Leveraging Global and Local Data Sources for Flood Hazard Assessment and Mitigation: An Application of Machine Learning to Manila

Masahiko Haraguchi,

Department of Earth and Environmental Engineering, Columbia University, New York, NY, USA; The Belfer Center for Science and International Affairs, Harvard Kennedy School, Harvard University, Cambridge, MA, USA

Fabio Cian,

Department of Economics, University of Venice Ca' Foscari, Venice, 30121, Italy

Upmanu Lall,

Department of Earth and Environmental Engineering, Columbia University, New York, NY, USA

Abstract

The objective of this paper is to illustrate how developing countries with limited datasets and capacity can utilize global hazard data to support risk-informed decision-making at the local level. Using urban hydrologic models for flood risk assessment requires the collection of intensive data for model calibration, and even after such an effort leads to considerable uncertainty for spatially specific risk assessment. The case study in this paper examines flooding events in Metro Manila, which routinely experiences the disasters. We explore whether relationships between flood occurrence in Manila and remotely-sensed environmental data, such as rainfall amounts and vegetation moisture could be established using machine learning techniques such as visualization, decision tree, and logistic regression. The study demonstrated that a model that uses an appropriate rainfall index is better than one that uses only daily rainfall amounts or adds a vegetation index. The best predictive models are found to be: (i) one that uses rainfall type and rainfall amount; (ii) and one that integrates all the information including, rainfall amount, rainfall type and vegetation index. The case study demonstrates that globally available data used with machine learning techniques can be effective for local flood management.

1. Introduction

The Dartmouth Flood Observatory (2018) indicates that floods caused 62,000 deaths and displaced 610 million people in the world since 1985. Recent floods in urban areas caused substantial economic damage and losses (Haraguchi & Kim, 2016; Haraguchi & Lall, 2015). Urban flooding is a phenomenon caused by multiple factors such as large rainfall amount, flood-prone topography, inadequate infrastructure for drainage and storage and poor water management. High population densities and informal settlements in waterfront areas as well as poor infrastructure and drainage systems make flood risk management a significant challenge in mega cities in developing countries. Often, detailed hydrologic and stream gage data is unavailable, which makes it challenging to mitigate and respond to flood risks at a local scale[1] in developing countries (Haraguchi, 2018).

Satellite data can be a useful resource for decision-makers to address disaster risks, both for emergency management and for disaster risk reduction (DRR) planning. In the era of big Earth Observation (EO) data, many of which are freely accessible, high-resolution, and frequently acquired, the applications for DRR and disaster risk management (DRM) are numerous and of great importance especially for those countries with limited national resources (Twele, Cao, Plank, & Martinis, 2016).

We present a case study of urban flooding in Manila Philippines to show how global remote sensing (RS) technology and EO data can be used to assess local urban flooding in developing countries. The Philippines is a typhoon-prone country. An average of 20 out of 36 tropical cyclones that develop over the Northwest Pacific basin per year impact the country. Metro Manila in the Philippines has experienced numerous flooding incidents for many years. It experienced massive flooding in September 2009, which caused the deaths of 420 and 20,000 evacuees, with damage and losses equivalent to about 2.7 % of Gross Domestic Product (Stoutjesdijk, 2017). Innovative programs to improve the capacity to respond to the effect of extreme weather conditions are in place. For example, University of the Philippine's Nationwide Operational Assessment of Hazards (UP NOAH) Center has developed hazard and risk maps that visualize the impacts of hazard events as a dynamic way (Cadiz, 2018). Specifically, the objective of this research

[1] In this paper, a local scale means a scale that covers from a city to a metropolitan area.

was to 1) show how global satellite data sets can help local decision-makers addressing flood risk and 2) to examine the relationship between flood generation in Metro Manila in the Philippines and RS environmental data.

2. Remote Sensing and Its Applicability to Local Floods Detection

Satellite RS has been used for decades for disaster management. Its applicability has been demonstrated for near-real-time mapping (Brisco, Schmitt, Murnaghan, Kaya, & Roth, 2013; Cossu, Schoepfer, Bally, & Fusco, 2009; Henry, Chastanet, Fellah, & Desnos, 2006; Horritt, Mason, & Luckman, 2001; Martinis, Kersten, & Twele, 2015; Matgen, Schumann, Henry, Hoffmann, & Pfister, 2007), and operational emergency response (Bessis, Bequignon, & Mahmood, 2004; Mahmood, 2012). The European Copernicus Emergency Management Service[2], the International Charter on Space and Major Disaster[3], and the UNOSAT Rapid Mapping Service[4], provide institutional examples. Optical multi-spectral data, acquired by different sensors such as Landsat, Sentinel-2 and MODIS, are employed for flood mapping (Brakenridge & Anderson, 2006; Cian, Marconcini, & Ceccato, 2018; Heimhuber, Tulbure, & Broich, 2018; Huang, Peng, Lang, Yeo, & McCarty, 2014; Islam, Bala, & Haque, 2010; Kordelas, Manakos, Aragonés, Díaz-Delgado, & Bustamante, 2018; Qi et al., 2009; Ryu, Won, & Min, 2002; Thomas, Kingsford, Lu, & Hunter, 2011; Tong et al., 2018; Y. Wang, Colby, & Mulcahy, 2002) with the usual limitation that cloud coverage, very likely to happen during floods especially in tropical areas, can affect the detection of the flood extent. In contrast, Synthetic Aperture Radar (SAR) is a powerful tool in the context of flood mapping since it can provide frequent observations (Alsdorf, Rodríguez, & Lettenmaier, 2007; Cian et al., 2018; Mertes, 2002; Ward et al., 2014) thanks to the capability to monitor land in almost any weather conditions (Marzano, Mori, Weinman, & Montopoli, 2012; Schumann, Bates, Horritt, Matgen, & Pappenberger, 2009) and during night-time (Waisurasingha, Aniya, Hirano, Kamusoko, & Sommut, 2007; Wilson & Rashid, 2005).

Until a few years ago, the availability of RS data for disaster monitoring was limited due to the low time frequency (technical limitation of the sensors, limited storing capacity on board of satellites or limited capacity of data downlink) of data acquisition by satellites. Moreover, many methodologies use archived data, i.e., data acquired

[2] http://emergency.copernicus.eu/mapping
[3] https://disasterscharter.org
[4] https://unitar.org/unosat/unosat-rapid-mapping-service

before the event, to perform a change detection analysis by comparing them with the images acquired during the disaster. This allows one to detect the extent of the flood or to assess damages occurred during the event. For the same reason mentioned, archived data has limitations of detecting changes in many cases. The advent of new EO constellations, such as the European Space Agency's (ESA) Sentinels (with optical and radar satellite acquiring data globally every week at a resolution of 10 m), or commercial constellations of hundreds of small satellites acquiring daily at a global scale, such as Planet[5] (optical at 3 m) and ICEYE[6] (radar at 10 m), have started the era of EO big data (Huang et al., 2014; Ma et al., 2015; Yang, Huang, Li, Liu, & Hu, 2017) promising a new era for disaster monitoring and EO data exploitation. These new sources of data increase the probability of observing the evolution of a disaster given the higher frequency of observation and the higher resolution.

RS can also be used to assess the impact caused by a flood. An event needs to be characterized by its extent, depth, and duration. The extent of events can be retrieved relatively with ease; a challenge is to extract its depth and duration. However, RS is regularly employed for post-disaster impact assessment (Haq, Akhtar, Muhammad, Paras, & Rahmatullah, 2012; Klemas, 2009; Van Westen, 2013), using optical sensors (Van der Sande, De Jong, & De Roo, 2003; Q. Wang, Watanabe, Hayashi, & Murakami, 2003; Yamagata & Akiyama, 1988) and radar sensors (Bhatt et al., 2017; Serpico et al., 2012) in an agricultural context (Kwak, Arifuzzanman, & Iwami, 2015; Pantaleoni, Engel, & Johannsen, 2007; Tapia-Silva, Itzerott, Foerster, Kuhlmann, & Kreibich, 2011) and in a socio-economic perspective (Oddo, Ahamed, & Bolten, 2018). These analyses allow to estimate the impacts of disasters on crops and critical infrastructures as well as to determine the number of people affected and the number of damaged buildings.

However, for the monitoring of flood extent in urban areas, some limitations remain. SAR data, due to their physical characteristics, are not ideal for monitoring urban flooding. For example, small areas in between buildings are difficult to detect as flooded using low (50-by-50 meter per a pixel) and middle-resolution (20–30 meter) images, because areas of interests are too small compared to the image resolution. High-resolution images also have difficulties due to the physical nature of the radar signal. It is hard to make a distinction between smooth surfaces of

[5] https://www.planet.com
[6] https://www.iceye.com

streets with smooth surfaces of water or tall buildings close to each other, which prevents the radar signal from sensing the water surface in between them. Optical data in the absence of cloud would allow monitoring urban flooding in case of high resolution. However, high-resolution (below 3–5 meter) data are expensive, and the significant limitation of cloud cover remains, which can prevent the observation of the event itself (Rahman & Di, 2017). Table 1 gives an overview of some of the existing EO satellite missions useful in the frame of disaster monitoring and post-disaster assessment.

