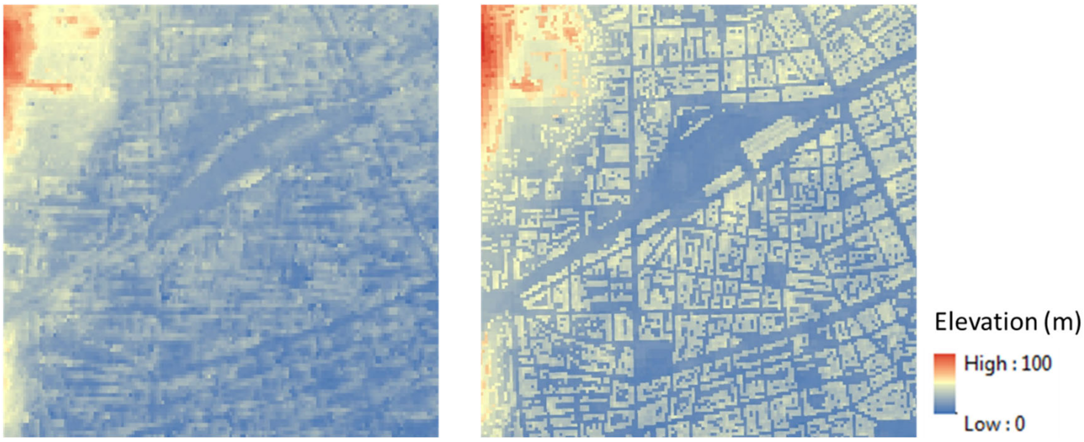
4.2.4 DEM Comparison between TanDEM-X DEM and iSRTM DEM

In this Section, the improved SRTM DEM is compared with the German Aerospace Center's DEM, TanDEM-X DEM (DLR's DEM), mentioned in Section 3.2.1.2 of Chapter 3. TanDEM-X DEM has 12 m resolution at the cost of € 30 per km^2. To note, particularly for dense urban area, is: even with 12 m resolution in TanDEM-X DEM, many tiles, in TanDEM-X DEM, each captured street and the neighboring tall buildings in one tile and an averaged elevation value for that tile is given. Averaging tall building with street elevations will surely result in significant discrepancy with the actual street elevation; this is quite alarming as an accurate street elevation is of great importance for flood study. The dense urban areas of Scenarios 1 and 2 are used below for comparisons.

Figure 4.21 showed the DEM comparison between TanDEM-X DEM and iSRTM_v2 DEM in dense urban area of Scenario 1 (Nice, France). iSRTM_v2 DEM showed clearer land shapes (i.e. buildings and roads) than the TanDEM-X DEM. The better performance is reflected in the scatter plots and the RMSE as shown in Figure 4.22. The RMSE of iSRTM_v2 DEM (5.18 m) is lower than that of TanDEM-X DEM (7.24 m) and it performs 28.5 % better than TanDEM-X DEM.



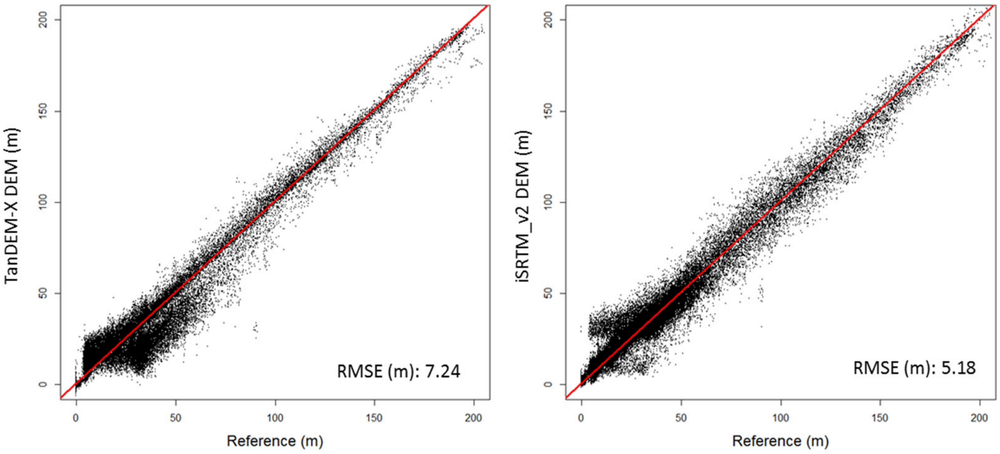
(a) 1 m reference DEM



(b) TanDEM-X DEM

(c) iSRTM_v2 DEM

Figure 4.21 DEM comparisons of (a) 1 m reference DEM, (b) TanDEM-X DEM and (c) iSRTM_v2 DEM in Nice, France



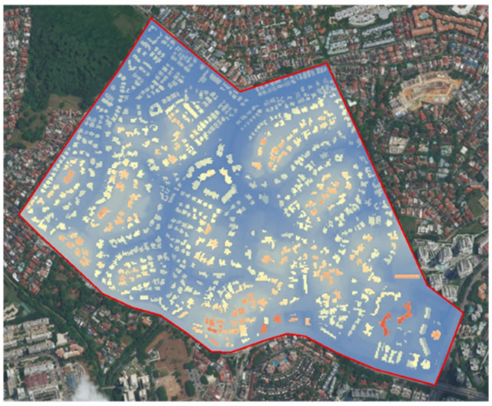
(a) TanDEM-X DEM vs Reference DEM

(b) iSRTM_v2 DEM vs Reference DEM

Figure 4.22 Scatter Plots and RMSE comparisons between (a) TanDEM-X DEM and (b) iSRTM_v2 DEM trained in Nice, France

Figure 4.23 showed the DEM comparison between TanDEM-X DEM and iSRTM_v2 DEM in dense urban area of Scenario 2 (Singapore). iSRTM_v2 DEM again showed clearer land shapes (i.e. buildings, roads and streams) than the TanDEM-X DEM. Significant performance is reflected also in the scatter plots and the RMSE as shown in Figure 4.24. The RMSE of iSRTM_v2 DEM (6.97 m) is lower than TanDEM-X DEM (8.69 m) and it performs 19.8 % better than TanDEM-X DEM.

(a) 1m reference DEM



(b) TanDEM-X DEM

(c) iSRTM_v2 DEM

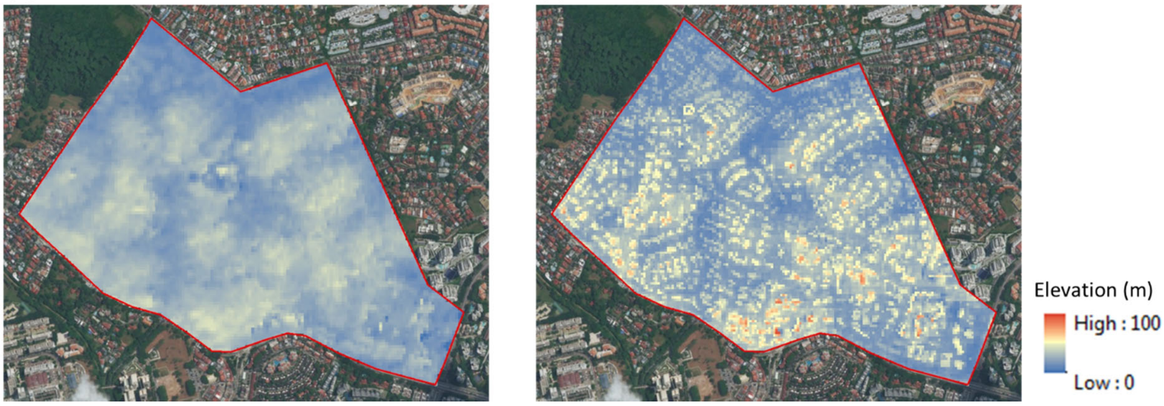


Figure 4.23 DEM comparisons of (a) 1 m reference DEM, (b) TanDEM-X DEM and (c) iSRTM_v2 DEM in Singapore

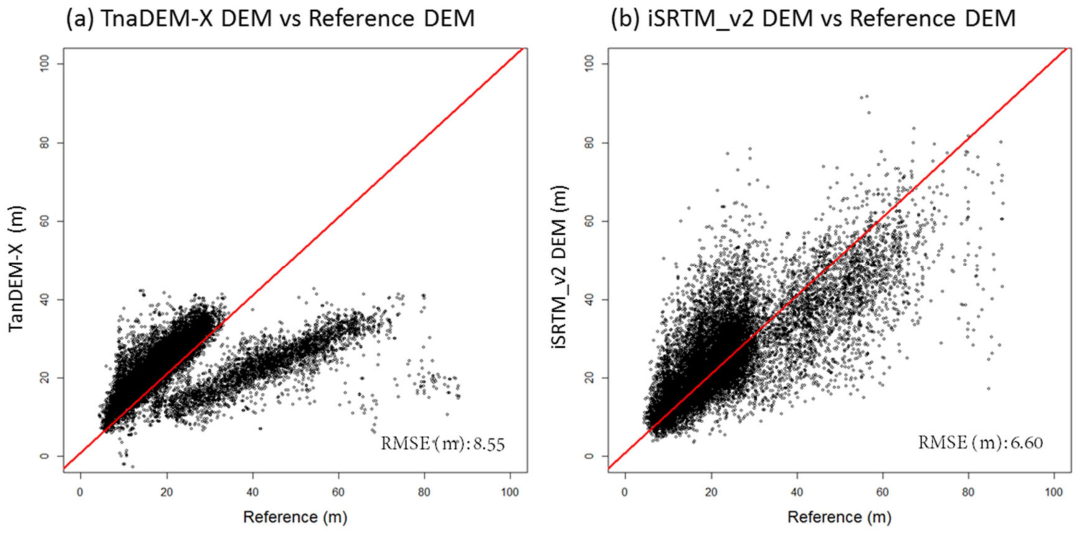


Figure 4.24 Scatter plots and RMSE comparisons between (a) TanDEM-X DEM and (b) iSRTM_v2 DEM in Singapore

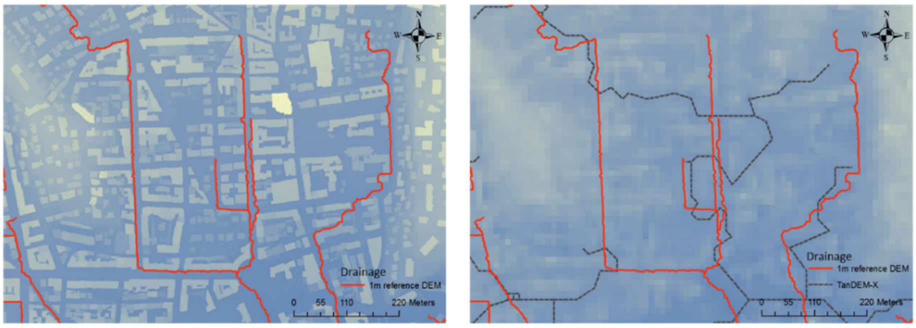
4.2.5 Drainage Networks Derivation from the DEMs

In flood simulation, the drainage networks are of great importance. This Section presents comparisons of drainage networks derived from the 1 m reference DEM, TanDEM-X DEM, SRTM DEM and iSRTM_v2 DEM; this is another matrix to evaluate the performance of various DEMs. The derivation of drainage networks from DEMs is basically determined by the flow direction, flow accumulation grid, variation of elevation and spatial resolution may influence the integrity of the resulting networks [*Thomas and Prasannakumar*, 2015]. The areas of Scenarios 1 and 2 in dense urban cites are considered.

Figure 4.25 (a) is the drainage network derived from the 1m reference DEM in Nice and this is used as the benchmark for comparison with networks resulting from other DEMs; Figure 4.25 (b) is the drainage network resulting from TanDEM-X DEM with 12 m resolution; Figure 4.25 (c) is the one from the original SRTM DEM with 30 m resolution; Figure 4.25 (d) is the one resulting from iSRTM_v2 DEM with 10 m resolution. In general, the streamlines are in a qualititive match but some parts in middle - left area, none follow the drainage line from that of the 1 m reference DEM. The drainage network derived from iSRTM_v2 DEM showed closest and good agreement with that of the 1m reference DEM drainage.

Figure 4.26 (a) is the drainage network derived from 1 m reference DEM in Singapore and this is the bench mark to be compared with other networks; Figure 4.26 (b) is the drainage network from TanDEM-X DEM with 12 m resolution; Figure 4.26 (c) is the one from the original SRTM DEM with 30 m resolution; Figure 4.26 (d) is the one resulting from iSRTM_v2 DEM with 10 m resolution. In general, the streamlines are in a good correspondence with drainage lines from 1 m reference DEM. The drainage network derived from and iSRTM_v2 DEM showed the best agreement with the one from 1 m reference DEM.

(a) 1m reference DEM (b) TanDEM-X DEM



(c) SRTM DEM (d) iSRTM_v2 DEM

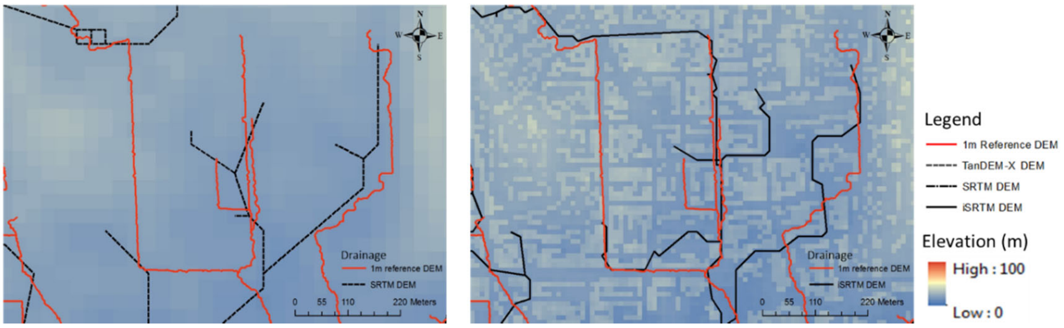
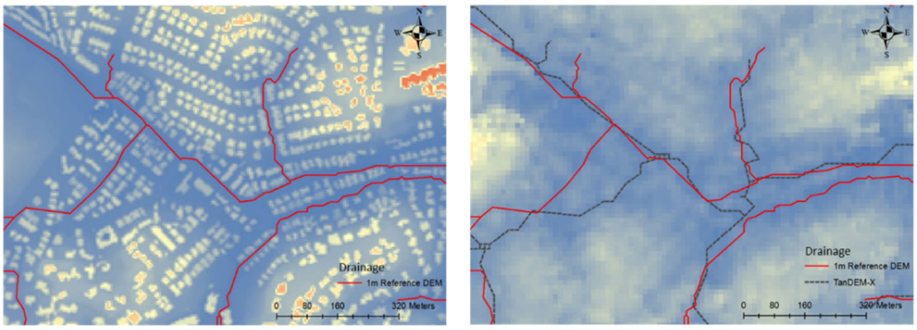


Figure 4.25 Comparisons of drainage networks derived from (a) 1 m reference DEM, (b) TanDEM-X DEM, (c) SRTM DEM and iSRTM_v2 DEM: Nice, France

(a) 1m reference DEM (b) TanDEM-X DEM



(c) SRTM DEM (d) iSRTM_v2 DEM

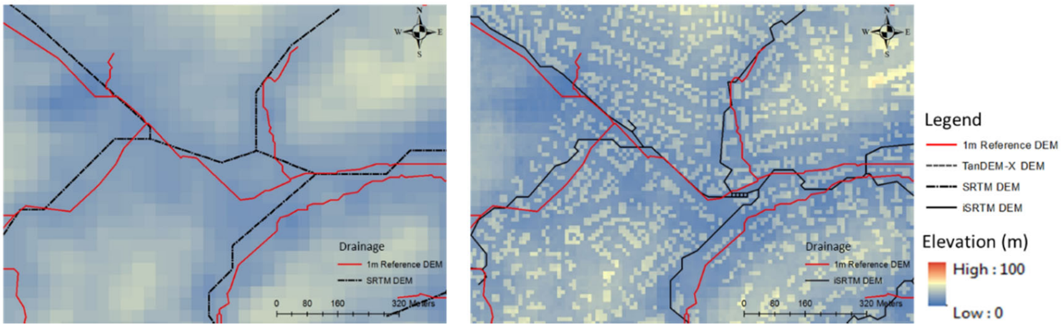


Figure 4.26 Comparisons of drainage networks derived from (a) 1 m reference DEM, (b) TanDEM-X DEM, (c) SRTM DEM and iSRTM_v2 DEM: Singapore

4.2.6 Summary

The performance of iSRTM DEM has been evaluated with various scenarios in 2 countries, France and Singapore. First the landuse type is the same for both training and test areas (i.e. dense urban areas; forested areas). The ANN for dense urban area considered two different approaches. Single training of the ANN (iSRTM_v1 DEM) and two separate ANN trainings of the ANN (iSRTM_v2 DEM, with buildings and no building) for dense urban areas; their performances were compared. iSRTM_v2 DEM which was trained from with buildings and no building heights separately, performed much better than the iSRTM_v1 DEM in which ANN was trained from the entire dense urban area without separating buildings from the rest.