Table 1: Existing EO satellite missions useful for disaster monitoring and post-disaster assessment.

Dataset	Characteristics	Application	Pros	Cons
Sentinel-1	SAR / High Res (10 m) / 6 days repeating cycle / free Available from 2014	Flood Mapping / damage assessment / land subsidence	High resolution / frequent acquisition globally (6 days) / free	
Sentinel-2	Optical Multispectral / High Res (10 m) / 5 days repeating cycle / free Available from 2015	Flood mapping / land cover / vegetation monitoring	High resolution / frequent acquisition globally (5 days) / free	
Landsat	Optical (Multispectral) / medium resolution (30 m) / free Available from 1984	Flood mapping / land cover / vegetation monitoring	Historical archive (since 1984)	Not frequent acquisition in the past / not high resolution
MODIS	Optical Multispectral / low resolution (250 m to 1 km) / Daily coverage / Free Available from 1999	Flood mapping / land cover / vegetation monitoring	Daily coverage	Low resolution
TerraSAR-X	SAR / High Res (<1 m) Available from 2007	Flood Mapping / damage assessment / land subsidence	Very high resolution	Not free / not frequent acquisition in certain countries / limited spatial coverage at high resolution
COSMO-SkyMed	SAR / High Res (<1 m) Available from 2007	Flood Mapping / damage assessment / land subsidence	Very high resolution / Frequent acquisition in many locations also in developing countries	Not free
Planet	Optical / High Res (3 m) / Daily coverage Available from 2014	Flood mapping / land cover / vegetation monitoring	The very high resolution, daily coverage	Not free
Iceye	SAR / High Res (3 m) / Daily coverage Available from 2018	Flood Mapping / damage assessment / land subsidence	The very high resolution, daily coverage	Not free

RS is useful not only in the response and recovery phases of DM, but also mitigation and preparedness phases. RS can support flood risk assessment monitoring rainfall by measuring river discharge and flood depth to

forecast flooding and assessing vulnerability. An exhaustive review of the applications of RS in flood management in all its phases is provided by Rahman & Di (2017). Precipitation measurement is useful in the pre-event phase and during the event. Several datasets are available based on different satellite mission and technology. Table 2 provides a summary of some of the existing datasets.

Table 2: Precipitation dataset which can be used for DRM

Dataset	Description	Characteristics	Citation
CHIRPS (Climate Hazards Group InfraRed Precipitation with Station data)	quasi-global rainfall dataset; gridded rainfall time-series obtained from infrared satellite data and in-situ station data	0.05° resolution, daily and 5-day composite, for trend and seasonal analysis, public, from Jan 1981 to present	Funk et al. (2015)
TRMM (Tropical Rainfall Measuring Mission)	Designed to monitor and study tropical rainfall; gridded precipitation estimate obtained from infrared satellite measurements	0.25° spatial resolution, 3-hour or 1-month temporal resolution, public, from Jan 1998 to present	Adler et al. (2003)
GPM (Global Precipitation Measurement)	Estimates based on the Integrated Multi-satellitE Retrievals for GPM (IMERG) algorithm, using passive-microwave and microwave-calibrated infrared instruments and gauge analyses.	0.1° resolution, 30 minutes, public, available from March 2014 to present	NASA PMM
GSMaP Operational (Global Satellite Mapping of Precipitation)	Estimation using a radiative transfer model with multi-band passive microwaves and infrared radiometers of the GPM Core Observatory satellite. Estimation is gauge-adjusted using NOAA/CPC gauge measurement	0.1° resolution, 1 hour, global, public, from March 2014 to present	Kubota et al. (2007)
PERSIANN-CDR (Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks-Climate Data Record)	Estimation produced using Gridded Satellite infrared data. Produced quarterly with a lag of three months.	0.25 ° resolution, daily, quasi-global, public, from Jan 1983 to present	Ashouri et al. (2015)

Finally, the advances in big data computation provided by Google Earth Engine, open up new possibilities in image processing and data analysis. The deluge of data freely available, which are already contained in the Engine, allows efficient processing without the need to be downloaded, thanks to an intuitive and user-friendly interface, and cloud computing capabilities offered by Google. Limitations of computing power, RAM memory and storage space are

now becoming part of the past. The world of RS has taken a revolutionary path at an unprecedented speed (Gorelick et al., 2017), which can be exploited for policy design, implementation and decision-making processes for DRM.

3. Case Study - Study Design

A case study of using RS and EO for urban flood management in Manila, Philippines is presented in this section. Several, previous studies discuss flooding in Manila. Bankoff (2003) examines the recorded history of floods in Manila. Flooding has been a part of daily life in Manila since at least the 19th century (Bankoff, 2003). Rapid urbanization magnified the extent of flooding in the metropolitan area. By the 1960s, almost 70% of the city was prone to floods with a depth from 3.6 to 4.5 meters (Bankoff, 2003). By the 1970s, massive floods disrupted essential lifeline services and forced many people to evacuate (Bankoff, 2003). Floodwater depths continued to rise until the 1990s. At present, around 20% of the city is classified as flood-prone (Bankoff, 2003). The root causes of vulnerability to floods in Manila lies in the dynamics between stakeholders, ecosystem, and natural hazards (Bankoff, 2003; Porio, 2011). Few studies, however, point out how to estimate actual precipitation amounts that lead to flooding in a rapid way, such as using RS. Garcia, Retamar, and Javier (2015) deploys a real-time urban flood monitoring system using Random Forest algorithms, but the paper has limited area coverages in Manila. Tablazon et al. (2015) developed a storm surge inundation map based on simulated historical tropical cyclones and a storm surge model. The produced maps show that Metro Manila is vulnerable to storm-surge with a flood depth up to 4 m and extent of up to 6.5 km from the coastline. A research question to be addressed here is the following:

How can floods be robustly identified given a lack of ground based data from stream gages or other local sensors?

Specifically, using machine learning techniques[7], this research attempts to answer the following research questions:

[7] Machine learning is a method of data analysis that systems can learn from data, identify patterns, and make predictions with minimal human explicit instructions. Specifically, we use supervised learning methods (Friedman et al., 2001).

a. *Can relationships between flooding in Manila and RS data such as rainfall and vegetation prior to or during the flood be established, so that the proxy RS information is effective for flood state assessment?*

b. *If so, can these be used to construct adequate models to predict floods in Manila?*

We explore the development of a model to predict and prepare for floods using machine learning techniques. First, the study looks at a time series of rainfall data and conducts an initial analysis with a baseline model. The study measures the intensity, volume, and duration of rainfall events. As the first analysis, the study used machine learning techniques such as visualization and a decision tree to improve the baseline model and find better models to predict floods in Manila, Philippines. Then, the models incorporate types of rainfall events and a vegetation index. The model improvement will be measured by the following indicators (Table 3):

Table 3: Measurement of model performances

	Actual Positive (=Flood occurred)	Actual Negative (=Flood did not occur)
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (TN)

Precision= TP/(TP+FP)

Sensitivity= TP/(TP+FN)

Specificity= TN/(FP+TN)

Accuracy= (TP+TN)/(TP+FP+FN+TN)

Since the accuracy is considered most comprehensive, the study will most value this indicator. The second analysis of the study conducted cross-validated ridge regression to evaluate models. The models are evaluated through the AIC, BIC, and ROC curve. The whole process is summarized in Table 4.

Table 4: Proposed steps and the application to our case study in Manila.

Proposed steps	**Examples in our case study in Manila, Philippines**
1. Initial Analysis	
1.1. Identify flood events in a specified area.	Specified the latitudes and longitudes of Manila and identified flood events from database from the Dartmouth Flood Observatory.
1.2. Assess if the satellite products can be used	CMORPH is identified based on measurement errors
1.3. Select vegetation index as proxy for vegetation moisture	NDVI is identified
2. Explanatory Analysis	
2.1. Identify a cut off value	Identified 0.5 mm / hour
2.2. Identify a rainfall event	Defined a rainfall event as 1 successive day of rainfall for one rainfall event
2.3. Calculate the volume and duration of each rainfall event	
2.4. Determine a baseline threshold	Rainfall intensity: 1.5 mm per hour Rainfall volume: 20 mm per event Rainfall duration: 2 days per event
3. Model Assessment, Selection and Evaluation	
3.1. Select predictors	4 models are built with the following selected predictors: Model 1 (Baseline): 1.5mm/hour of rainfall intensity Model 2: Adding the type of rainfall to the baseline Model3: Adding NDVI to the baseline Model4: Adding type of rainfall and NDVI to the baseline
3.2. Select methods	1. Decision trees and visualization 2. Cross-validated ridge regression
3.3. Evaluate models	1. Precision, sensitivity, specificity, accuracy (ROC curve) 2. AIC and BIC

4. Data, Background Information, and Preparation of the Analysis

The first attempt was to examine an accurate rainfall data that is observed at the ground level. However, nineteen out of twenty-two ground stations did not have a complete time series of data. Thus, a ground-data-based approach, even using a hydrologic model lacks sufficient data to construct a model to prepare for floods. Thus, as a proxy for rainfall, we used gridded satellite rainfall estimates, which are publicly available online. Additionally, vegetation indices, which are a proxy for the state of ground moisture, were used as additional information. The reason for this is that antecedent soil moisture also influences the occurrence of floods. For example, when the ground retains moisture from preceding rainfall events, its capacity to hold water is exceeded by new rainfall. Therefore, a series of rainfall events can lead to floods even if the event rainfall is not extreme. We look at the indices in the following way:

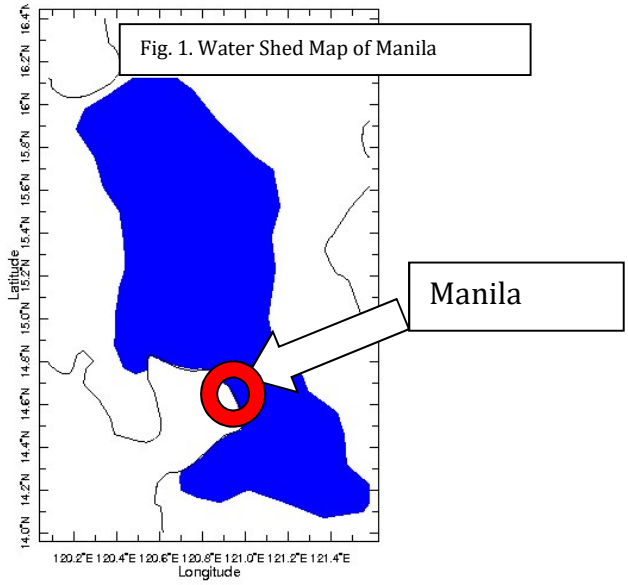
Dependent variable (DV): flood occurrence (binary)

Independent variables (IV): precipitation intensity, volume, and duration (numerical); type of rainfall (binary: tropical cyclone or not); and vegetation index (numerical).

Specifically, data are taken from the following various sources.

Spatial Domain

This project considers a watershed that includes Metro Manila because of the assumption that the flooding is induced by precipitation that has fallen in the larger watershed. The spatial domain is 14 Degree of North to 16.4 Degree of North and 120.0833 Degree of East to 121.5833 Degree of East (Fig. 1).



Floods data (Dependent variable)

Data of flood events were retrieved from the Dartmouth Flood Observatory. Seven flooding events were recorded in Manila areas since 2006 (Table 5). This study focused on from 2006 - 2011. From the perspective of risk analysis, the flood risk considered is that of a lower likelihood, but a high-impact event.

Table 5. Historical Flood Events Since 2006

Starting Date	Ending Date	Main Cause	Number of Casualties	Number of Displaced People
27-Sep-06	06-Oct-06	Tropical cyclone	260	250,000
8-Aug-07	13-Aug-07	Tropical cyclone	11	12,000
17-Aug-07	24-Aug-07	Tropical cyclone	42	600,000
16-Jul-09	18-Jul-09	Tropical cyclone	5	n/a
25-Sep-09	01-Oct-09	Tropical storm Ketsana/Ondoy	420	200,000
2-Oct-09	17-Oct-09	Typhoon Parma	438	40,000
30-Oct-09	04-Nov-09	Tropical Storm Mirinae	n/a	98

Source: Dartmouth Flood Observatory (2011)[8].

4.1. Pre-assessment of Rainfall Data: Comparison between Station Data and Satellite Estimates

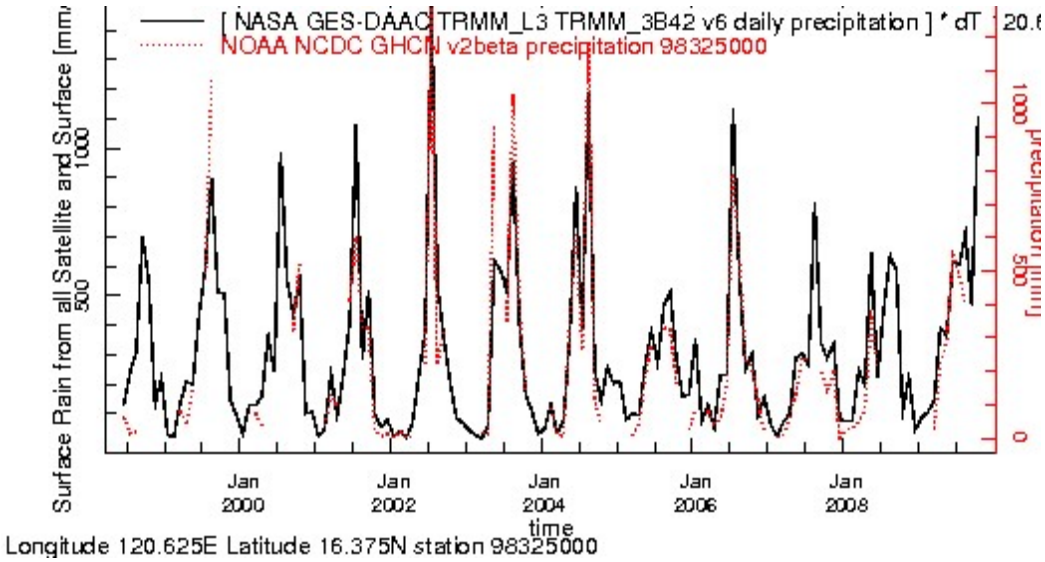
There are 22 stations in the area of focus. However, only three stations have complete data for the periods between 2006–2009. They are Dagupan (IWMO#: 98325000, 120.3E 16.1N), Baguios (IWMO#: 98328000, 120.6E 16.4N), and Cabanatuan (IWMO#: 98330000, 120.97E 15.48N). The time period from January 2006 to October 2009 was used. As a satellite estimate, the study used the Climate Prediction Center Morphing Technique (CMORPH) updated daily by the National Oceanic and Atmospheric Administration (NOAA) at 0.25° latitude/longitude spatial resolution. The temporal resolution of this product is daily with units of mm/hour. The results show that the gridded satellite estimates are quite reliable (Fig 2). Therefore, CMORPH was used in this study to estimate precipitation. Also, the study relates a flood event with rainfall events with two days before the flood event throughout the whole study.

In order to verify the gridded satellite estimates, we calculated the mean error (bias), root mean square error (RMSE), and correlation in a comparison between a station monthly precipitation time series and a time series of the satellite precipitation estimates of the pixels that contain the station location (Table 6). Station data was drawn from the NOAA NCDC GHCN v2beta station precipitation dataset. The temporal resolution is monthly and the database is composed of 7280 station data.

[8] http://www.dartmouth.edu/~floods/Archives/index.html

Table 6. Satellite and Station Precipitation Error Statistics (mm/hours)			
Station Name	Station #	Location	CMORPH
Bias			
Dagupan Philippine	98325000	16.1N 120.3E	-42.00
Baguio Philippine	98328000	16.4N 120.6E	-31.93
Cabanatuan	98330000	15.48N 120.97E	-53.35
RMSE			
Dagupan Philippine	98325000	16.1N 120.3E	124.90
Baguio Philippine	98328000	16.4N 120.6E	134.28
Cabanatuan	98330000	15.48N 120.97E	191.63
Correlation			
Dagupan Philippine	98325000	16.1N 120.3E	0.97
Baguio Philippines	98328000	16.4N 120.6E	0.93
Cabanatuan	98330000	15.48N 120.97E	0.76

Fig. 2. Comparison of Satellite and Station



4.2. Vegetation Moisture

As a proxy for vegetation moisture, the Normalized Difference Vegetation Index (NDVI) offered by USGS's MODIS NDVI was employed in this study. NDVI has been used for many years to measure and monitor plant growth (vigor) and vegetation cover. The study considered using other indices such as the Enhanced Vegetation Index (EVI), which is also offered by USGS's MODIS and the Normalized Differenced Water Index (NDWI). However, since the

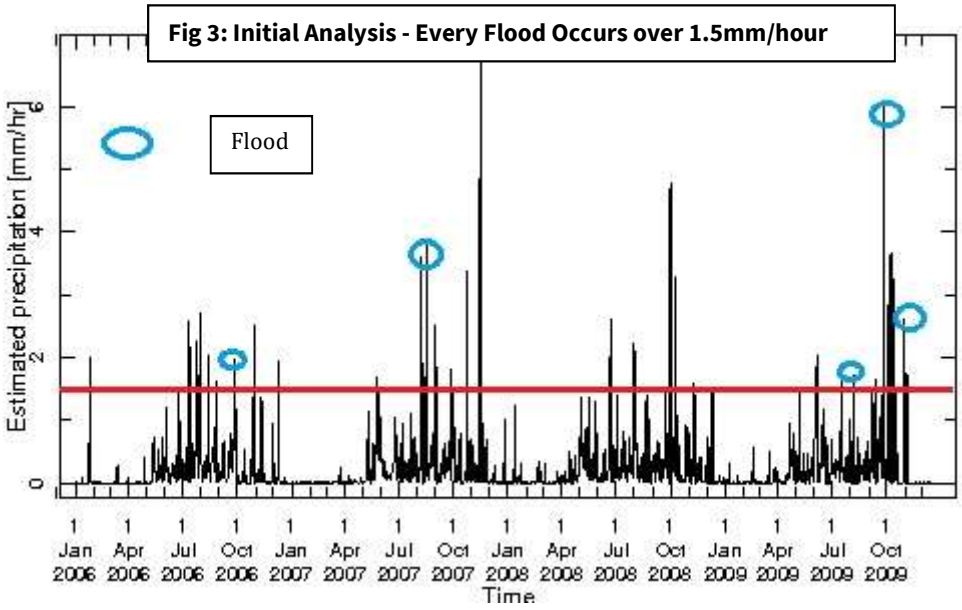
preliminary analysis showed that these indices have strong correlations (0.98-0.99) over the study's time and spatial range, this study will use only NDVI.