It is of great interest to test the DEM quality of ANN, trained with various landuse patterns and characteristics, when it is applied in a faraway country within the range of the same landuse pattern and characteristics. The test performance showed its applicability with SRTM DEM improved significantly.

Table 4.4 Summary of performances of various scenarios

Scenarios	Description		Training (km^2)	Test (km^2)	RMSE (m)	
					SRTM DEM	iSRTM DEM
1	Trained and validated in Nice, France (landuse type: forested area)		16	8	14.37	7.84 (45.4 % ↓)
	Trained and validated in Nice, France (dense urban area)	Version 1	12	5.2	8.36	7.82 (6.5 % ↓)
		Version 2	12	5.2	8.36	5.18 (38.0 % ↓)
2	Trained and validated in Singapore (forested area)		1.84	1.54	14.13	6.01 (57.5 % ↓)
	Trained and validated in Singapore (dense urban area)		2.08	1.62	10.70	6.60 (38.3 % ↓)
3	Trained in Nice and validated in Singapore (forested area)		42.2	1.54	14.13	9.24 (34.6 % ↓)
	Trained in Nice and validated in Singapore (dense urban area)		42.2	1.62	10.70	6.93 (35.2 % ↓)

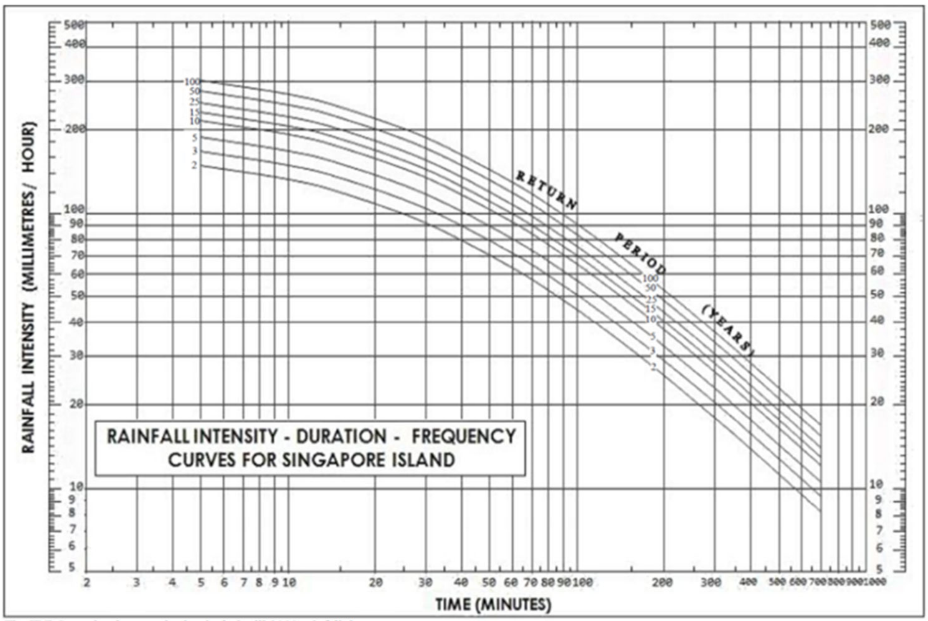
The iSRTM_v2 DEM was then compared with the German Aerospace Center's TanDEM-X DEM which has high-resolution (12 m) and high accuracy (10 m). The comparison showed that iSRTM_v2 DEM matches with the 1 m reference DEM closer than that of TanDEM-X DEM.

The drainage networks, another matrix to evaluate the performance of DEM, resulting from iSRTM_v2 DEM showed best agreement with the streamlines of the reference DEM. Table 4.4 summarizes the performances of the various scenarios.

4.3 IDF Curves Resulting from WRF/ERAI, CHIRPS and COP35

The rainfall data from WRF/ERAI and CHIRPS, for Singapore domain, are separately used to construct their respective IDF curves. They are then compared with their counterpart IDF curves constructed from rainfall data collected by Singapore's national water agency, Public Utilities Board [*PUB*, 2012].

Precipitation data extracted from WRF/ERAI and CHIRPS are separately checked on their homogeneity using the Regional Frequency Analysis (RFA). Then they are used to derive their respective IDF curves. Note that PUB has published IDF curves [*PUB*, 2012], with rainfall data from 35 rainfall stations, in Code of Practice (COP) on Surface Water Drainage (hereafter referred as COP35). COP35's IDF curves, shown in Figure 4.27, can be found in following link: https://www.pub.gov.sg/Lists/AppendicesDrawings/Attachments/5/app2.pdf.



The IDF Curve has been revised to include till 2009 rainfall data.

Figure 4.27 Singapore's IDF curves with rainfall record up to 2009 (extracted from PUB, 2012)

Figure 4.28 and Table 4.5 showed the comparisons between IDF curves resulting from WRF/ERAI, CHIRPS and COP35 at rainfall durations of 6, 12, 18 and 24 hours. It should be noted that COP35 provides the IDF curves only up to 12 hours; WRF/ERAI yields only data at 6, 12, 18 and 24 hours; CHIRPS provides only one data daily, thus at 24th hour. The IDF data of WRF/ERAI reasonably match their counterparts from COP35 albeit a slight overestimation (16.6 - 19.6 %).

CHIRPS' IDF data, however, can only be compared with its counterpart, WRF/ERAI, at 24th hour; CHIRPS significantly underestimated WRF/ERAI by 50.9 %. Although CHIRPS cannot directly be compared with COP35, Figure 4.28 showed CHIRPS would significantly underestimate the IDF curves of COP35.

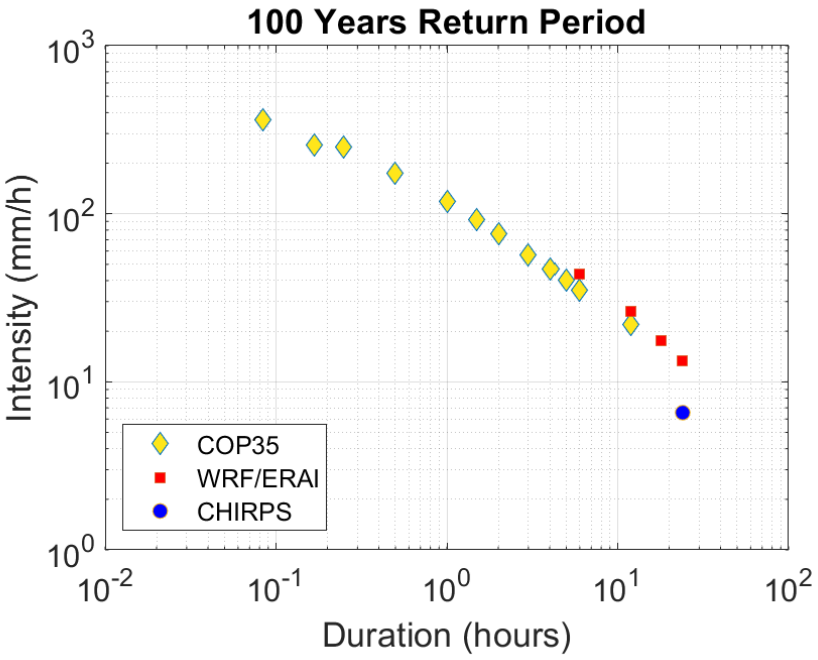


Figure 4.28 Comparisons of IDF curves resulting from WRF/ERAI, CHIRPS and COP35

Table 4.5 Extreme rainfall intensities resulting from WRF/ERAI, CHIRPS and COP35

Data source	**Rainfall intensity (mm/hr) at different rainfall durations**			
	6 hour	**12 hour**	**18 hour**	**24 hour**
WRF/ERAI	43.88	26.15	17.49	13.34
CHIRPS	-	-	-	6.55
COP35	35.29	21.81	-	-

4.4 Summary

This chapter evaluated the performances of the DEM improvement scheme and IDF curves from downscaled climate model. The following is a salient summary of this chapter:

- SRTM DEM, a publicly accessible elevation data, was significantly improved using a trained ANN with Sentinel 2 multispectral imagery and the high accuracy 1 m reference DEM. The performances were evaluated through various matrices: visual clarity, scatter plots, RMSE and drainage networks. Also, the improved SRTM DEM performs better than the high-resolution (12 m) German Aerospace Center's DEM, TanDEM-X DEM, in all performance matrices. The derived DEM from iSRTM_v2 DEM will then be used, in Chapter 5, as the input data of a numerical flood model.
- The performance of IDF curves from WRF/ERAI has been compared with the existing IDF curves in Singapore (COP35) derived from observation station data, and gridded observation CHIRPs data. IDF curves from WRF/ERAI match the IDF curves of COP35 reasonably well. Thus, IDF curves, for areas considered in Chapter 5, will be derived from WRF/ERAI to conduct the flood modelling.

5 Application of Proposed Approaches: Flood Simulation and Mapping of Greater Jakarta, Indonesia

5.1 Overview

This chapter serves as an application of the aforementioned proposed approaches to places where data scarcity could be a serious issue. Greater Jakarta, Indonesia, has been selected to generate the flood map with the following input data and model:

- **DEM derived from the improved SRTM DEM, iSRTM_v2 DEM**: the ANN was trained in Nice, France, and validated in Singapore, both with high spatial resolution and high accuracy reference DEMs. The visual comparison is conducted with the satellite imagery. The DEM resulting from iSRTM_v2 DEM is then used as the bathymetry data in the flood model.
- **IDF Curves derived from rainfall proxies resulting from WRF/ERAI**: Regional Frequency Analysis (RFA) is first applied on rainfall data, from WRF/ERAI, at various grids in and surrounding Greater Jakarta; heterogeneity and the discordancy tests are conducted. The IDF curves are then derived and distributed using the Chicago design storm method. The data is then used as the precipitation data of precipitation in the flood model.
- **Numerical model setup:** MIKE 21 Flow Model FM is used with input data from the above mentioned DEM and precipitation data. The simulation duration in this study is set at the Time Concentration (TC) as of the main interest is the maximum flood depths in the study domain. TC is calculated using the Passini method. Two return periods of 50- and 100-years are applied to the model and flood maps are then generated.

As presented in Section 1.3 and delineated in Figure 1.8, Greater Jakarta, the capital city of Indonesia, has an area of approximately 2,976 km^2 and is the most densely populated and urbanised city in Indonesia. Overall the catchment is sloping towards the northern Jakarta Bay. The upper part of catchment consists of mountains with rather steep slopes while the lower part the slope is very mild. Two different Times of Concentration (TCs) are considered based on the catchment characteristics. These will be further elaborated later in Section 5.4.

5.2 DEM Derivation

This section applies the developed methodology of improved SRTM (iSRTM_v2 DEM) to the study area, the Greater Jakarta, Indonesia. The ANN, applied to the Greater Jakarta, was trained in Nice, France, and validated in Singapore as presented in Section 4.2. As the study area does not have high accuracy DEM to be compared with, only visual comparisons are conducted with Google satellite imagery. Table 4.1 showed the data required for development of iSRTM_v2 DEM. The data are freely accessible by the publics through the listed websites. The SRTM DEM can be downloaded from USGS Earth Explorer and two tiles cover the Greater Jakarta. Sentinel 2 also can be downloaded from the same website of SRTM DEM and four tiles cover the Greater Jakarta. These data are merged in ArcGIS into one tile and the spatial resolutions are standardized at 30 m. Building footprints are available at OpenStreetMap website and it is used to filter the building areas in the DEM.

Table 5.1 Summary table of the data obtained for study area

Data	Source	Description
SRTM DEM	USGS Earth Explorer (https://earthexplorer.usgs.gov/)	ID: SRTM1S07E106V3 SRTM1S07E107V3 Acquisition date: 11 Feb 2000 Resolution: 1 arc (≈ 30 m)
Sentinel 2	USGS Earth Explorer (https://earthexplorer.usgs.gov/)	ID: L1C_T48MYT_A015616_20180619T031538 L1C_T48MYU_A014901_20180430T031909 L1C_T48MXU_A015616_20180619T031538 L1C_T48MXT_A015044_20180510T030910 Platform: Sentinel 2A Resolution: 10, 20, 60 m
Building footprint	OpenStreetMap (https://www.openstreetmap.org/export)	Building footprint shape file

The trained ANN, trained in Nice, France and validated in Singapore, is then applied to Greater Jakarta with data, for the input nodes, from SRTM DEM and Sentinel 2. Figure 4.1 showed the satellite imagery of 4 areas of Jakarta Metropolitan; they are used for visual comparisons. Figure 5.2 to Figure 5.6 showed the comparison between satellite imagery, SRTM DEM and iSRTM_v2 DEM for each of the 4 areas.