NDVI is derived from measurements made by the USGS .LandDAAC .MODIS .version_005 .SEAS .reflectance. The time resolution is 16-day daily, and special resolution is 250 meters. The number of data and locations that are used in this study are summarized in Table A-1 in the Appendix.

5. Exploratory Analysis

To assess critical values that cause floods and construct a prediction model of floods in Metro Manila, the time series data of a weighted-average daily precipitation values from CMORPH for 14 flood-causing rainfall events in the area of focus (Fig 3) are plotted over the period of January 2006 – December 2009.

Fig 3: Initial Analysis - Every Flood Occurs over 1.5mm/hour



Since the remote-sensing data records a very small amount of rainfall, we cut off the value below 0.5mm/hour. Then, we calculated one successive day of rainfall to one rainfall event. Thereafter, the volume and duration of each rainfall event have been calculated. During the period of the seven flood events, it can be estimated from the CMORPH that every flood event was accompanied by more than 1.5 mm/hour of rainfall intensity, which can be considered as a potential critical value that could indicate flood events in Manila. Also, most of the floods are similar with 20mm/event of rainfall volume and two days/event of rainfall duration (Table 7).

Table 7. Flood Events and Associated Rainfall Intensity, Volume, and Duration

Flood Starting Date	Flood Ending Date	Related Rainfall Event(s)	Highest Precipitation (mm/hour)	Rainfall Volume (mm/event)	Rainfall Duration (days/event)
27-Sep-06	06-Oct-06	Sept 27-28	1.16	46.64	2
8-Aug-07	13-Aug-07	August 5	0.84	20.07	1
		August 7-9	1.86	96.70	3
17-Aug-07	24-Aug-07	August 14-15	1.55	50.02	2
		August 17	2.25	54.02	1
		August 20	0.88	21.08	1
		August 22-24	1.48	89.69	3
16-Jul-09	18-Jul-09	July 16-July17	2.17	68.22	2
25-Sep-09	01-Oct-09	Sept 22- Sept23	2.11	74.88	2
		Sept 25	1.83	43.90	1
		Sept 27	0.58	13.93	1
2-Oct-09	17-Oct-09	Oct 2- Oct 4	6.86	215.58	3
		Oct 7- Oct8	12.34	376.88	2
30-Oct-09	04-Nov-09	Oct 30	0.81	19.41	1

Note that there is a two-day lag between flood and rainfall events.

As seen in Table 8, the reliability of 1.5 mm/hour of rainfall intensity is not high (e.g., accuracy is 69%). Tables 9 and 10 show that the critical value of two-day duration of one rainfall event has more accuracy than that of the volume (20mm/event). Therefore, none of these three models can provide an accurate forecast. In addition, the possibilities of false negatives can have devastating effects (i.e., if the prediction is wrong and then a flood occurs). Thus, it is necessary to improve the model. Since the highest rainfall intensity of 1.5 mm/hour has the largest accuracy, this study will continue to use this value. Then, using this value of 1.5 mm/hour, the study found rainfall events that did not cause flooding in Manila. Using a value of 1.5 mm/hour of rainfall is considered as the baseline model in this study and is compared with other models in terms of how other data will improve the model.

Table 8. A matrix of Predicted Flood or True Flood Using Rainfall Intensity

Highest rainfall intensity (mm/hour)	Actual Positive (=Flood occurred)	Actual Negative (=Flood did not occur)
Predicted Positive (>=1.5 mm/hour)	9	35
Predicted Negative (<1.5 mm/hour)	6	81
Precision	20%	
Sensitivity	60%	
Specificity	70%	
Accuracy	69%	

Table 9. A matrix of Predicted Flood or True Flood Using Rainfall Volume

Volume (mm/event)	Actual Positive (=Flood occurred)	Actual Negative (=Flood did not occur)
Predicted Positive (>20.0mm/event)	12	82
Predicted Negative (<20.0mm/event)	2	35
Precision	13%	
Sensitivity	86%	
Specificity	30%	
Accuracy	36%	

Table 10. A matrix of Predicted Flood or True Flood Using Rainfall Duration

Duration (days/event)	Actual Positive (=Flood occurred)	Actual Negative (=Flood did not occur)
Predicted Positive (>=2 days)	8	40
Predicted Negative (<2 days)	6	77
Precision	17%	
Sensitivity	57%	
Specificity	66%	
Accuracy	65%	

6. Visualization and Decision Tree

6.1. Methodologies to Compare Different Models

With this critical value in mind, the study looked at rainfall events that were not associated with floods in Manila, even if they were over the 1.5 mm/hour critical value. The results from CMORPH, with weighted average and over all areas of the watershed, are depicted in Figures A-1 to A-10 in the Appendix. There were 35 rainfall events of 1.5 mm/hour or greater that were not associated with flooding. Thus, we tried to find other data to be added using

visualization and a decision tree analysis. The models from the decision trees were evaluated in terms of sensitivity, precision, specificity, and accuracy. This method compares performance in the following models:

Model 1 (Baseline): 1.5mm/hour of rainfall intensity

Model 2: Adding the type of rainfall to the baseline

Model3: Adding NDVI to the baseline

Model4: Adding type of rainfall and NDVI to the baseline

6.2. Model 2: Adding the Type of Rainfall (Binary: Tropical Cyclone or Not)

Visualization

First, the type of rainfall was then added from UNISYS for every rainfall event that was more than 1.5 mm/hour. Since the website of UNISYS provides information regarding storms such as hurricanes and typhoons from all over the world, the study tried to determine what kind of tropical cyclones were associated with each rainfall event with over 1.5 mm/hour of rainfall. Figures A-1 to A-10 (in the appendix) and Table 5 show that floods were always associated with some type of tropical cyclones, while 81% of rainfall events with greater than 1.5 mm/hour rainfall that did not cause floods were not related with any tropical cyclone.

Based on the above observation, the study added a binary value to the decision tree (Fig. 4), and constructed a decision tree with a size of 3. By adding this binary value, the mode accuracy improved to 88% and specificity to 94% (Table 11). Consequently, if rainfall events are tropical cyclones, more attention should be paid to prepare for floods. However, 36% in sensitivity is considered very low given its catastrophic impact. Therefore, the study will look at adding another data type: vegetation index.

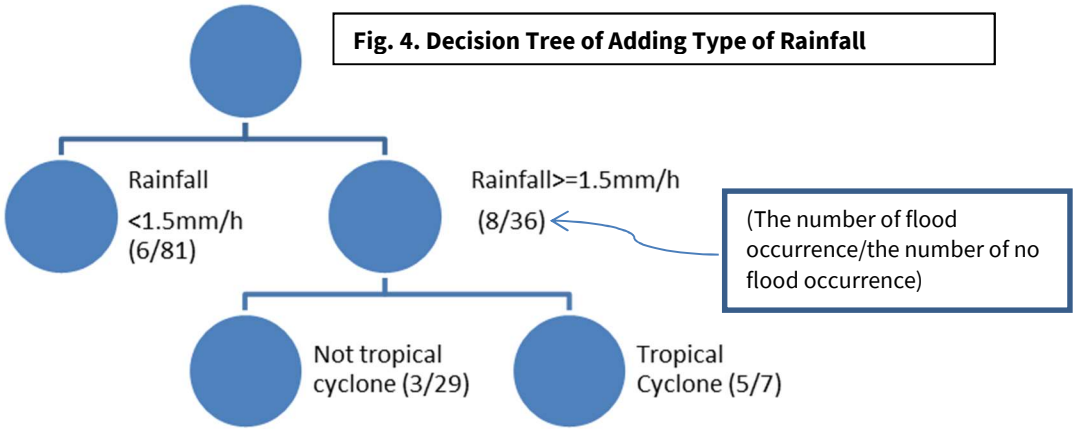


Fig. 4. Decision Tree of Adding Type of Rainfall

Table 11. 2 x 2 Matrix of Predicted Flood or True Flood Using Type of Rainfall

Highest Intensity and Rainfall Type	Actual Positive (=Flood occurred)	Actual Negative (=Flood did not occur)
Predicted Positive	5	7
Predicted Negative	9	110
Precision	42%	
Sensitivity	36%	
Specificity	94%	
Accuracy	88%	

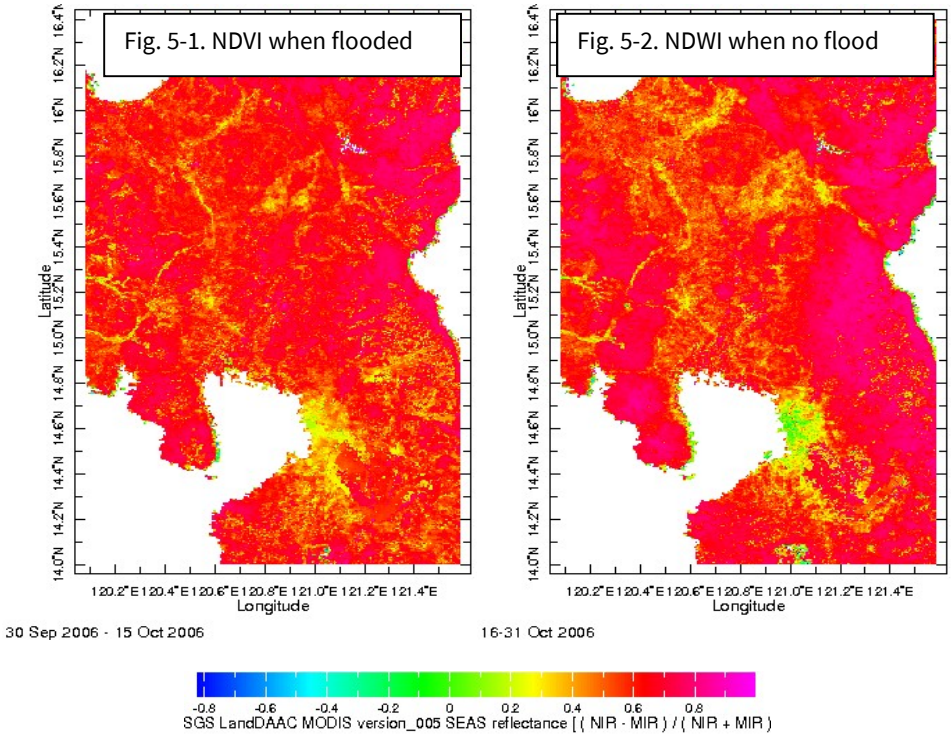
6.3. Model 3: Adding Vegetation Greenness and Moisture Index

Since only the rainfall amount and type do not provide a reliable value that determines whether floods would occur or not, the study also conducted vegetation analysis using NDVI. This index was plotted for both the watershed areas of focus (120.0833E 121.5833E, 14N - 16.4N) and for smaller areas surrounding Manila (121E-121.5E, 14N-15N). The study looked at the mapping, decision analysis using vegetation indices, and the time series of these indices over the studied period of January 2006 – March 2011.

Mapping

We compare close periods with floods and without floods. We assume that differences in NDVI values indicate differences in soil moisture for the following reason. Considering that NDVI measures vegetation and its closeness of two periods (for example, in Fig 5-1, two images are within one month), it is reasonable to assume that differences in

NDVI values between two images would be not due to differences in vegetation but due to differences in soil moistures. It is clear that every time a flood occurred, NDVI indicated that soil moistures were high over the region (for example, in Fig. 5 the red region is more likely to have floods.) However, NDVI was high even when floods did not occur (for example, **Fig. 5-1 and 5-2). Therefore, there were no significant differences in the NDVI map visualization for determining a critical metric to project a flood.**



As can be seen in Figures 6 and Table 12, the index will improve the accuracy and specificity if added. More than a 0.5 NDVI value has an accuracy of 73% and specificity 74%. However, the accuracy of this model has declined to 73% from 88% of the one that uses the types of rainfall. Thus, this index as a single source cannot provide a critical value to provide early warning for floods in Manila.

Fig. 6. Decision Tree

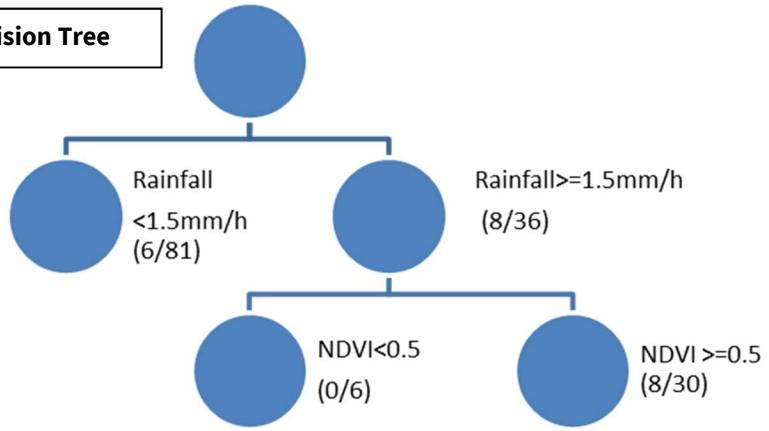


Table 12. A matrix of Predicted Flood or True Flood Using NDVI

NDVI>=0.5	Actual Positive (=Flood occurred)	Actual Negative (=Flood did not occur)
Predicted Positive	8	30
Predicted Negative	6	87
Precision	21%	
Sensitivity	57%	
Specificity	74%	
Accuracy	73%	

6.4. Model 4: Integration of Rainfall Amount, Rainfall Type, and Vegetation Indices

Comparing these two decision trees, adding information on the type of rainfall largely improved accuracy and specificity while adding the vegetation index improved sensitivity. As can be seen, the model with the added type of rainfall and the vegetation index performs best in terms of accuracy, precision, specificity (Tables 13 and 14). In terms of accuracy, the study places the most value in Models 2 and 4. However, given the devastating effect of false negative, shown in the value of sensitivity, a policymaker should also look at the volume of a rainfall event whose value of sensitivity is high (86%).

Table 13. 2 x 2 Matrix of Predicted Flood or True Flood Using Rainfall type and NDVI

Type of rain Every index	Actual Positive (=Flood occurred)	Actual Negative (=Flood did not occur)
Predicted Positive	5	7
Predicted Negative	9	110
Precision	42%	
Sensitivity	36%	
Specificity	94%	
Accuracy	88%	

Table 14. Summary of the Evaluation Values for Each Model

	Accuracy	Precision	Sensitivity	Specificity
Baseline 1.5 mm/h	69%	20%	60%	70%
Volume >20.0 mm/event	36%	13%	86%	36%
Duration >=2 days/ event	65%	17%	57%	66%
Model 2: Adding type of rainfall	88%	42%	36%	94%
Model 3: Adding NDVI	73%	21%	57%	73%
Model 4: Adding type of rainfall and NDVI	88%	42%	36%	94%

6.5. Times Series Visualization

Next, we looked at the time series of NDVI from January 2006 – March 2011 (Fig. A-11 – A-13 in the Appendix). There are weak tendencies in the high vegetation index (NDVI) before flooding events (indicated by red vertical lines in Fig. A-11 – A-13 in the Appendix). However, there continue to be no apparent differences in the indices between the rainfall events leading to floods with more than 1.5 mm/hour rainfall and the ones that did not cause flood events (indicated by green vertical lines in Fig. A-11 – A-13 in the Appendix). Hence, it is critical to look at various indices when attempting to predict floods.

7. Cross-validated Ridge Regression

To predict a response variable, the study next conducted a logistic regression using cross-validation. First, data were split into training and test sets using an 80/20 split, and then conducting logistic regressions for four models:

(1) Baseline model (IV: Rainfall amount)

(2) Adding the type of rainfall to the baseline model (IV: rainfall amount, type of rainfall)

(3) Adding vegetation indices to the baseline (IV: rainfall amount, NDVI)

(4) Adding both types of rainfall and vegetation indices (IV: rainfall amount, type of rainfall, NDVI).