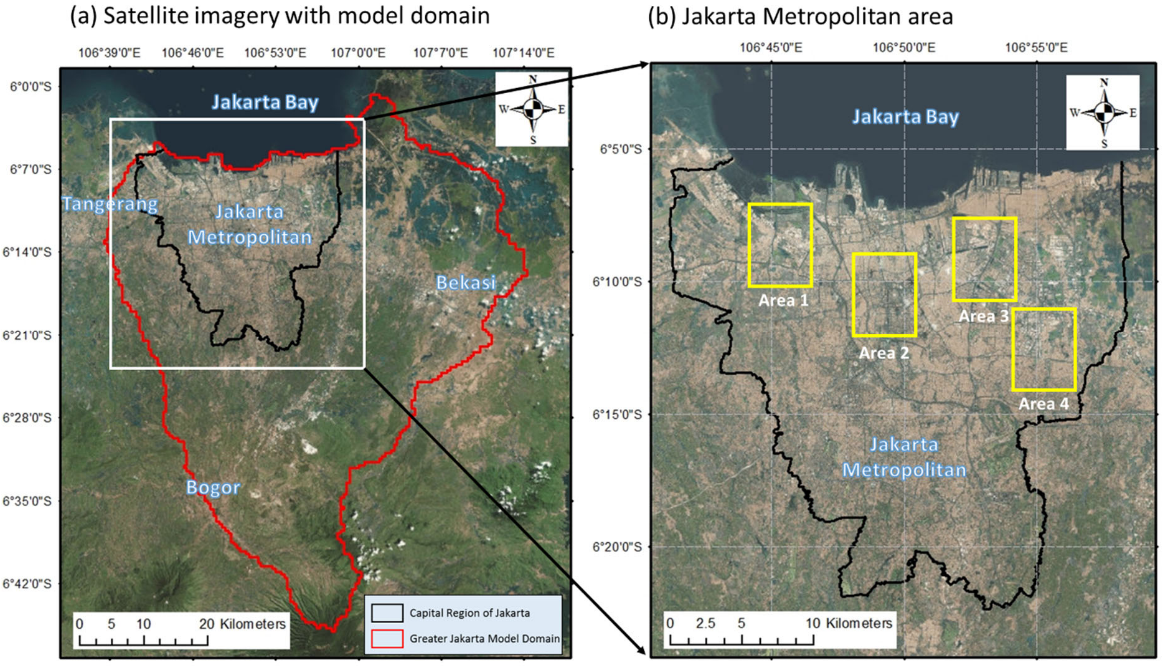


Figure 5.1 (a) Satellite imagery of study area; (b) Jakarta Metropolitan with four areas chosen for comparisons

Figure 5.2 showed the industrial area in the western side of Jakarta. The iSRTM_v2 DEM showed much clearer topography shapes than SRTM DEM, especially the river networks, buildings and roads. The elevations of buildings are higher than the roads and rivers in iSRTM_v2 DEM. These land surface characteristics can surely affect the water flow; thus, it produces more reliable simulated flood data in the area.

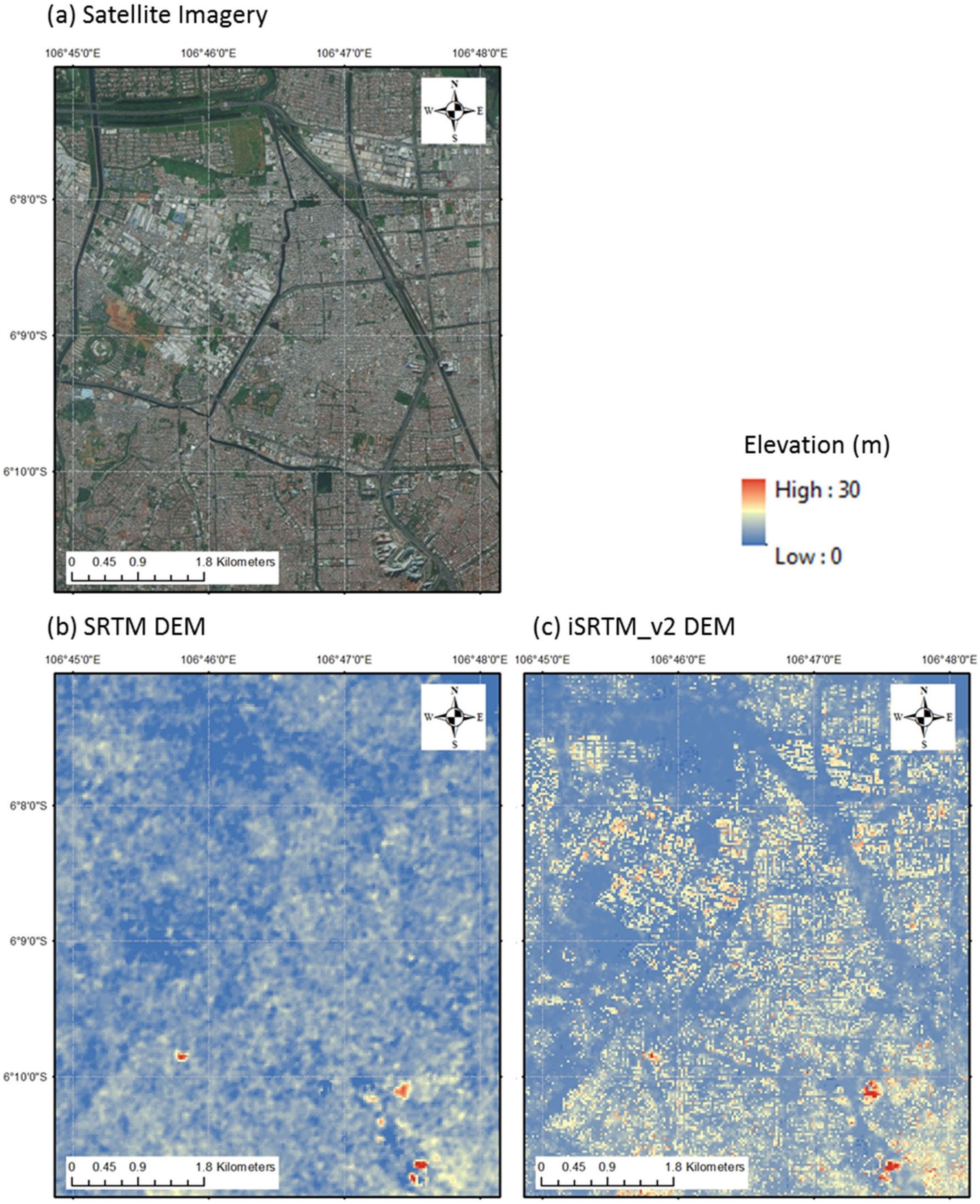


Figure 5.2 Comparisons between (a) satellite imagery, (b) SRTM DEM and (c) iSRTM_v2 DEM: area 1

(a) Satellite Imagery

(b) SRTM DEM (c) iSRTM_v2 DEM

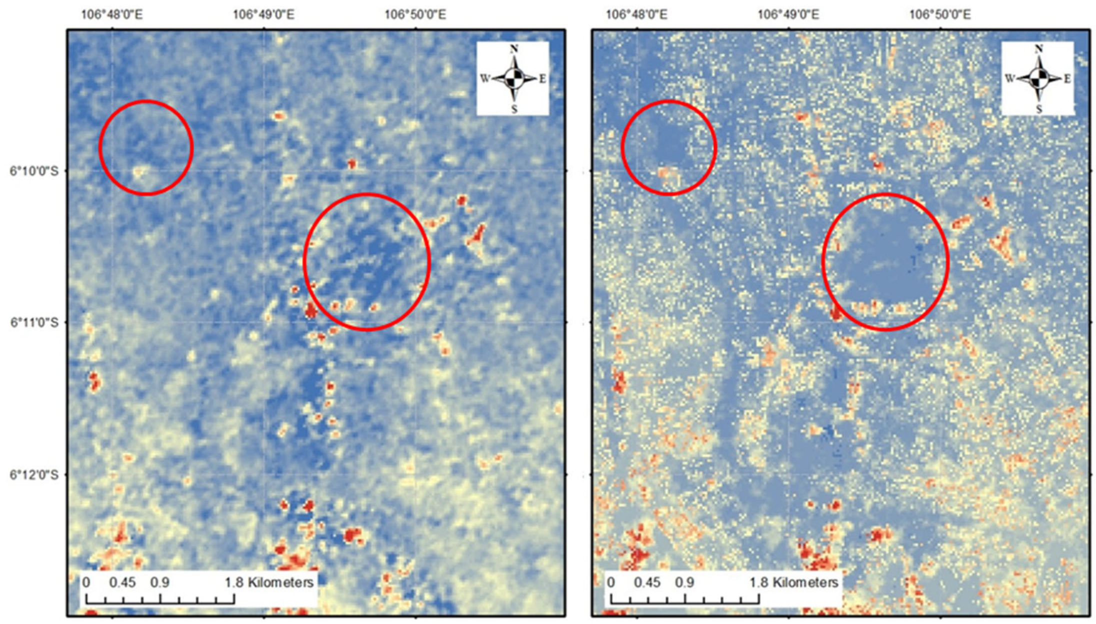


Figure 5.3 Comparisons between (a) satellite imagery, (b) SRTM DEM and (c) iSRTM_v2 DEM: area 2

Figure 5.3 showed the central area of Jakarta near National Monument Park. The high buildings and the roads are much clearer in iSRTM_v2 DEM; the circled area consists mainly of vegetation and flat area as shown in Figure 5.4. In the SRTM DEM, it has scattered high elevations from the trees. These canopies are removed in the iSRTM_v2 DEM and it is flat and has lower elevation than the surrounding areas. This may affect the flood map with underestimated flood extent in the vegetated area.

(a) Google street view of circle 1 in Figure 5.3

(b) Google street view of circle 2 in Figure 5.3

Figure 5.4 Google street (a) view of circle 1 and (b) view of circle 2 in Figure 5.3

Figure 5.5 and Figure 5.6 showed the industrial area in the northern side of Jakarta and residential area in the eastern side of Jakarta respectively. The buildings, roads and canals are much clearer in the iSRTM_v2 DEM. Also, the vegetated areas are having lower elevations than surrounding areas in the iSRTM_v2 DEM due to the removal of the canopy.

(a) Satellite Imagery

(b) SRTM DEM

(c) iSRTM_v2 DEM

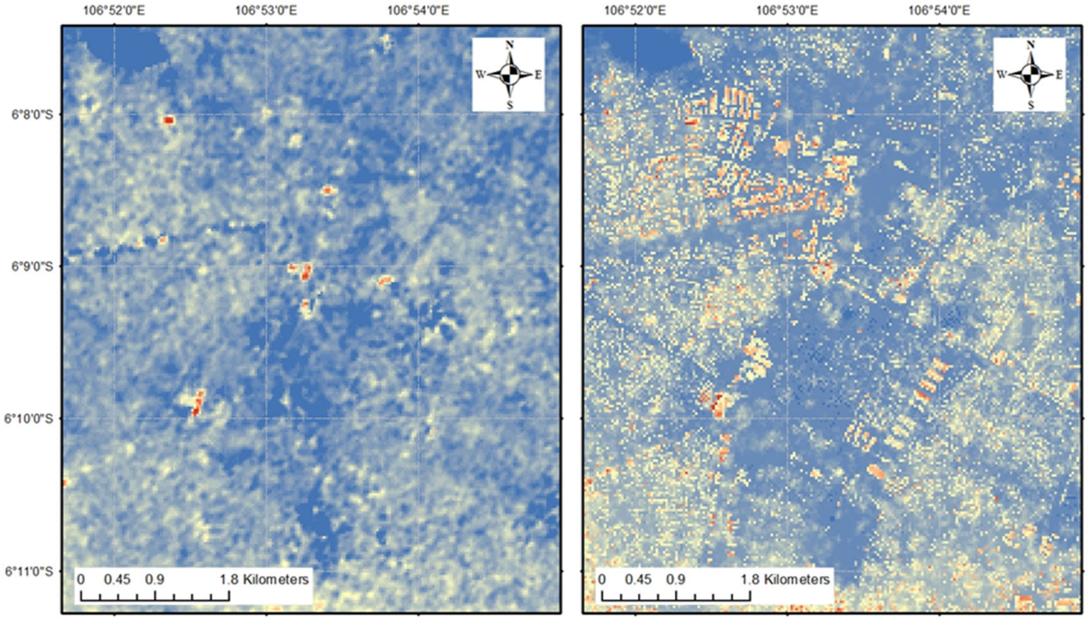


Figure 5.5 Comparisons between (a) satellite imagery, (b) SRTM DEM and (c) iSRTM_v2 DEM: area 3