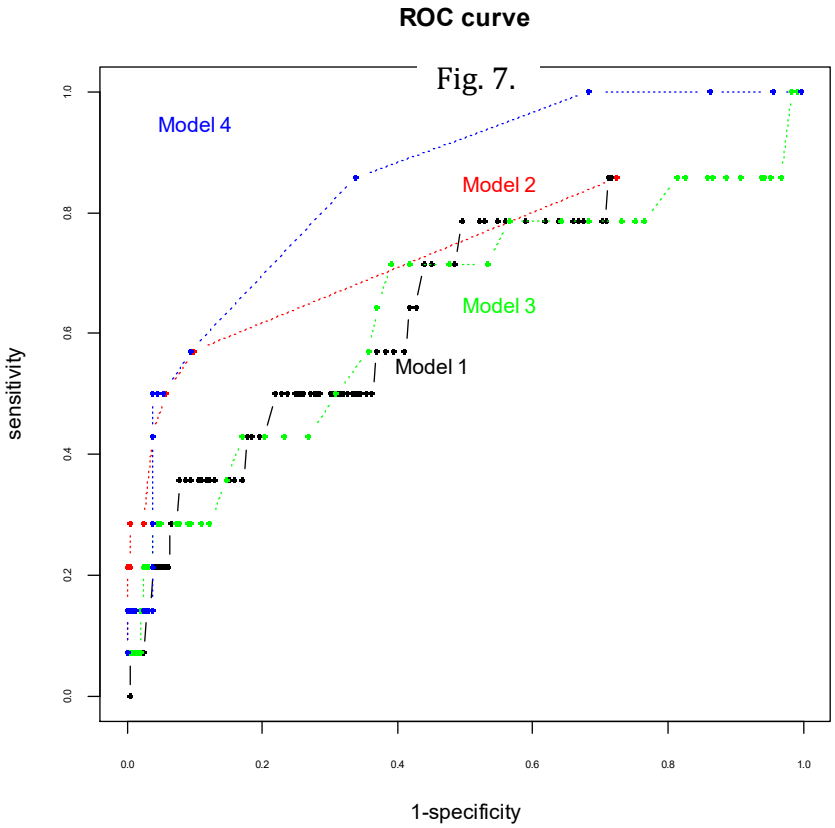
A comparison of the Akaike information criterion (AIC) and Bayesian Information Criterion (BIC) (Friedman, Hastie, & Tibshirani, 2001), which evaluates in-sample modeling, Table 15 indicates that the best models are Model 2 and Model 4.

Table 15. AIC and BIC Values for Each Model

Model	AIC	BIC
(1) Baseline model (IV: Rainfall amount)	351.8	392.2
(2) Adding the type of rainfall to the baseline model (IV: rainfall amount, type of rainfall)	279.4	320.7
(3) Adding vegetation indices to the baseline (IV: rainfall amount, NDVI)	347.5	396.9
(4) Adding both types of rainfall and vegetation indices (IV: rainfall amount, type of rainfall, NDVI)	281.4	327.7

We then went on to use these models to make predictions for the test set. These models gave a probability of flood for each observation in the test set. A False Positive Rate (1-specificity) and True Positive Rate (sensitivity) were plotted as Receiver Operating Characteristic curves (ROC) (Friedman et al., 2001)in Fig. 7. In ROC curve, curves in the upper left indicates better models (Friedman et al., 2001). As can be seen, there is a clear indication that both Models 2 and 4 show better performances in the test set than the Models 1 or 3. Yet, there were no clear differences between Models 2 and 4. This result is consistent with the result that is gained through the previous decision tree analysis. Namely, Model 2, which uses rainfall intensity and types, and Model 4, which utilizes all rainfall intensity, type, and vegetation index, can predict most accurately.

Fig. 7.



8. Conclusions and Study Limitations

We demonstrated an effective way of utilizing global satellite data for local risk-informed decision-making for urban flood management. The RS and EO sector is rapidly changing as the private sector enter the field and develop new technologies. CubeSats, Unmanned Aerial Vehicles (UAVs), smartphones are bringing a deluge of data and new measurements, which results in helping overcoming the limitations and complementing the potential of RS in DRM: real-time high-definition video, storm-cell development, flood propagation, precipitation monitoring, measurements of snow depth, and floods and estimated evaporation at sub-meter resolution from small unmanned drones. Petabytes of new data that may be a game-changing in hydrological sciences, such as in flood risk and DRM, if usefully exploited (McCabe et al., 2017).

Our case study demonstrated that there were useful relationships between flooding and remotely-sensed environmental data, such as rainfall amount and vegetation moisture using several machine learning techniques, such

as visualization, decision tree, and logistic regression. The initial analysis of rainfall data indicated that most flooding events had more than 1.5 mm/hour of rainfall, 20 mm/event of rainfall volume, and two days/event of rainfall duration over the studied watershed area. Yet, we found that many rainfall events of 1.5 mm/hour or higher did not cause floods.

To improve the performance of the prediction models, we explored the types of rainfall were associated with floods. All floods in our observation period were associated with tropical cyclones. In contrast, rainfall events of more than 1.5 mm/hour that did not lead to floods were unlinked at 81% with tropical cyclones. We also tried to find relationships between vegetation moisture and greenness and flood events. Prior to floods, there were tendencies for higher indices of NDVI. The decision tree shows that accuracy and specificity improved because of these indices. However, the study did not find a single critical value of these indices that can provide a reliable indication for the early prediction of floods. In contrast, as the result of decision trees, the classification rate and other indices such as sensitivity and specificity show that the best prediction models are the one that uses rainfall types, and the one that integrates rainfall amount, rainfall type, and vegetation indices. This result was also supported by findings from the cross-validated logistic regression. The study demonstrated that a model adding rainfall types is better than one that only utilizes rainfall amount or adds the vegetation index. Yet, the results did not show which is better, the model that uses only rainfall type or the one that integrates all the information including rainfall type and various vegetation moisture indices.

In summary, it is essential for local policymakers to comprehensively look at such indices and rainfall intensity, volume, and duration to provide a flood prediction early-warning system in Manila. Local policymakers must consider rainfall amount and duration, type of rainfall, and vegetation indices along with other important indicators such as water height at local rivers and dams. Near term weather forecasts have improved dramatically in accuracy in many places, and several global weather forecasting products are now available and could be used together with an established model using global indicators to assess flooding threats. Eventually, addressing DRM has substantial synergies with sustainable development (Haraguchi & Lall, 2019).

As with any study, there are some limitations. First, it considers only the time series of rainfall dated from January 2006 to December 2009 because of the lack of data on floods in Manila. With current data gathering now in progress, future researchers will have a broader time domain to examine. Second, this study only examined three types

of predictors (rainfall amount, rainfall type, and three vegetation indices). Future work should consider other factors that cause floods in an urban area, such as any human-made and natural infrastructure considerations.

Bibliography

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., . . . Bolvin, D. (2003). The version-2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–present). *Journal of hydrometeorology, 4*(6), 1147-1167.

Alsdorf, D. E., Rodríguez, E., & Lettenmaier, D. P. (2007). Measuring surface water from space. *Reviews of Geophysics, 45*(2).

Ashouri, H., Hsu, K.-L., Sorooshian, S., Braithwaite, D. K., Knapp, K. R., Cecil, L. D., . . . Prat, O. P. (2015). PERSIANN-CDR: Daily precipitation climate data record from multisatellite observations for hydrological and climate studies. *Bulletin of the American Meteorological Society, 96*(1), 69-83.

Bankoff, G. (2003). Constructing vulnerability: the historical, natural and social generation of flooding in metropolitan Manila. *Disasters, 27*(3), 224-238.

Bessis, J.-L., Bequignon, J., & Mahmood, A. (2004). Three typical examples of activation of the International Charter "Space and Major Disasters". *Advances in Space Research, 33*(3), 244-248.

Bhatt, C., Rao, G., Farooq, M., Manjusree, P., Shukla, A., Sharma, S., . . . Diwakar, P. (2017). Satellite-based assessment of the catastrophic Jhelum floods of September 2014, Jammu & Kashmir, India. *Geomatics, Natural Hazards and Risk, 8*(2), 309-327.

Brakenridge, R., & Anderson, E. (2006). MODIS-based flood detection, mapping and measurement: the potential for operational hydrological applications. In *Transboundary floods: reducing risks through flood management* (pp. 1-12): Springer.

Brisco, B., Schmitt, A., Murnaghan, K., Kaya, S., & Roth, A. (2013). SAR polarimetric change detection for flooded vegetation. *International Journal of Digital Earth, 6*(2), 103-114.

Cadiz, N. R. (2018). UP NOAH in Building Resilient Philippines; Multi-hazard and Risk Mapping for the Future. *Procedia Engineering, 212*, 1018-1025.

Cian, F., Marconcini, M., & Ceccato, P. (2018). Normalized Difference Flood Index for rapid flood mapping: Taking advantage of EO big data. *Remote Sensing of Environment, 209*, 712-730.

Cossu, R., Schoepfer, E., Bally, P., & Fusco, L. (2009). Near real-time SAR-based processing to support flood monitoring. *Journal of Real-Time Image Processing, 4*(3), 205-218.

Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1): Springer series in statistics New York.

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., . . . Hoell, A. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific data, 2*, 150066.

Garcia, F. C. C., Retamar, A. E., & Javier, J. C. (2015). *A real time urban flood monitoring system for metro Manila.* Paper presented at the TENCON 2015-2015 IEEE Region 10 Conference.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment, 202*, 18-27.

Haq, M., Akhtar, M., Muhammad, S., Paras, S., & Rahmatullah, J. (2012). Techniques of remote sensing and GIS for flood monitoring and damage assessment: a case study of Sindh province, Pakistan. *The Egyptian Journal of Remote Sensing and Space Science, 15*(2), 135-141.