(a) Satellite Imagery

(b) SRTM DEM

(c) iSRTM_v2 DEM

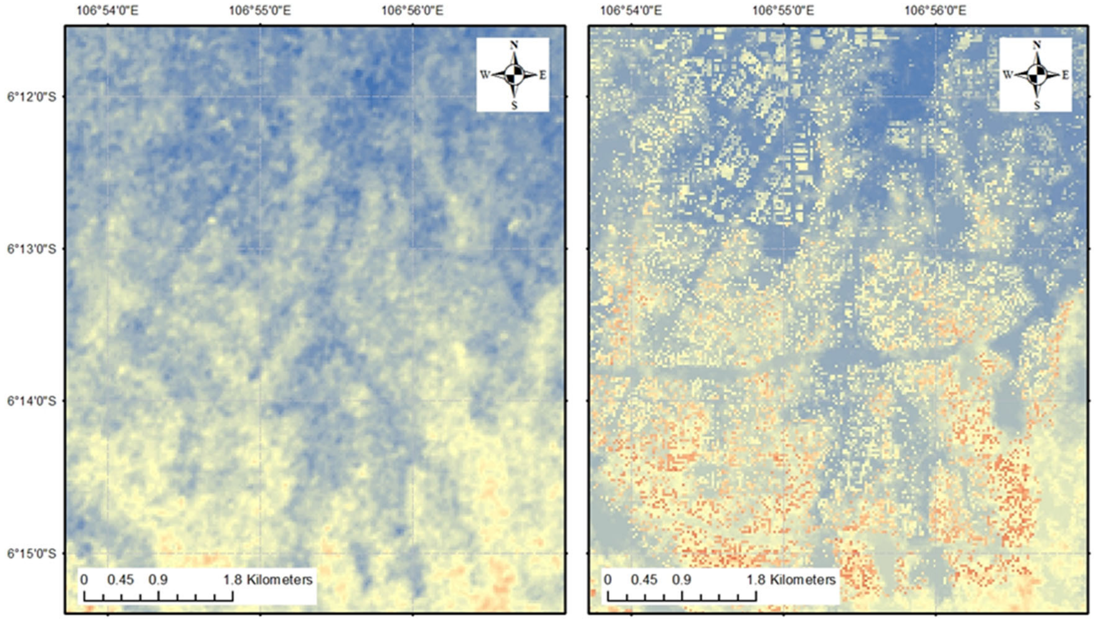
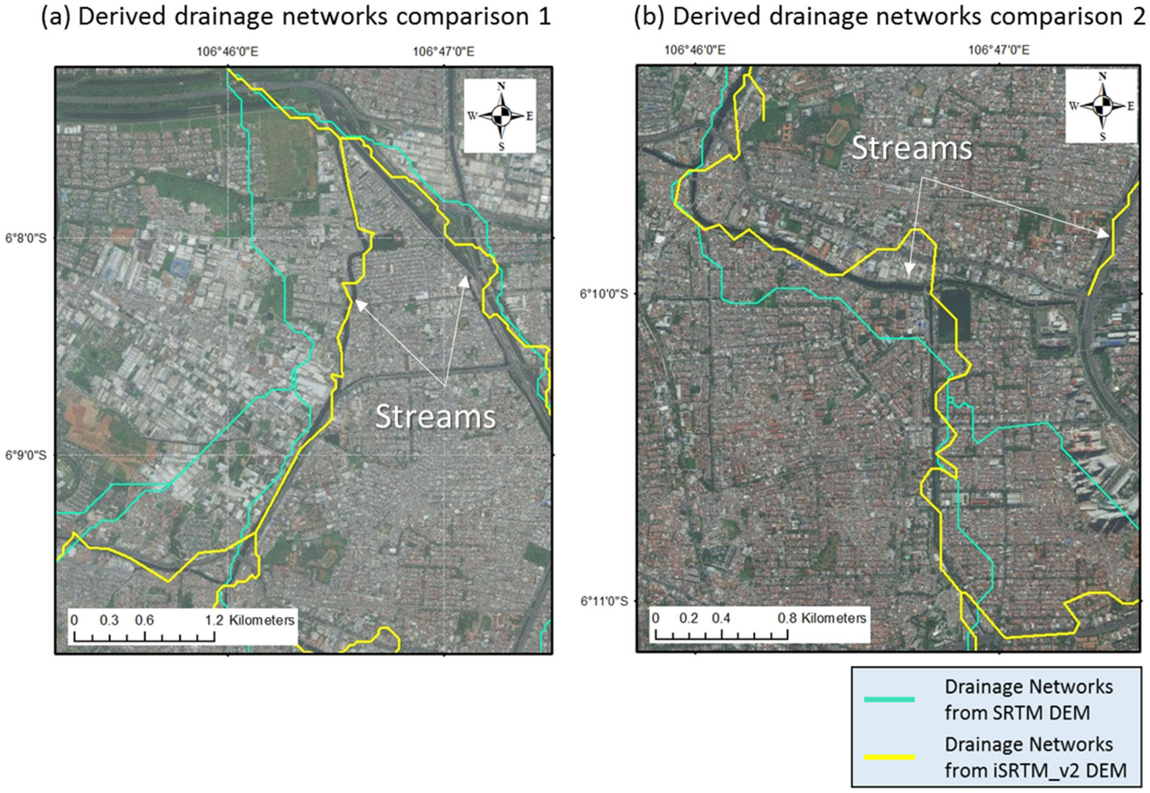


Figure 5.6 Comparisons between (a) satellite imagery, (b) SRTM DEM and (c) iSRTM_v2 DEM: area 4

The drainage networks derived from SRTM and iSRTM_v2 DEM are compared with the drainage networks of satellite imagery. Figure 5.7 showed the drainage networks of each DEM. The mint green colour is from SRTM DEM while yellow is from iSRTM_v2 DEM. These networks are compared with their counterparts in the satellite imagery. The drainage network resulting from iSRTM_v2 DEM showed the closest agreement with the drainage network in the satellite image. Again, the accuracy of the drainage networks would certainly affect the flood map significantly.

The derived iSRTM_v2 DEM is later used in the flood model as the bathymetry.



(a) Derived drainage networks comparison 1

(b) Derived drainage networks comparison 2

Figure 5.7 Drainage network comparisons between SRTM (mint green) and iSRTM_v2 DEM (yellow) in (a) area 1 and (b) area 2

5.3 IDF Curve Derivation

5.3.1 Regional Frequency Analysis

This section investigates the frequency and magnitudes of extreme rainfall events resulting from WRF/ERAI for greater Jakarta, as mentioned in Section 3.3.4.3. Nine grid points, circled in Figure 5.8, are the rainfall stations which met the discordancy and heterogeneity tests of the Regional Frequency Analysis (RFA). The RFA is conducted for up to 7 days.

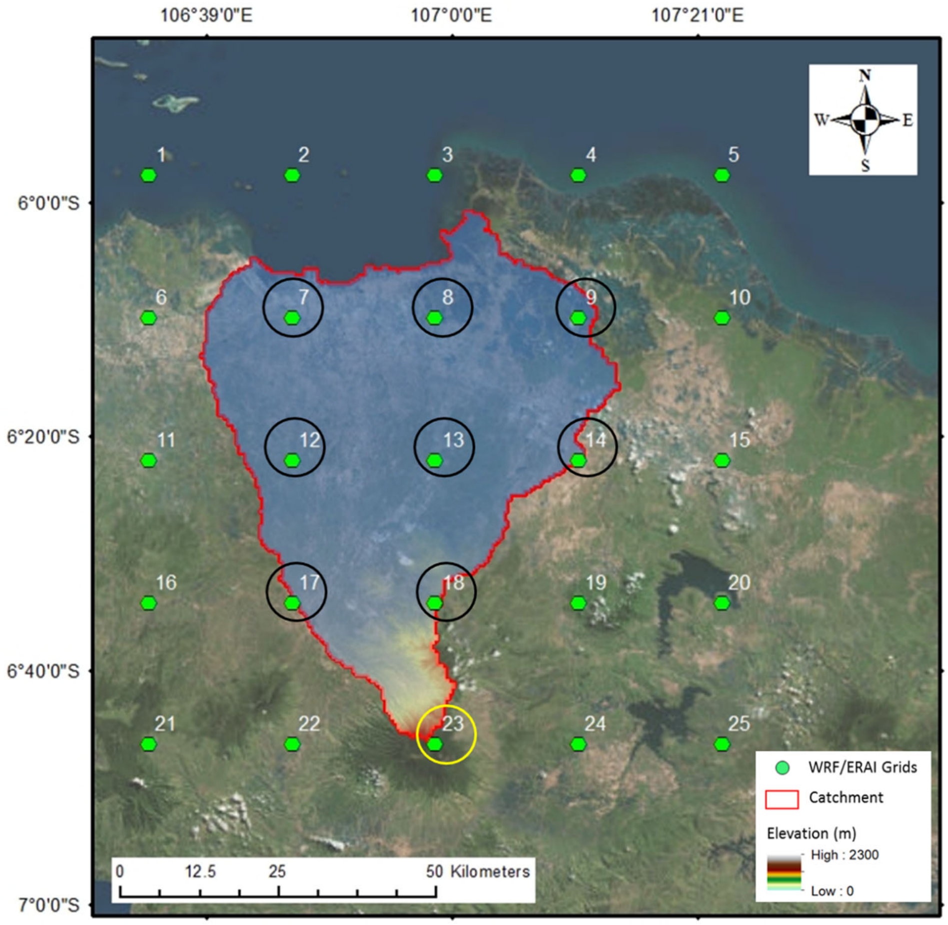


Figure 5.8 WRF/ERAI grid points considered for RFA

The criteria established by *Hosking and Wallis* [1997] are the discordancy (D_i) and heterogeneity (H) measures as described in the equation (3-1) and (3-2). The critical values of D_i for regions of various sizes, presented in the Table 3.8, and the data at all sites with large values D_i should be carefully scrutinized before deciding whether the sites are discordant. The heterogeneity tests check the value of H, with H < 1 considered as acceptably homogeneous region, 1 ≤ H < 2 possibly heterogeneous region, and H ≥ 2 definitely heterogeneous region.

Based on the discordancy measures in Table 5.2, grid point number 23 (see Figure 5.8) is excluded.

Table 5.2 Discordancy (D_i) measure of study area

	Grid Point P7	P8	P9	P12	P13	P14	P17	P18	P23
1 day	1.48	1.89	1.27	0.63	0.56	0.77	0.07	0.01	2.33
2 days	0.71	1.89	1.59	0.19	0.66	0.69	0.19	1.3	1.78
3 days	0.99	1.67	0.15	0.54	0.64	0.98	1.31	0.82	1.9
4 days	1.11	1.46	0.11	0.5	0.75	1.29	1.43	0.3	2.05
5 days	1.29	1.95	0.09	0.53	0.61	1.14	1.15	0.22	2.01
6 days	1.55	2.06	0.46	0.44	0.69	0.61	0.95	0.28	1.97
7 days	1.54	1.85	0.6	0.5	0.68	0.37	1.06	0.4	2

Table 5.3 Heterogeneity (H) measure of study area

Rainfall duration 1 day	2 days	3 days	4 days	5 days	6 days	7 days
1.11	1.41	1.41	1.06	1.2	1.16	1.13

5.3.2 IDF Curves and Design Storms

Using RFA method, the rain gauges are statistically resembling the rainfall characteristics into a single homogeneous region; they are then used to estimate the various representative regional IDF curves. The general procedure of IDF curve derivation is described below.

Step 1: Derive Annual Maximum Rainfall (AMR) series from the output of RFA over Greater Jakarta

Step 2: Fit a Cumulative Distributed Function (CDF) to the data of Step 1 with a specific rainfall duration (1-, 2-, 3-, 4-, 5-, 6- or 7-days) for each return period (50- or 100-years)

Step 3: Calculate the rainfall intensity from the fitted CDF of different rainfall durations derived from Step 1 and Step 2.

Step 4: Repeat Step 3 for all return periods to construct IDF curves.

As described in Section 3.3.4.2, Chicago Design Storm method is used to derive the design storms. Figure 5.9 and Table 5.4 showed the IDF curves and their fitted parameters for Greater Jakarta. Table 5.5 showed the rainfall intensities for different durations and different return periods.

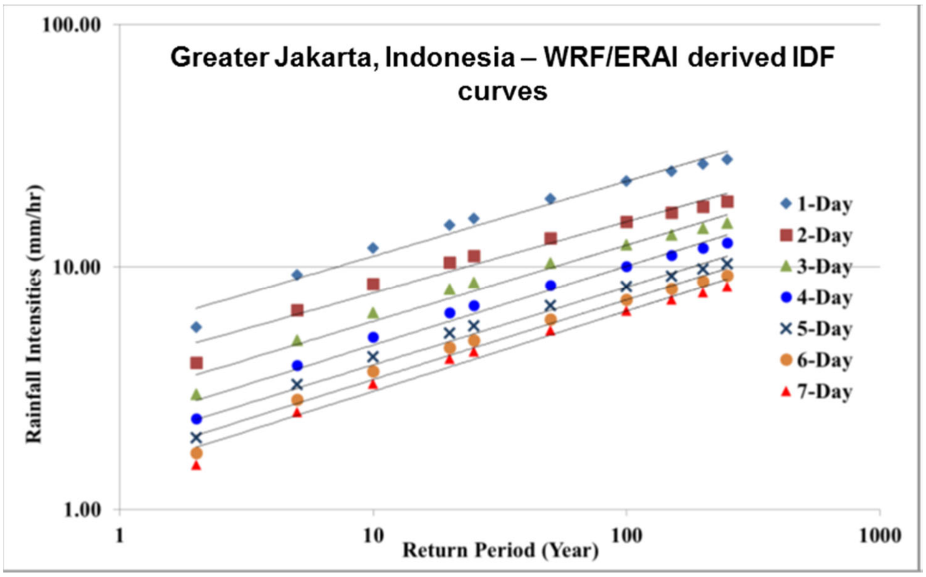


Figure 5.9 WRF/ERAI derived IDF curves for Greater Jakarta: duration of 1 to 7 days, return periods of 50- and 100-years

Table 5.4 Chicago Design Storm method fitted equations for Greater Jakarta IDF curves

Parameter	Fitted Equation ($I = A/(t_d + B)$)	
	50 year	100 year
A	53,712	40,979
B	2,271.2	2,650.1

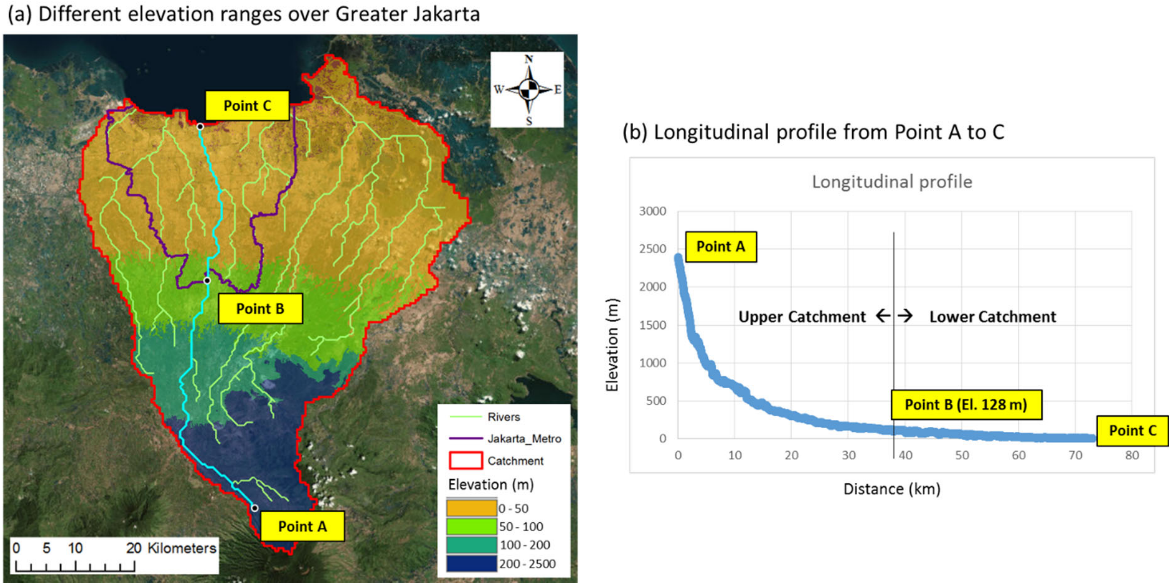
Table 5.5 Rainfall intensity (mm/hr) derived from WEF/ERAI for different return periods of 1 to 7 days storm durations

Return Period (Year)	Duration (days)						
	1 day	2 days	3 days	4 days	5 days	6 days	7 days
50	8.59	6.41	5.12	4.26	3.65	3.19	2.83
100	10.02	7.41	5.88	4.87	4.16	3.63	3.22

5.4 Flood Model Setup

Time of Concertation (TC) is important in defining the required precipitation duration in flood modelling. When TC of an upstream subcatchment is relatively short while the entire catchment length is long, the high designed rainfall intensity of upstream subcatchment may not have insignificant or no impact on the downstream subcatchment. On the other hand, when TC is relatively long, the low rainfall intensity will may have impacts on the downstream subcatchment with underestimated maximum flood depths.

To calculate the TC, using the Passini method, for the study area the catchment is divided into upstream and downstream subcatchments based on its topography. The Passini equation was described in Section 3.4.2.2. Based on the TC calculation the upstream subcatchment has relatively short duration due to its steep slope, while the downstream subcatchment has relatively long rainfall duration due to its flat topography as shown in Figure 5.10. In this study, the peaks of the two design storms are placed at the same time in the middle of rainfall durations, Figure 5.11, to generate the most conservative flood depth scenario. Table 5.6 showed the parameters for the calculation of TC.


(a) Different elevation ranges over Greater Jakarta

(b) Longitudinal profile from Point A to C

Figure 5.10 (a) Different elevation ranges over Greater Jakarta and (b) longitudinal profile from point A to point B

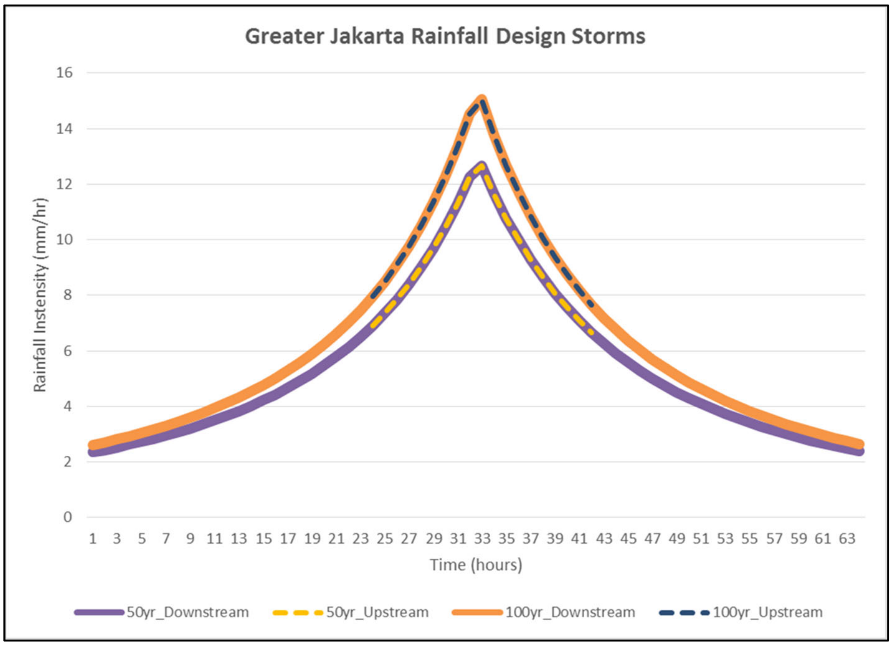


Figure 5.11 Rainfall design storms for Greater Jakarta upstream and downstream catchments

Table 5.6 Parameters to calculate the Time of Concentration (TC)

Catchment	Elevation Range (m)	Highest (m)	Lowest (m)	Length (m)	Slope (m/m)	Area (km^2)	Time of Concentration TC (min)
Upstream	128 – 2,856	2,856	128	44,605	0.061	1,785.6	1,123
Downstream	0 – 128	128	0	36,196	0.004	1,190.4	3,806

As mentioned in Section 3.4.2.1, different computational mesh sizes are applied to the catchment. For the city and industrial areas, a fine mesh size of 900 m^2 is applied which is equivalent to the grid resolution of 30 m (maximum). For the other areas such as mountains and plain areas, a course mesh size of 8,100 m^2 (the equivalent of 90 m grid resolution) is used. Figure 5.12 showed the model domain with the fine and course mesh zones.

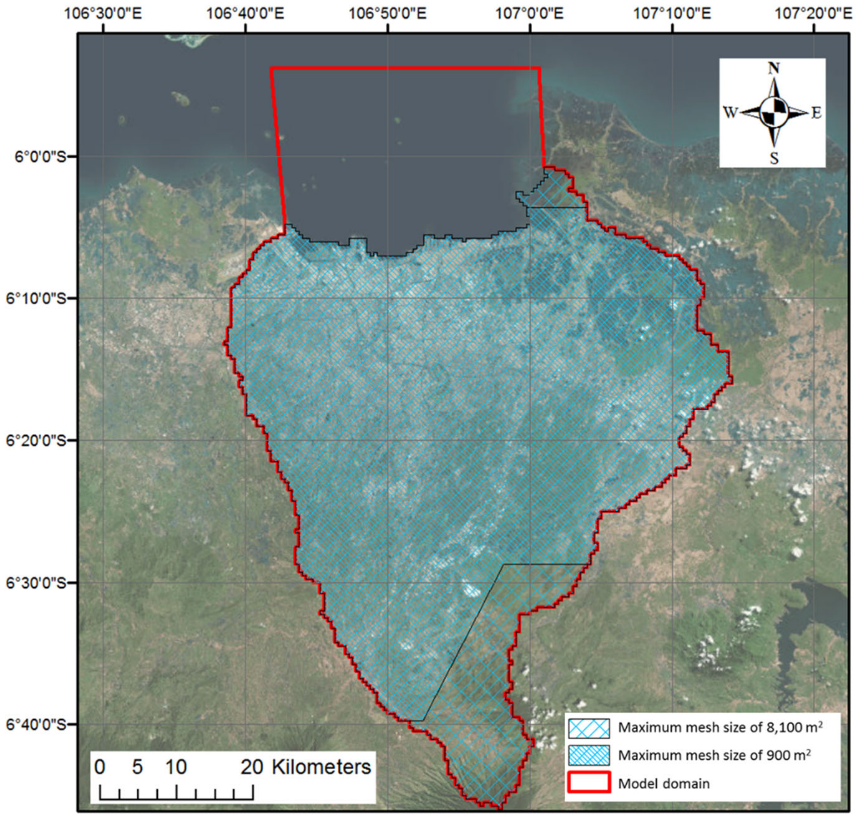


Figure 5.12 Areas with fine and coarse meshes

Table 5.7 summarizes the flood model setup of MIKE 21 FM. The computational area of the model is 4,807 km^2 including sea area. The simulation time is set at 63.4 hours based on the calculated TC. The model time step is dynamic varying between 0.02 second to 1 second. The hourly tide level is adopted from Global Tide Model as mentioned in Section 3.4.2.3. The variation of tide level is from -0.33 m to 0.44 m. Two design storms (50- and 100-year return periods) are simulated from the IDF curves derived from WRF/ERAI for the study area. Improved SRTM_v2 DEM is used as the bathymetry. Each scenario simulation takes about 28 hour computational time. The generated flood maps are presented in next section.

Table 5.7 MIKE 21 FM model setup

Input type	Remarks
Computation area	4,807 km^2 (Inclusive the sea area)
Simulation time	63.4 hours
Time step	Dynamic time step: 0.02 - 1 sec
Tide	Hourly tidal level extracted from Global Tide Model

Rainfall	IDF Design Storms with 50- and 100-year return periods
DEM	30 x 30 m resolution Improved SRTM_v2 DEM
Mesh information	Number of meshes: 5,961,648 Number of nodes: 2,985,072
Computational time	28 hours per simulation

5.5 Flood Maps of Different Scenarios

Flood maps of 2 return periods of 50- and 100-years are generated. For the 100-year return period, two flood maps, one with SRTM DEM while the other with iSRTM_v2 DEM, are compared. In the final process of flood mapping, water bodies are masked using river data from OpenStreetMap and NDWI (Normalised Difference Water Index) derived from Sentinel 2. The OpenStreetMap provides water body polygons for lakes, rivers and canals with shapefile format. It is converted to raster format with 30 m resolution. Note that no raster is generated for places with river width of less than 30 m.

NDWI is calculated based on two remote sensing indexes related to liquid water [*Gao*, 1996]. The Green spectrum (band 3 of Sentinel 2) and Near Infra-Red spectrum (NIR; band 8 of Sentinel 2) are used in the calculation as shown in following equation:

$$\mathbf{NDWI} = \frac{X_{Green} - X_{NIR}}{X_{Green} - X_{NIR}} \qquad (5\text{-}1)$$

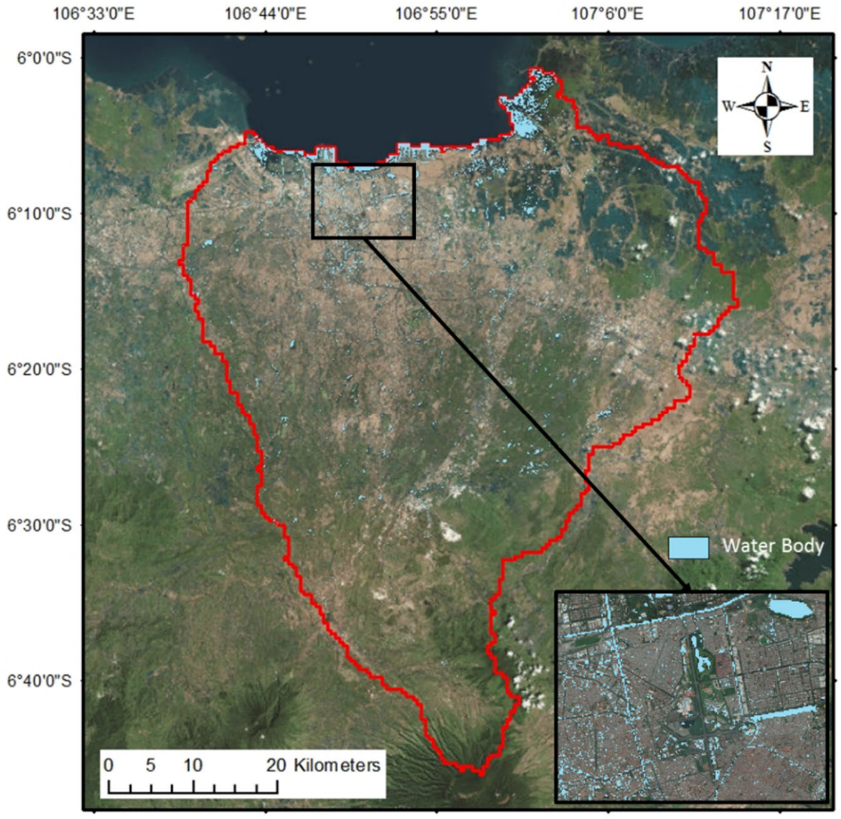


Figure 5.13 Water bodies in Greater Jakarta, Indonesia

The reason for using the two data for water body masking is because OpenStreetMap does not capture the fisheries and artificial ponds while NDWI does. Figure 5.13 showed the water body of the study area. The area of waterbody is removed in the final flood map.

Figure 5.14 showed the maximum flood depths (m) in Greater Jakarta with 50- and 100-year return periods resulting from iSRTM_v2 DEM and 100-year return period resulting from SRTM DEM. In Figure 5.14 (a) and (b), the inundation depths and areas increase with increasing return period increases – for the case with iSRTM_v2 DEM. It should be noted that for the case with the flood map of SRTM.DEM, Figure 5.14 (c), its inundated area is less than that of its counterpart with iSRTM_v2 DEM. This is expected as, in comparison to iSRTM_v2 DEM (1) the longitudinal slopes in SRTM DEM are slightly steeper towards the open sea; and (2) many more depression storages in SRTM DEM.

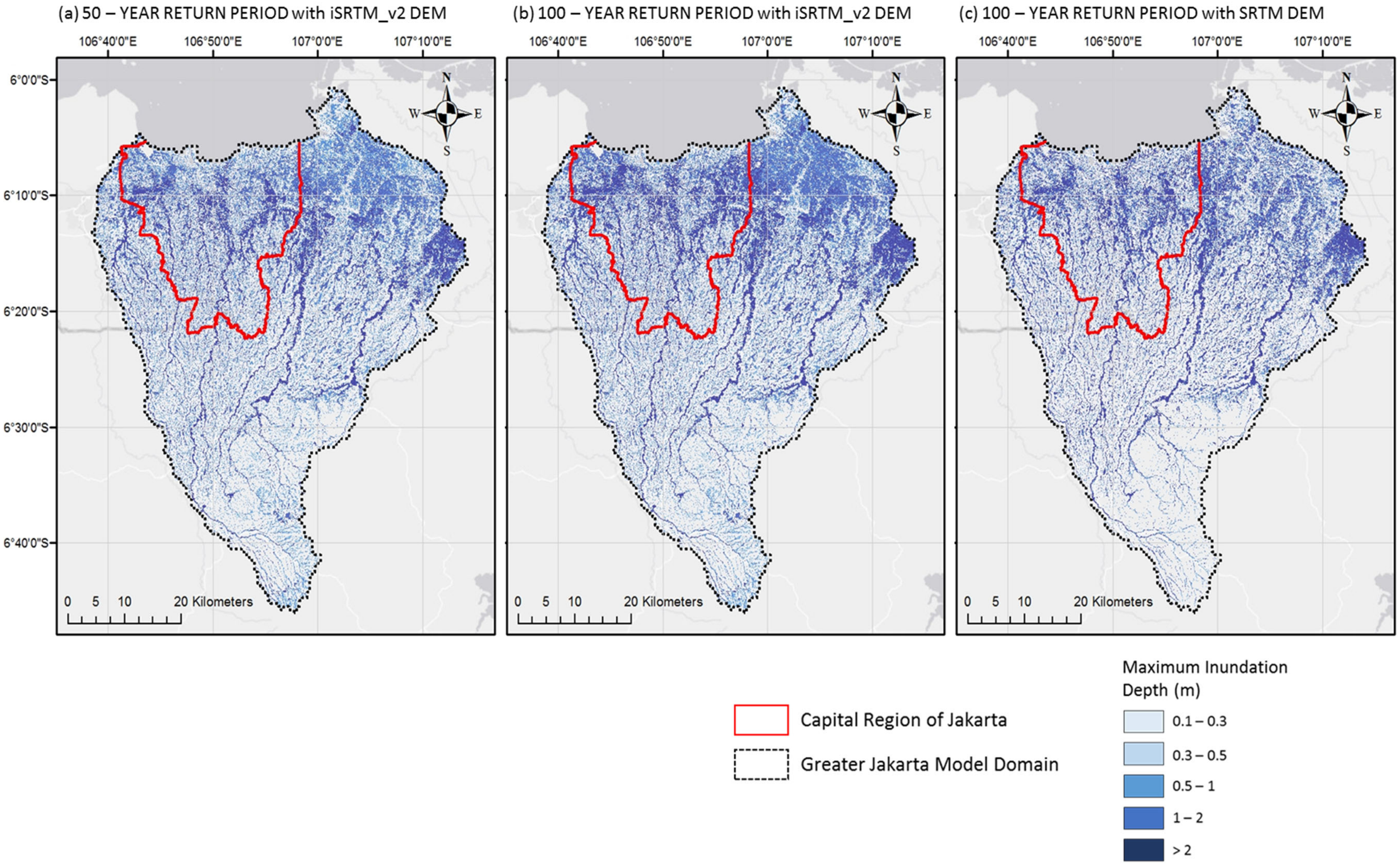


Figure 5.14 Maximum flood depths (m) of Greater Jakarta with (a) 50- and (b) 100-year return period from iSRTM_v2 DEM and (c) 100-year return period from SRTM DEM

The statistics of the inundated areas based on different water depths are shown in Table 5.8. It can be seen that total inundated areas increases with increasing return period. The inundation area of 0.1 - 1 m flood depth is decreasing with higher return period, while the areas with greater than 1 m are increasing. In the flood extent resulting from iSRTM_v2 DEM, 34.3 % of the catchment is inundated with 50-year return period and 36.3 % of the area is inundated with 100-year return period. While the flood map of 100-year return period from SRTM DEM has 27.1 % of the inundated area which is less than the flood maps from iSRTM_v2 DEM. The flooded area, in the range between 0.1 m to 2 m, of 100-year return period from SRTM DEM showed less areas than iSRTM_v2 DEM's flood maps. However, the flood depth of greater than 2 m, flood map from SRTM DEM has more areas than flood map of iSRTM_v2 DEM. This implies that the presence of canopy level in DEM produces less flooded area but has greater water depths due to lack of connectivity in the DEM.