Haraguchi, M. (2018). *Innovations towards Climate-Induced Disaster Risk Assessment and Response.* Columbia University,

Haraguchi, M., & Kim, S. (2016). Critical infrastructure interdependence in New York City during Hurricane Sandy. *International Journal of Disaster Resilience in the Built Environment, 7*(2), 133-143.

Haraguchi, M., & Lall, U. (2015). Flood risks and impacts: A case study of Thailand's floods in 2011 and research questions for supply chain decision making. *International Journal of Disaster Risk Reduction, 14*, 256-272.

Haraguchi, M., & Lall, U. (2019). *Linkage Between Disaster Risk Management, Climate Change, and Sustainable Development: Concepts, Frameworks and Solutions*. Working Paper.

Heimhuber, V., Tulbure, M., & Broich, M. (2018). Addressing spatio-temporal resolution constraints in Landsat and MODIS-based mapping of large-scale floodplain inundation dynamics. *Remote Sensing of Environment, 211*, 307-320.

Henry, J. B., Chastanet, P., Fellah, K., & Desnos, Y. L. (2006). Envisat multi-polarized ASAR data for flood mapping. *International Journal of Remote Sensing, 27*(10), 1921-1929.

Horritt, M., Mason, D., & Luckman, A. (2001). Flood boundary delineation from synthetic aperture radar imagery using a statistical active contour model. *International Journal of Remote Sensing, 22*(13), 2489-2507.

Huang, C., Peng, Y., Lang, M., Yeo, I.-Y., & McCarty, G. (2014). Wetland inundation mapping and change monitoring using Landsat and airborne LiDAR data. *Remote Sensing of Environment, 141*, 231-242.

Islam, A. S., Bala, S. K., & Haque, M. (2010). Flood inundation map of Bangladesh using MODIS time-series images. *Journal of Flood Risk Management, 3*(3), 210-222.

Klemas, V. V. (2009). The role of remote sensing in predicting and determining coastal storm impacts. *Journal of Coastal Research*, 1264-1275.

Kordelas, G., Manakos, I., Aragonés, D., Díaz-Delgado, R., & Bustamante, J. (2018). Fast and Automatic Data-Driven Thresholding for Inundation Mapping with Sentinel-2 Data. *Remote Sensing, 10*(6), 910.

Kubota, T., Shige, S., Hashizume, H., Aonashi, K., Takahashi, N., Seto, S., . . . Nakagawa, K. (2007). Global precipitation map using satellite-borne microwave radiometers by the GSMaP project: Production and validation. *IEEE Transactions on Geoscience and Remote Sensing, 45*(7), 2259-2275.

Kwak, Y., Arifuzzanman, B., & Iwami, Y. (2015). Prompt proxy mapping of flood damaged rice fields using MODIS-derived indices. *Remote Sensing, 7*(12), 15969-15988.

Ma, Y., Wu, H., Wang, L., Huang, B., Ranjan, R., Zomaya, A., & Jie, W. (2015). Remote sensing big data computing: Challenges and opportunities. *Future Generation Computer Systems, 51*, 47-60.

Mahmood, A. (2012). Monitoring disasters with a constellation of satellites–type examples from the International Charter 'Space and Major Disasters'. *Geocarto International, 27*(2), 91-101.

Martinis, S., Kersten, J., & Twele, A. (2015). A fully automated TerraSAR-X based flood service. *ISPRS Journal of Photogrammetry and Remote Sensing, 104*, 203-212.

Marzano, F. S., Mori, S., Weinman, J. A., & Montopoli, M. (2012). Modeling polarimetric response of spaceborne synthetic aperture radar due to precipitating clouds from X-to Ka-band. *IEEE Transactions on Geoscience and Remote Sensing, 50*(3), 687-703.

Matgen, P., Schumann, G., Henry, J.-B., Hoffmann, L., & Pfister, L. (2007). Integration of SAR-derived river inundation areas, high-precision topographic data and a river flow model toward near real-time flood management. *International Journal of applied earth observation and geoinformation, 9*(3), 247-263.

McCabe, M. F., Rodell, M., Alsdorf, D. E., Gonzalez Miralles, D., Uijlenhoet, R., Wagner, W., . . . Franz, T. E. (2017). The future of Earth observation in hydrology. *Hydrology and Earth System Sciences, 21*(7), 3879-3914.

Mertes, L. A. (2002). Remote sensing of riverine landscapes. *Freshwater biology, 47*(4), 799-816.

Oddo, P. C., Ahamed, A., & Bolten, J. D. (2018). Socioeconomic Impact Evaluation for Near Real-Time Flood Detection in the Lower Mekong River Basin. *Hydrology, 5*(2), 23.

Pantaleoni, E., Engel, B., & Johannsen, C. (2007). Identifying agricultural flood damage using Landsat imagery. *Precision Agriculture, 8*(1-2), 27-36.

Porio, E. (2011). Vulnerability, adaptation, and resilience to floods and climate change-related risks among marginal, riverine communities in Metro Manila. *Asian Journal of Social Science, 39*(4), 425-445.

Qi, S., Brown, D. G., Tian, Q., Jiang, L., Zhao, T., & Bergen, K. M. (2009). Inundation extent and flood frequency mapping using LANDSAT imagery and digital elevation models. *GIScience & Remote Sensing, 46*(1), 101-127.

Rahman, M. S., & Di, L. (2017). The state of the art of spaceborne remote sensing in flood management. *Natural Hazards, 85*(2), 1223-1248.

Ryu, J.-H., Won, J.-S., & Min, K. D. (2002). Waterline extraction from Landsat TM data in a tidal flat: a case study in Gomso Bay, Korea. *Remote Sensing of Environment, 83*(3), 442-456.

Schumann, G., Bates, P. D., Horritt, M. S., Matgen, P., & Pappenberger, F. (2009). Progress in integration of remote sensing–derived flood extent and stage data and hydraulic models. *Reviews of Geophysics, 47*(4).

Serpico, S. B., Dellepiane, S., Boni, G., Moser, G., Angiati, E., & Rudari, R. (2012). Information extraction from remote sensing images for flood monitoring and damage evaluation. *Proceedings of the IEEE, 100*(10), 2946-2970.

Stoutjesdijk, J. (2017). *Philippines - Metro Manila Flood Management Project: environmental and social impact assessment : Environmental and social impact assessment*. Retrieved from Washington DC: http://documents.worldbank.org/curated/en/932441468285638424/Environmental-and-social-impact-assessment

Tablazon, J., Caro, C., Lagmay, A., Briones, J., Dasallas, L., Lapidez, J., . . . Gonzalo, L. (2015). Probabilistic storm surge inundation maps for Metro Manila based on Philippine public storm warning signals. *Natural Hazards & Earth System Sciences, 15*(3).

Tapia-Silva, F.-O., Itzerott, S., Foerster, S., Kuhlmann, B., & Kreibich, H. (2011). Estimation of flood losses to agricultural crops using remote sensing. *Physics and Chemistry of the Earth, Parts A/B/C, 36*(7-8), 253-265.

The Dartmouth Flood Observatory. (2018). Space-based Measurement, Mapping, and Modeling of Surface Water For Research, Humanitarian, and Water Management Applications. Retrieved from http://floodobservatory.colorado.edu/

Thomas, R. F., Kingsford, R. T., Lu, Y., & Hunter, S. J. (2011). Landsat mapping of annual inundation (1979–2006) of the Macquarie Marshes in semi-arid Australia. *International Journal of Remote Sensing, 32*(16), 4545-4569.

Tong, X., Luo, X., Liu, S., Xie, H., Chao, W., Liu, S., . . . Jiang, Y. (2018). An approach for flood monitoring by the combined use of Landsat 8 optical imagery and COSMO-SkyMed radar imagery. *ISPRS Journal of Photogrammetry and Remote Sensing, 136*, 144-153.

Twele, A., Cao, W., Plank, S., & Martinis, S. (2016). Sentinel-1-based flood mapping: a fully automated processing chain. *International Journal of Remote Sensing, 37*(13), 2990-3004.

Van der Sande, C., De Jong, S., & De Roo, A. (2003). A segmentation and classification approach of IKONOS-2 imagery for land cover mapping to assist flood risk and flood damage assessment. *International Journal of applied earth observation and geoinformation, 4*(3), 217-229.

Van Westen, C. J. (2013). Remote sensing and GIS for natural hazards assessment and disaster risk management. *Treatise on geomorphology, 3*, 259-298.

Waisurasingha, C., Aniya, M., Hirano, A., Kamusoko, C., & Sommut, W. (2007). Application of c-band synthetic aperture radar data and digital elevation model to evaluate the conditions of flood-affected paddies: Chi river basin, thailand. *Asian Association on Remote Sensing, Proceedings ACRS*.

Wang, Q., Watanabe, M., Hayashi, S., & Murakami, S. (2003). Using NOAA AVHRR data to assess flood damage in China. *Environmental monitoring and assessment, 82*(2), 119-148.