Table 5.8 Inundation statistics of different scenarios

Return periods	**Inundated area (km^2) at different flood depths (m)**											
	0.1 – 0.3		**0.3 – 0.5**		**0.5 – 1**		**1 – 2**		**> 2**		**Total**	
	km^2	**%**	**km^2**	**%**	**km^2**	**%**	**km^2**	**%**	**km^2**	**%**	**km^2**	**%**
50-yr (iSRTM)	209.1	7.0	137.5	4.6	296.5	10.0	271.6	9.1	106.6	3.6	1,021.27	34.3
100-yr (iSRTM)	200.0	6.7	130.8	4.4	288.0	9.7	318.5	10.7	142.4	4.8	1,079.65	36.3
100-yr (SRTM)	94.2	3.2	133.9	4.5	155.9	5.2	228.3	7.7	194.2	6.5	806.57	27.1

Figure 5.15 (a) showed the flood footprints of Jakarta flooding in 2013 [*InaSafe*, 2015]; no return period was, however, mentioned. The main purpose of presenting it is to check whether the flood prone areas of flood map with iSRTM_v2 DEM (Figure 5.15 (b)) agree more with the flood footprints of Jakarta flooding than its counterpart with SRTM DEM (Figure 5.15 (c)). Generally, more flood extent and higher flood depths (the three boxed areas) are observed in the flood maps of iSRTM_v2 DEM; which are the case shown in Figure 5.15(a). This again demonstrates the higher accuracy of iSRTM_v2 DEM over SRTM DEM. More flood map comparisons can be found in the appendix.

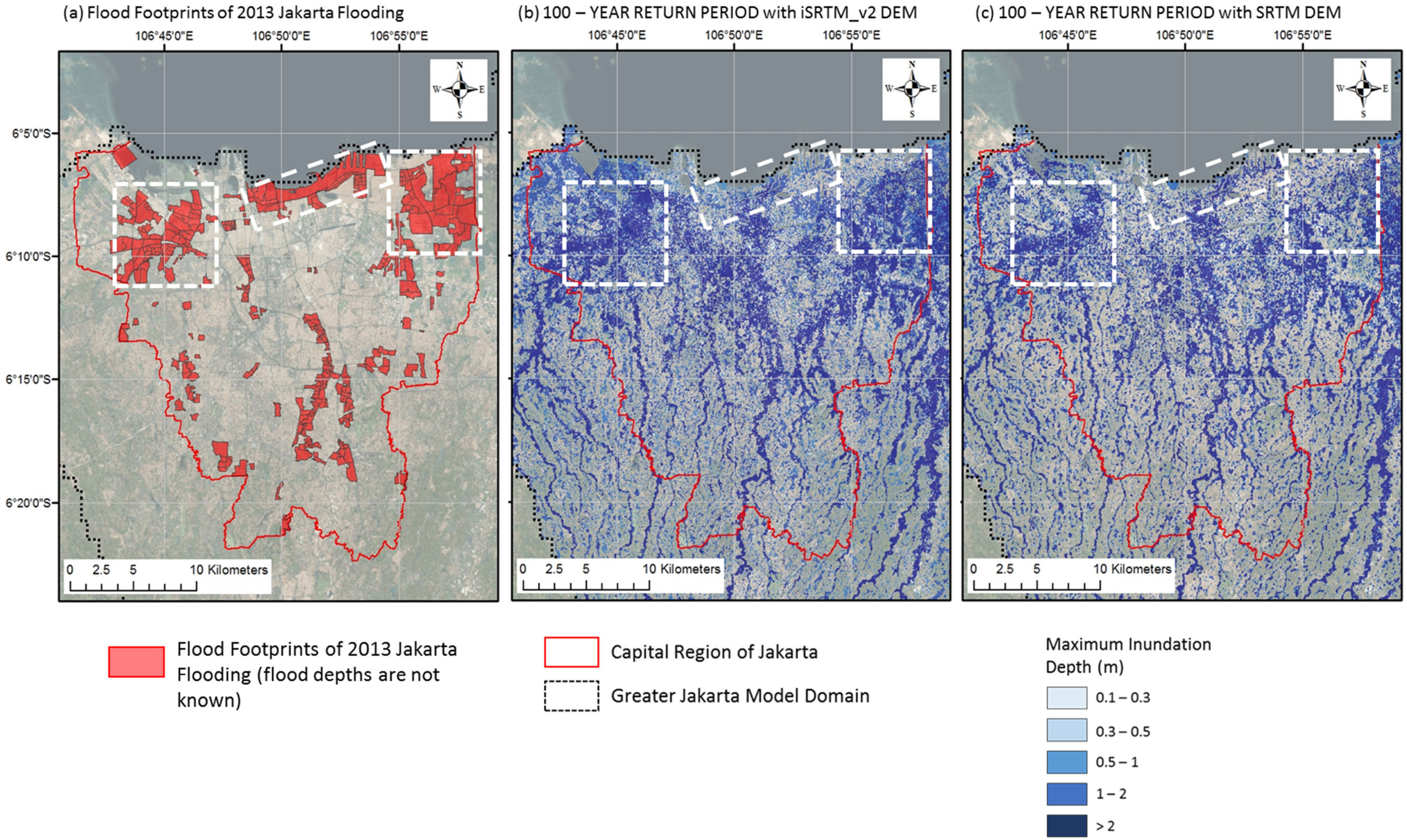


Figure 5.15 Comparison of (a) flood footprints of 2013 Jakarta flooding; 100-year return period of flood maps from (b) iSRTM_v2 DEM; and (c) SRTM DEM

5.6 Summary

This chapter applied the developed methods, the improved SRTM_v2 DEM and IDF curved using rainfall proxies from WRF/ERAI, to generate the flood maps of the Greater Jakarta, Indonesia. The following are the salient points of this chapter:

- Improved SRTM_v2 DEM, trained in Nice, France and validated in Singapore, significantly improved the original SRTM DEM of the study area where no high accuracy DEM was available. Visual comparisons of satellite imagery, SRTM DEM and iSRTM_v2 DEM clearly categorically showed much clearer bathymetry in iSRTM_v2 DEM. Similarly, comparison of drainage networks resulting from SRTM DEM and iSRTM_v2 DEM was conducted with, again, a much better drainage network resulting from iSRTM_v2 DEM. Thus, iSRTM_v2 DEM is strongly recommended as the bathymetry input for the flood model.
- IDF curves were derived, from WRF/ERAI rainfall data, after going through the various test criteria of Regional Frequency Analysis. Chicago Design Storm method was chosen to derive the design storms (of 50- and 100-year return periods). The different durations of design storms were applied in the upstream and downstream subcatchments with their respective Time of Concentrations, calculated from the Passini equation. The data was then used as the rainfall input in the flood model.
- The maximum flood depths were generated for different scenarios. For 100-year return period , the flood map with SRTM DEM and one with iSRTM_v2 DEM were compared with the flood footprints from 2013 Jakarta flooding. The flood map from iSRTM_v2 DEM was able to capture the better qualitative agreement with flood footprints and reasonable flood extent.

6 Summary, Conclusions and Recommendations

6.1 Introduction

The main objective of this research is to provide high accuracy flood simulation assessments, to policy makers, of areas challenged by the scarcity or low quality of physical data (e.g. Digital Elevation Model) and climatic data (rainfall). Collating high-resolution and accuracy DEM is not only costly, but also time consuming. Rainfall data are often not available and/or not sufficiently long to yield rainfall design curves (Intensity-Duration-Frequency (IDF) curves) of high accuracy required for urban storm drainage designs.

To resolve the aforementioned challenges on DEM and rainfall data, the study considers a publicly accessible satellite DEM (SRTM DEM) and rainfall proxy resulting from a high-resolution regional climate model (RCM) downscaling. Section 6.2 summarizes the DEM improvement scheme developed in this study and its performance evaluations. Section 6.3 summarizes the accuracy of rainfall proxies from RCM from which IDF curves are derived. Section 6.4 showed how useful the proposed approaches are through demonstration of its application on Jakarta, the densely populated and urbanized capital city of Indonesia. Recommendations is presented in Section 6.5.

6.2 Development of DEM Improvement Scheme

The DEM improvement scheme was developed using ANN with SRTM DEM, together with Sentinel 2 multispectral imagery, as input nodes and high-resolution and high accuracy surveyed DEM (the reference DEM) as the target layer. The trained ANN was able to classify the landuses and land covers with the assistance of 8 bands of Sentinel 2. Based on the various land characteristics, different weight were calculated to reduce the error between the elevation of SRTM DEM and reference DEM.

A series of scenarios was taken into consideration for calibration and of validation: (1) forested areas (Nice, France and Singapore), and (2) dense urban areas (Nice, France and Singapore). The performance of improved SRTM (iSRTM DEM) was shown to be significantly better, in all scenarios, than the original SRTM DEM. In forested areas the performance of iSRTM DEM showed 45.4 % and 57.5 % RMSE reduction in Nice, France and Singapore, respectively. In dense urban areas, the RMSE reduction of the iSRTM DEM is about 38 % for both Nice, France and Singapore. It is interesting to note that for dense urban areas the improvement of SRTM DEM showed also: (1) much clearer view of landshapes, buildings and roads; (2) closer agreement with the drainage networks of reference DEM.

As one of the interests in the study was to improve the SRTM DEM of faraway countries where no high quality DEM (like those DEMs in Nice and Singapore, in this study) is available to train ANN, the study looked into whether the ANN trained in Nice can provide reasonably high quality DEM in Singapore which was used for testing. The test performance showed significant improvement with RMSE reduction of 34.6 % and 35.2 % in forested and dense urban areas respectively. Upon successful experiment, the ANN trained in Nice was then used with high confidence to improve SRTM DEM elsewhere, the Greater Jakarta in this study.

In the study, the iSRTM DEM was also compared with German Aerospace Center's TanDEM-X, a high-resolution and accuracy DEM (12 m resolution; 10 m accuracy vertically; at a cost of €30/km^2 and a minimum purchase of 500 km^2). Comparisons, done in both cities (Nice, France and Singapore) showed that iSRTM DEM is in better agreement with the reference DEM than TanDEM-X DEM. The main conclusion is: even a high-resolution DEM from satellite, ANN's DEM improvement scheme has been convincingly demonstrated that it can still significantly improve the high-resolution satellite's raw DEM; high-resolution and high-quality surveyed DEM must be a part of the DEM improvement training scheme.

6.3 Derivation of IDF Curves from Regional Climate Model

The high-resolution Regional Climate Model (RCM; 20 Km resolution) WRF driven by reanalysis climate data, ERAI (WRF/ERAI), yields 6-hourly precipitation data from 1986-2005 over Southeast Asia domain (80E-125E, 15S-26N). The rainfall data has been compared reasonably well with gridded observation data, e.g. CHIRPS. The precipitation data extracted from WRF/ERAI and CHIRPS were separately checked on their homogeneity using the Regional Frequency Analysis (RFA). They were then compared with the IDF curves of Singapore (COP35). The IDF data of WRF/ERAI reasonably match their counterparts from COP35 albeit a slight overestimation (16.6 - 19.6 %). CHIRPS' IDF data, however, significantly underestimated COP35's IDF by 50.9 %. Thus, the main conclusion is: when high quality rainfall data, both temporal and spatial resolutions, are not available, one possible data source to extract rainfall proxies is dynamically downscaled regional climate model driven by reanalysis data such as ERA-Interim, NCEP-RA.

Upon successful testing of the applicability of WRF/ERAI data in Singapore, the IDF curves for Greater Jakarta were derived for later usage in flood modelling. Rainfall data from 8 grid points of

WRF/ERAI, after testing the various measures of Regional Frequency Analysis, were then used to derive the IDF curves. Chicago Design Storm method was then applied to derive the design storms of 50- and 100-year return periods.

6.4 Flood Analysis and Mapping over Greater Jakarta, Indonesia

With improved SRTM DEM and designed storms, MIKE 21 FM model was set up for Greater Jakarta. Passini method was used to compute the Time of Concentration (TC) of the upstream and downstream subcatchments; these TC values determine the design storms. Different durations of design storms were considered. Different mesh sizes were applied to the model domain for its efficient computational demands.

Flood maps of 2 return periods (50- and 100-years), for demonstration purposes, were generated and were compared with the flood footprints from 2013 Jakarta flooding. These flood maps give the necessary information on the flood extents and flood depths in the catchment. This information will then be able to use to take necessary and appropriate flood mitigation measures.

6.5 Recommendations for Future Study

- **Application of ANN DEM improvement scheme:** Conduct more tests on other satellite data of higher resolutions and accuracy, such as AW3D (Advanced World 3D map; 0.5 – 2 m resolution; 0.5 – 2m accuracy vertically; USD 95 – 190/km^2; minimum purchase of 25 km^2) whether the proposed ANN DEM improvement scheme can further improve their raw products. When significant improvement is noted, one clear benefit is that the trained ANN can be used to produce high accuracy DEM elsewhere at relatively very low cost.
- **Application of downscaled climate data with Regional Climate Model:** Conduct more tests of IDF curves derived from WRF/ERAI how well they compare with other areas' IDF curves obtained from high quality rain gauge station data. With rapid urbanization taking place in Southeast Asia, finding reliable rainfall proxies is of great importance. The downscaled rainfall data resulting from WRF/ERAI could be considered as reliable rainfall proxies.

Bibliography

[1] Abburu, S., and S. Golla (2015), Satellite Image Classification Methods and Techniques: A Review, *International Journal of Computer Applications*, *119*, 20-25.

[2] Abily, M., O. Delestre, L. Amossé, N. Bertrand, Y. Richet, C.-M. Duluc, P. Gourbesville, and P. Navaro (2015), Uncertainty related to high resolution topographic data use for flood event modeling over urban areas: toward a sensitivity analysis approach, *ESAIM: Proceedings and Surveys*, *48*, 385-399.

[3] Allen, M. R., V. R. Barros, J. Broome, W. Cramer, R. Christ, J. A. Church, L. Clarke, Q. Dahe, P. Dasgupta, and N. K. Dubash (2014), IPCC fifth assessment synthesis report-climate change 2014 synthesis report.

[4] Andersen, O. B., P. L. Woodworth, and R. A. Flather (1995), Intercomparison of recent ocean tide models, *Journal of Geophysical Research: Oceans*, *100*(C12), 25261-25282.

[5] Andres, L., W. A. Salas, and D. Skole (1994), Fourier analysis of multi-temporal AVHRR data applied to a land cover classification, *Remote Sensing*, *15*(5), 1115-1121.

[6] Andresen, S. L. (2002), John McCarthy: father of AI, *IEEE Intelligent Systems*, *17*(5), 84-85.

[7] Asharyanto, H., B. Soeksmantono, and K. Wikantika. (2015). *Three Dimensional City Building Modelling With Lidar Data (Case Study: Ciwaruga, Bandung),* . Paper presented at the INA-Rxiv. https://www.researchgate.net/publication/326135657_THREE_DIMENSIONAL_CITY_BUILDING_MODELLINGWITH_LIDAR_DATA_CASE_STUDY_CIWARUGA_BANDUNG

[8] Ashish, D., R. W. McClendon, and G. Hoogenboom (2009), Land-use classification of multispectral aerial images using artificial neural networks, *International Journal of Remote Sensing*, *30*(8), 1989-2004.

[9] Axelberg, P. (2007), *On tracing flicker sources and classification of voltage disturbances*, Department of Signals and Systems, Chalmers University of Technology.

[10] Bangkokpost (2012), edited, https://www.bangkokpost.com/.

[11] Bartosh, C. D. (2012), Integrating Land Survey Data into Measurement-Based GIS: An Assessment of Challenges snd Practical Solutions for Surveyors in Texas, University of Southern California.

[12] Benestad, R. E., I. Hanssen-Bauer, and D. Chen (2008), *Empirical-Statistical Downscaling*, 228 pp., WORLD SCIENTIFIC, https://doi.org/10.1142/6908.

[13] Bernard, M. M. (1932), Formulas For Rainfall Intensities of Long Duration, *Transactions of the American Society of Civil Engineers*, *96(1)*, 592-606.

[14] Bevere, L., R. Enz, J. Mehlhorn, and T. Tamura (2012), *Natural catastrophes and man-made disasters in 2011: historic losses surface from record earthquakes and floods*, Swiss Reinsurance Company Ltd.

[15] Bhoochaoom, A., and A. Dixon (2012), Thai flood 2011: rapid assessment for resilient recovery and reconstruction planning, Washington, DC, World Bank.

[16] Bladé Castellet, E., M. Valentin, Dolz, J. Aragón-Hernández, G. Corestein Poupeau, and M. Juny (2012), *Integration of 1D and 2D finite volume schemes for computations of water flow in natural channels*.

[17] Burlando, P., and R. Rosso (2002), Effects of transient climate change on basin hydrology. 1. Precipitation scenarios for the Arno River, central Italy, *Hydrological Processes*, *16*(6), 1151-1175.

[18] Casas, A., G. Benito, V. R. Thorndycraft, and M. Rico (2006), The topographic data source of digital terrain models as a key element in the accuracy of hydraulic flood modelling, *Earth Surface Processes and Landforms*, *31*(4), 444-456.

[19] Coles, S. (2001), *An introduction to statistical modeling of extreme values,* , Springer, 10.1007/978-1-4471-3675-0 doi.

[20] Cunnane, C. (1978), Unbiased plotting positions — A review, *Journal of Hydrology*, *37*(3), 205-222.

[21] Daniels, A. E., J. F. Morrison, L. A. Joyce, N. L. Crookston, S. C. Chen, and S. G. McNulty (2012), Climate projections FAQ, 32 pp, U.S. Department of Agriculture, https://doi.org/10.2737/RMRS-GTR-277.

[22] Das, S., and C. Cunnane (2012), Performance of flood frequency pooling analysis in a low CV context, *Hydrological Sciences Journal*, *57*(3), 433-444.

[23] Dee, D., d. Rosnay P, Uppala, Simmons, Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. Balmaseda, G. Balsamo, Bauer, Bechtold, Beljaars, v. de Berg, J. Bidlot, N. Bormann, Delsol, R. Dragani, M. Fuentes, and F. Vitart (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Quarterly Journal of the Royal Meteorological Society*, *137*, 553–597.

[24] Deltares (2019), SOBEK Hydrodynamics, Rainfall Runoff and Real Time Control.

[25] DHI (2014), MIKE 21 Flow Model FM Parallelisation using GPU.

[26] DHI (2017), MIKE FLOOD.

[27] Dimitriadis, P., A. Tegos, A. Oikonomou, V. Pagana, A. Koukouvinos, N. Mamassis, D. Koutsoyiannis, and A. Efstratiadis (2016), Comparative evaluation of 1D and quasi-2D hydraulic models based on benchmark and real-world applications for uncertainty assessment in flood mapping, *Journal of Hydrology*, *534*, 478-492.

[28] Drusch, M., U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, and P. Martimort (2012), Sentinel-2: ESA's optical high-resolution mission for GMES operational services, *Remote sensing of Environment*, *120*, 25-36.

[29] Duong, V. N., and P. Gourbesville (2016), Model Uncertainty in Flood Modelling. Case Study at Vu Gia Thu Bon Catchment - Vietnam, *Procedia Engineering*, *154*, 450-458.

[30] Elsebaie, I. H. (2012), Developing rainfall intensity–duration–frequency relationship for two regions in Saudi Arabia, *Journal of King Saud University - Engineering Sciences*, *24*(2), 131-140.

[31] Entwistle, N., G. Heritage, and D. Milan (2018), Recent remote sensing applications for hydro and morphodynamic monitoring and modelling, *Earth Surface Processes and Landforms*, *43*(10), 2283-2291.

[32] ESRI (2018), Environmental Systems Research Institute.

[33] Fatichi, S., V. Y. Ivanov, and E. Caporali (2011), Simulation of future climate scenarios with a weather generator, *Advances in Water Resources*, *34*(4), 448-467.

[34] Feser, F., B. Rockel, H. v. Storch, J. Winterfeldt, and M. Zahn (2011), Regional Climate Models Add Value to Global Model Data: A Review and Selected Examples, *Bulletin of the American Meteorological Society*, *92*(9), 1181-1192.

[35] Floodlist. (2008).

[36] Foody, G. M. (1999), The significance of border training patterns in classification by a feedforward neural network using back propagation learning, *International Journal of Remote Sensing*, *20*(18), 3549-3562.

[37] Fowler, H. J., S. Blenkinsop, and C. Tebaldi (2007), Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling, *International Journal of Climatology*, *27*(12), 1547-1578.

[38] François, K. (2001), *Modelling topography with SAR interferometry: Illustrations of a favourable and less favourable environment*, 1039-1050 pp.

[39] Funk, C., P. Peterson, M. Landsfeld, D. Pedreros, J. Verdin, S. Shukla, G. Husak, J. Rowland, L. Harrison, A. Hoell, and J. Michaelsen (2015), The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes, *Scientific Data*, *2*, 150066.

[40] Gao, B.-c. (1996), NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space, *Remote Sensing of Environment*, *58*(3), 257-266.

[41] Gatti, A., and A. Bertolini (2018), Sentinel-2 products specification document, https://earth.esa.int/documents/247904/685211/Sentinel-2+Products+Specification+Document.(2-PDGS-TAS-DI-PSD).

[42] 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.

[43] Goetz, A. F., G. Vane, J. E. Solomon, and B. N. Rock (1985), Imaging spectrometry for earth remote sensing, *science*, *228*(4704), 1147-1153.

[44] Gourbesville, P., M. Abily, O. Delestre, L. Amossé, N. Bertrand, Y. Richet, C.-M. Duluc, and P. Navaro (2015), Uncertainty related to high resolution topographic data use for flood

event modeling over urban areas: toward a sensitivity analysis approach, ESAIM Proceedings.

[45] Graf, L., M. Moreno-de-las-Heras, M. Ruiz, A. Calsamiglia, J. García-Comendador, J. Fortesa, J. López-Tarazón, and J. Estrany (2018), Accuracy Assessment of Digital Terrain Model Dataset Sources for Hydrogeomorphological Modelling in Small Mediterranean Catchments, *Remote Sensing*, *10*(12), 2014.

[46] Green, K., and M. Jackson (2009), Timeline of key developments in platforms and sensors for Earth observations, *American Society for Photogrammetry and Remote Sensing (ASPRS)*, *1.1*, 1-44.

[47] Gringorten, I. I. (1963), A plotting rule for extreme probability paper, *Journal of Geophysical Research (1896-1977)*, *68*(3), 813-814.

[48] Gruber, A., B. Wessel, M. Huber, and A. Roth (2012), Operational TanDEM-X DEM calibration and first validation results, *Photogrammetric Engineering & Remote Sensing*, *73*, 39-49.

[49] Gupta, R. P. (2018), *Remote Sensing Geology*, Springer, Berlin, Heidelberg.

[50] Gurney, K. (2014), *An introduction to neural networks*, CRC press, https://www.inf.ed.ac.uk/teaching/courses/nlu/assets/reading/Gurney_et_al.pdf.

[51] Guth, P. (2010), Geomorphometric comparison of ASTER GDEM and SRTM, paper presented at A special joint symposium of ISPRS Technical Commission IV & AutoCarto in conjunction with ASPRS/CaGIS.

[52] Hajnsek, I., T. Busche, H. Fiedler, G. Krieger, S. Buckreuss, M. Zink, A. Moreira, B. Wessel, A. Roth, and T. Fritz (2010), TanDEM-X Ground Segment TD-PD-PL-0069 TanDEM-X Science Plan, DLR.

[53] Haykin, S. (1994), *Neural Networks: A Comprehensive Foundation*, 768 pp., Prentice Hall PTR.

[54] Haylock, M. R., T. C. Peterson, L. M. Alves, T. Ambrizzi, Y. M. T. Anunciação, J. Baez, V. R. Barros, M. A. Berlato, M. Bidegain, G. Coronel, V. Corradi, V. J. Garcia, A. M. Grimm, D. Karoly, J. A. Marengo, M. B. Marino, D. F. Moncunill, D. Nechet, J. Quintana, E. Rebello, M. Rusticucci, J. L. Santos, I. Trebejo, and L. A. Vincent (2006), Trends in Total and Extreme South American Rainfall in 1960–2000 and Links with Sea Surface Temperature, *Journal of Climate*, *19*(8), 1490-1512.

[55] Hensley, S., P. Rosen, and E. Gurrola (2000), The SRTM topographic mapping processor, paper presented at IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium. Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment. Proceedings (Cat. No.00CH37120), 24-28 July 2000.

[56] Hervouet, J. (2007), Applications, *Hydrodynamics of Free Surface Flows*.

[57] Hewitson, B. C., and R. G. Crane (1996), Climate downscaling: techniques and application, *Climate Research*, *07*(2), 85-95.

[58] Hosking, J. R. M., and J. R. Wallis (1993), Some statistics useful in regional frequency analysis, *Water Resources Research*, *29(2)*(271-281).

[59] Hosking, J. R. M., and J. R. Wallis (1997), *Regional Frequency Analysis: An Approach Based on L-Moments*, Cambridge University Press, Cambridge.

[60] Hundecha, Y., B. Arheimer, C. Donnelly, and I. Pechlivanidis (2016), A regional parameter estimation scheme for a pan-European multi-basin model, *Journal of Hydrology: Regional Studies*, *6*, 90-111.

[61] III. (2018). Facts + Statistics: Global catastrophes. Retrieved from https://www.iii.org/fact-statistic/facts-statistics-global-catastrophes

[62] InaSafe. (2015). A flood in Jakarta like 2013, http://inasafe.org/. Retrieved from http://inasafe.org

[63] Jacobs, S. A., S. D. Taverna, Y. Zhang, S. D. Briggs, J. Li, J. C. Eissenberg, C. D. Allis, and S. Khorasanizadeh (2001), Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3, *The EMBO Journal*, *20*(18), 5232-5241.

[64] JakartaOpenData (2015), Data Jumlah Penduduk DKI Jakarta, edited, Pemerintah Provinsi DKI Jakarta, https://twitter.com/jakartaopendata?lang=en.

[65] Kawabata, D., and J. Bandibas (2009), Landslide susceptibility mapping using geological data, a DEM from ASTER images and an Artificial Neural Network (ANN), *Geomorphology*, *113*(1), 97-109.

[66] Keifer, C. J., and H. H. Chu (1957), Synthetic storm pattern for drainage design, *Journal of the hydraulics division*, *83(4)*, 1-25.

[67] Kim, D., P. Gourbesville, and S.-Y. Liong (2019), Overcoming data scarcity in flood hazard assessment using remote sensing and artificial neural network, *Smart Water*, *4*(1), 2.

[68] Kim, D., Y. Sun, D. Wendi, Z. Jiang, S.-Y. Liong, and P. Gourbesville (2018), Flood modelling framework for Kuching City, Malaysia: overcoming the lack of data, in *Advances in Hydroinformatics*, edited, pp. 559-568, Springer.

[69] Koutsoyiannis, D., D. Kozonis, and A. Manetas (1998), A mathematical framework for studying rainfall intensity-duration-frequency relationships, *Journal of Hydrology*, *206*(1), 118-135.

[70] Krieger, G., I. Hajnsek, P. Lopez Dekker, S. Baumgartner, M. Rodriguez-Cassola, F. De Zan, P. Prats-Iraola, K. Papathanassiou, M. Zink, and A. Moreira (2011), *TanDEM-X Scientific Results and Future Formation Flying SAR Missions*, 1-10 pp.

[71] Kugler, Z. (2012), Remote sensing for natural hazard mitigation and climate change impact assessment, *Quarterly Journal of the Hungarian Meteorological Service*, *116*, 21-38.

[72] Kumar, G., and P. U. Iyer (2010), Management of perioperative low cardiac output state without extracorporeal life support: What is feasible?, *Annals of pediatric cardiology*, *3*(2), 147-158.

[73] Levenberg, K. (1944), A Method for the Solution of Certain Non-Linear Problems in Least Squares, *Quarterly of Applied Mathematics*, *2*(2), 164-168.

[74] Li, P., C. Shi, Z. Li, J. P. Muller, J. Drummond, X. Li, T. Li, Y. Li, and J. Liu (2012), *Evaluation of ASTER GEM Ver2 using GPS measurements and SRTM Ver4.1 in China*, 181-186 pp.

[75] Lillesand, T., R. W. Kiefer, and J. Chipman (2015), *Remote Sensing and Image Interpretation*, 7th Edition ed., Wiley.

[76] Liu, J. (2017), Flood Projection and Analysis Through Stochastic Downscaling, PhD thesis, National University of Singapore, Singapore.

[77] Luk, K. C., J. E. Ball, and A. Sharma (2000), A study of optimal model lag and spatial inputs to artificial neural network for rainfall forecasting, *Journal of Hydrology*, *227*(1), 56-65.

[78] Mackay, C., S. Suter, N. Albert, S. Morton, and K. Yamagata. (2015). *Large scale flexible mesh 2D modelling of the Lower Namoi Valley*. Paper presented at the Floodplain Conference.

[79] Maraun, D., F. Wetterhall, A. M. Ireson, R. E. Chandler, E. J. Kendon, M. Widmann, S. Brienen, H. W. Rust, T. Sauter, M. Themeßl, V. K. C. Venema, K. P. Chun, C. M. Goodess, R. G. Jones, C. Onof, M. Vrac, and I. Thiele-Eich (2010), Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, *Reviews of Geophysics*, *48*(3).

[80] Marquardt, D. W. (1963), An Algorithm for Least-Squares Estimation of Nonlinear Parameters, *Journal of the Society for Industrial and Applied Mathematics*, *11*(2), 431-441.

[81] Martone, M., P. Rizzoli, C. Wecklich, C. González, J.-L. Bueso-Bello, P. Valdo, D. Schulze, M. Zink, G. Krieger, and A. Moreira (2018), The global forest/non-forest map from TanDEM-X interferometric SAR data, *Remote Sensing of Environment*, *205*, 352-373.

[82] McCarthy, J. (1956), The inversion of functions defined by Turing machines, *Automata studies*, 177-181.

[83] McGahey, C., and P. Samuels. (2004). *A Practical Approach to Uncertainty in Conveyance Estimation*. Paper presented at the 39th DEFRA Flood and Coastal Management Conference, London.

[84] Mirosław-Świątek, D., S. Szporak-Wasilewska, R. Michałowski, I. Kardel, and M. Grygoruk (2016), *Developing an algorithm for enhancement of a digital terrain model for a densely vegetated floodplain wetland*, 1-16, 16 pp., SPIE.

[85] Moody, D. I., S. P. Brumby, J. C. Rowland, and G. L. Altmann (2014), *Land cover classification in multispectral imagery using clustering of sparse approximations over learned feature dictionaries*, 1-19, 19 pp., SPIE.

[86] Navalgund, R. R., V. Jayaraman, and P. S. Roy (2007), Remote sensing applications: an overview, *Current Science*, *93 (12)*, 1747-1766.

[87] Neal, J. C., T. J. Fewtrell, P. D. Bates, and N. G. Wright (2010), A comparison of three parallelisation methods for 2D flood inundation models, *Environmental Modelling & Software*, *25*(4), 398-411.

[88] Neelz, S., and G. Pender (2009), Grid resolution dependency in inundation modelling, edited, pp. 109-117, Tayler & Francis Group.

[89] Nelli, N. R., V. R. Madineni, B. Ghouse, and R. Varaha (2018), Cloud vertical structure over a tropical station obtained using long-term high-resolution radiosonde measurements, *Atmospheric Chemistry and Physics*, *18*(16), 11709-11727.

[90] Ngongondo, C. S., C.-Y. Xu, L. M. Tallaksen, B. Alemaw, and T. Chirwa (2011), Regional frequency analysis of rainfall extremes in Southern Malawi using the index rainfall and L-moments approaches, *Stochastic Environmental Research and Risk Assessment*, *25*(7), 939-955.

[91] Nhat, L. M., Y. Tachikawa, and K. Takara. (2006). *Establishment of Intensity-Duration-Frequency Curves for Precipitation in the Monsoon Area of Vietnam*. Paper presented at the Annuals of Disaster.

[92] Nunez, J., K. Verbist, J. R Wallis, M. Schaefer, L. Morales-Salinas, and W. Cornelis (2011), Regional frequency analysis for mapping drought events in north-central Chile, *Journal of Hydrology*, *405*, 352-366.

[93] Pande, D. C., D. K. Moharir, and S. Khadri (2018), *Study of land use classification in an arid region using multispectral satellite images*, 1-11 pp., Applied Water Science.

[94] Pender, G. (2006), Briefing: Introducing the Flood Risk Management Research Consortium, *Proceedings of the Institution of Civil Engineers - Water Management*, *159*(1), 3-8.

[95] Prinadiastari, I., and P. Bahri (2018), *Quick Assessment of Padangsidimpuan's Flash Flood, North Sumatera, Indonesia Using 2D Debris Flow Simulation SIMLAR v2.1.*

[96] PUB (2012), Code of Practice 35, *Public Utility Board.*

[97] Radiomobile. (2018). Background on DEM.

[98] Renaldi, E., and T. Shelton (2018), Thousands killed as Indonesia devastated by nearly 2,000 natural disasters in 2018, in *ABC, https://www.abc.net.au/news/2018-12-28/the-five-most-deadly-natural-disasters-in-indonesia-this-year/10668480*, edited.

[99] Rencz, A. N., C. Bowie, and B. C. Ward (1996), Application of thermal imagery from LANDSAT data to locate kimberlites, Lac de Gras area, district of Mackenzie, N.W.T., *Geological Survey of Canada*, *3228*, 255-257.

[100] Reuter, H. I., A. Nelson, and A. Jarvis (2007), An evaluation of void-filling interpolation methods for SRTM data, *International Journal of Geographical Information Science*, *21*(9), 983-1008.

[101] Reuter, H. I., A. Neison, P. Strobl, W. Mehl, and A. Jarvis (2009), A first assessment of Aster GDEM tiles for absolute accuracy, relative accuracy and terrain parameters, paper

presented at 2009 IEEE International Geoscience and Remote Sensing Symposium, 12-17 July 2009.

[102] Robert, W., S. Charles, E. Zorita, B. Timbal, P. Whetton, and L. Mearns (2004), *Guidelines For Use of Climate Scenarios Developed From Statistical Downscaling Methods*.

[103] Rodriguez, E., C. S. Morris, and J. E. Belz (2006), A global assessment of the SRTM performance, *Photogrammetric Engineering & Remote Sensing*, *72*(3), 249-260.

[104] Rosen, P. A., S. Hensley, E. Gurrola, F. Rogez, S. Chan, J. Martin, and E. Rodriguez (2001), SRTM C-band topographic data: quality assessments and calibration activities, paper presented at IGARSS 2001. Scanning the Present and Resolving the Future. Proceedings. IEEE 2001 International Geoscience and Remote Sensing Symposium (Cat. No. 01CH37217), IEEE.

[105] Rosenblatt, F. (1961), Principles of neurodynamics. perceptrons and the theory of brain mechanisms, *Cornell Aeronautical Lab Inc Buffalo NY*, *VG-1196-G-8*.

[106] Rossman, L. A. (2010), *Storm Water Management Model User's Manual Version 5.0*, EPA (United States Environmental Protection Agency).

[107] Roy, D. P., J. Li, H. K. Zhang, and L. Yan (2016), Best practices for the reprojection and resampling of Sentinel-2 Multi Spectral Instrument Level 1C data, *Remote sensing letters*, *7*(11), 1023-1032.

[108] Roy, D. P., J. Li, H. K. Zhang, L. Yan, H. Huang, and Z. Li (2017), Examination of Sentinel-2A multi-spectral instrument (MSI) reflectance anisotropy and the suitability of a general method to normalize MSI reflectance to nadir BRDF adjusted reflectance, *Remote Sensing of Environment*, *199*, 25-38.

[109] Sarve, A. N., M. N. Varma, and S. S. Sonawane (2015), Response surface optimization and artificial neural network modeling of biodiesel production from crude mahua (Madhuca indica) oil under supercritical ethanol conditions using CO2 as co-solvent, *RSC Advances*, *5*(85), 69702-69713.

[110] Schmidhuber, J. (2015), Deep learning in neural networks: An overview, *Neural Networks*, *61*, 85-117.

[111] Sherman, C. W. (1931), Frequency and Intensity of Excessive Rainfalls at Boston, *Transactions of the American Society of Civil Engineers*, *95(1)*, 951-960.

[112] Sibson, R. (1981), A Brief Description of Natural Neighbor Interpolation, *Interpreting Multivariate Data*, *In: Barnett, V., Ed*, 21-36.

[113] Sim, W. (2018), Earthquakes, rains, heatwave, typhoon: Japan's brutal summer of 2018, in *The Straitstimes*, edited.

[114] Soares, P. M. M., R. M. Cardoso, P. M. A. Miranda, P. Viterbo, and M. Belo-Pereira (2012), Assessment of the ENSEMBLES regional climate models in the representation of precipitation variability and extremes over Portugal, *Journal of Geophysical Research: Atmospheres*, *117*(D7).

[115] Sun, Y., D. Wendi, D. E. Kim, and S.-Y. Liong (2016), Technical note: Application of artificial neural networks in groundwater table forecasting-a case study in a Singapore swamp forest, *Hydrology and Earth System Sciences*, *20*, 1405-1412.

[116] Sun, Y., C. D. Doan, A. T. Dao, J. Liu, and S.-Y. Liong (2014), Improving numerical forecast accuracy with ensemble Kalman filter and chaos theory: Case study on Ciliwung river model, *Journal of Hydrology*, *512*, 540-548.

[117] Tachikawa, T., M. Hato, M. Kaku, and A. Iwasaki (2011), Characteristics of ASTER GDEM version 2, paper presented at 2011 IEEE International Geoscience and Remote Sensing Symposium, 24-29 July 2011.

[118] Thenkabail, P. S. (2015), *Remote Sensing Handbook - Three Volume Set*, CRC Press

[119] Thomas, J., and V. Prasannakumar (2015), Comparison of basin morphometry derived from topographic maps, ASTER and SRTM DEMs: an example from Kerala, India, *Geocarto International*, *30*(3), 346-364.

[120] UNISDR (2010), Synthesis report on ten ASEAN countries disaster risk assessment, United Nations Office for Disaster Risk Reduction, Geneva.

[121] USGS. (2000). United States Geological Survey.

[122] USGS (2003), SRTM Water Body Data Product Specific Guidance.

[123] Weiss, J., and P. Bernardara (2013), Comparison of local indices for regional frequency analysis with an application to extreme skew surges, *Water Resources Research*, *49*(5), 2940-2951.

[124] Wendi, D., S.-Y. Liong, Y. Sun, and C. D. Doan (2016), An innovative approach to improve SRTM DEM using multispectral imagery and artificial neural network, *Journal of Advances in Modeling Earth Systems*, *8*(2), 691-702.

[125] Wendleder, A., B. Wessel, A. Roth, M. Breunig, K. Martin, and S. Wagenbrenner (2013), TanDEM-X Water Indication Mask: Generation and First Evaluation Results, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *6*(1), 171-179.

[126] Wessel, B., T. Fritz, T. Busche, P. Rizzoli, and G. Krieger (2016), TanDEM-X Ground Segment DEM Products Specification Document, DLR.

[127] Widrow, B., and M. E. Hoff (1962), Associative Storage and Retrieval of Digital Information in Networks of Adaptive "Neurons", in *Biological Prototypes and Synthetic Systems: Volume 1 Proceedings of the Second Annual Bionics Symposium sponsored by Cornell University and the General Electric Company, Advanced Electronics Center, held at Cornell University, August 30–September 1, 1961*, edited by E. E. Bernard and M. R. Kare, pp. 160-160, Springer US, Boston, MA.

[128] Wilby, R. L., and C. W. Dawson (2013), The Statistical DownScaling Model: insights from one decade of application, *International Journal of Climatology*, *33*(7), 1707-1719.

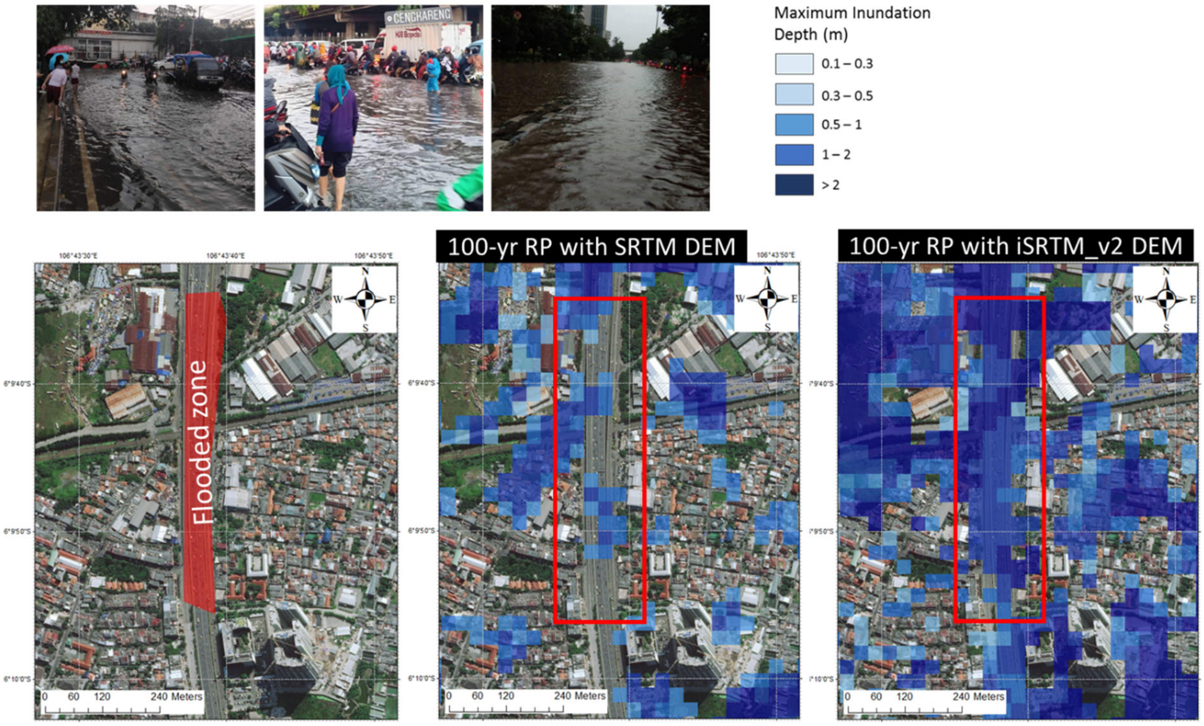
[129] Wulder, M. A., J. C. White, T. R. Loveland, C. E. Woodcock, A. S. Belward, W. B. Cohen, E. A. Fosnight, J. Shaw, J. G. Masek, and D. P. Roy (2016), The global Landsat archive: Status, consolidation, and direction, *Remote Sensing of Environment*, *185*, 271-283.

[130] Yusuf, A., and H. Francisco (2009), *Climate Change Vulnerability Mapping for Southeast Asia*, Economy and Environment Program for Southeast Asia (EEPSEA).

[131] Zhang, W., and D. R. Montgomery (1994), Digital elevation model grid size, landscape representation, and hydrologic simulations, *Water Resources Research*, *30*(4), 1019-1028.

Appendix

(1) Flooding at Cengkareng, West Jakarta (Apr 2019)



(2) Flooding at Cawang Soetoyo, East Jakarta (Apr 2019)