Wang, Y., Colby, J., & Mulcahy, K. (2002). An efficient method for mapping flood extent in a coastal floodplain using Landsat TM and DEM data. *International Journal of Remote Sensing, 23*(18), 3681-3696.

Ward, D., Petty, A., Setterfield, S., Douglas, M., Ferdinands, K., Hamilton, S., & Phinn, S. (2014). Floodplain inundation and vegetation dynamics in the Alligator Rivers region (Kakadu) of northern Australia assessed using optical and radar remote sensing. *Remote Sensing of Environment, 147*, 43-55.

Wilson, B. A., & Rashid, H. (2005). Monitoring the 1997 flood in the Red River Valley using hydrologic regimes and RADARSAT imagery. *Canadian Geographer/Le Géographe canadien, 49*(1), 100-109.

Yamagata, Y., & Akiyama, T. (1988). Flood damage analysis using multitemporal Landsat Thematic Mapper data. *TitleREMOTE SENSING, 9*(3), 503-514.

Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2017). Big Data and cloud computing: innovation opportunities and challenges. *International Journal of Digital Earth, 10*(1), 13-53.

Appendix

Table A-1: Summary of Data Used in This Study			
	Type of Data	Number of data	Location
Flood occurrence	Binary	7	Dartmouth Flood Observatory.[9]
Rainfall Amount : Ground station	Numerical	365 (days) x3 (years) x22 (stations) = 24090	*NOAA NCDC GHCN v2beta* station precipitation dataset[10]
Rainfall Amount :CMORPH	Numerical	365 (days) x3 (years) = 1095	NOAA
Rainfall Amount :TRMM	Numerical	365 (days) x3 (years) = 1095	NOAA
Rainfall Type	Binary (Tropical cyclone or not)	365 (days) x3 (years) = 1095	UNISYS[11]
Vegetation Indices: NDVI	Numerical	52 weeks x 3 (years) = 156	USGS .LandDAAC .MODIS .version_ 005 .SEAS .reflectance.

[9] http://www.dartmouth.edu/~floods/Archives/index.html
[10] http://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCDC/.GHCN/
[11] http://weather.unisys.com/hurricane/index.php

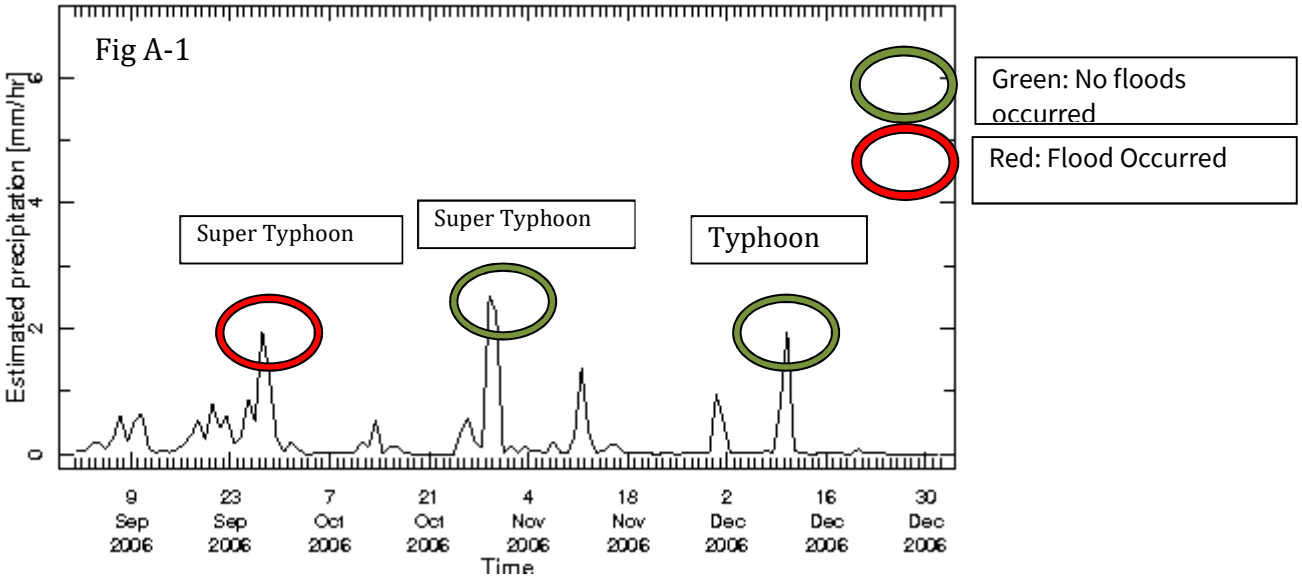


Fig A-1
Estimated precipitation [mm/hr]
0
2
4
6
Super Typhoon
Super Typhoon
Typhoon
9 Sep 2006
23 Sep 2006
7 Oct 2006
21 Oct 2006
4 Nov 2006
18 Nov 2006
2 Dec 2006
16 Dec 2006
30 Dec 2006
Time
Green: No floods occurred
Red: Flood Occurred

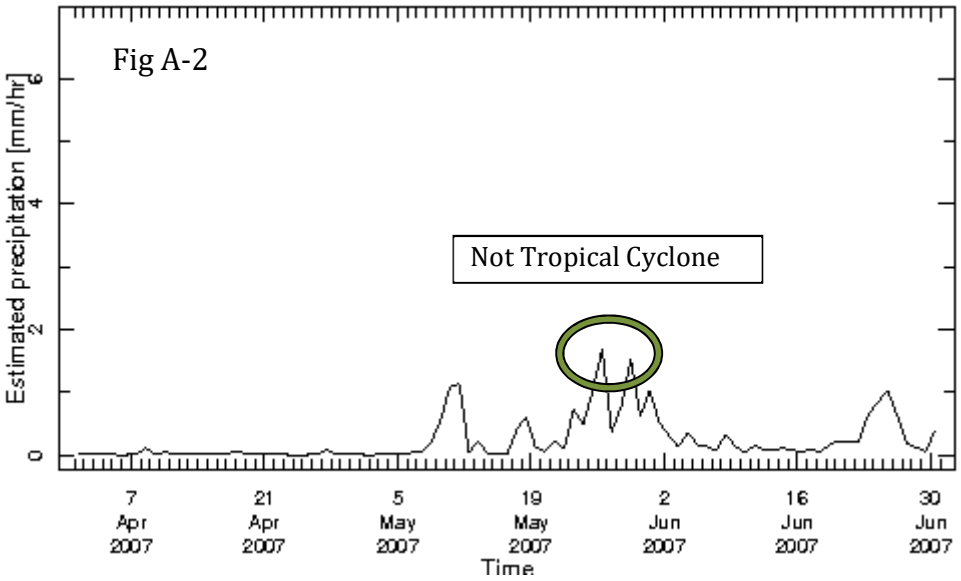


Fig A-2
Estimated precipitation [mm/hr]
0
2
4
6
Not Tropical Cyclone
7 Apr 2007
21 Apr 2007
5 May 2007
19 May 2007
2 Jun 2007
16 Jun 2007
30 Jun 2007
Time

Fig A-3

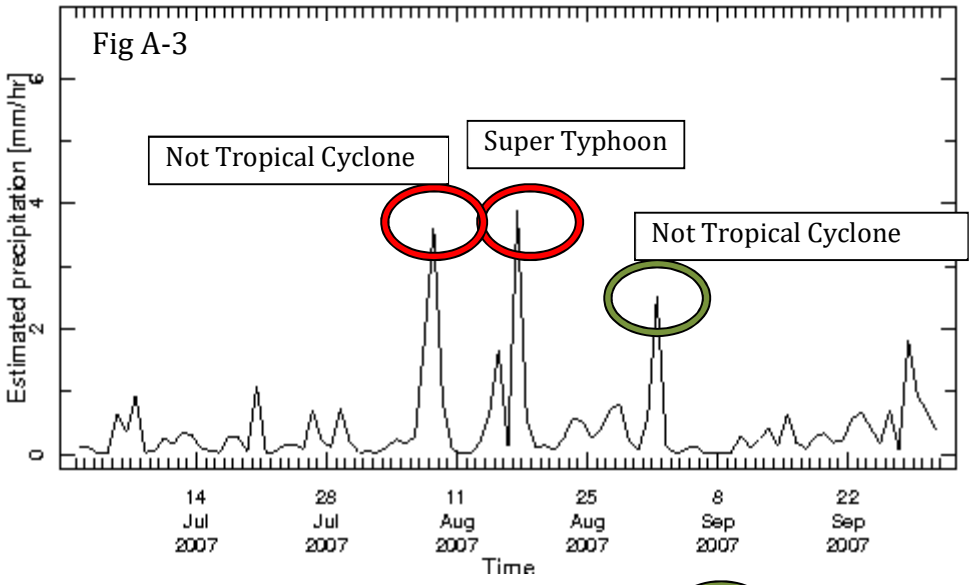
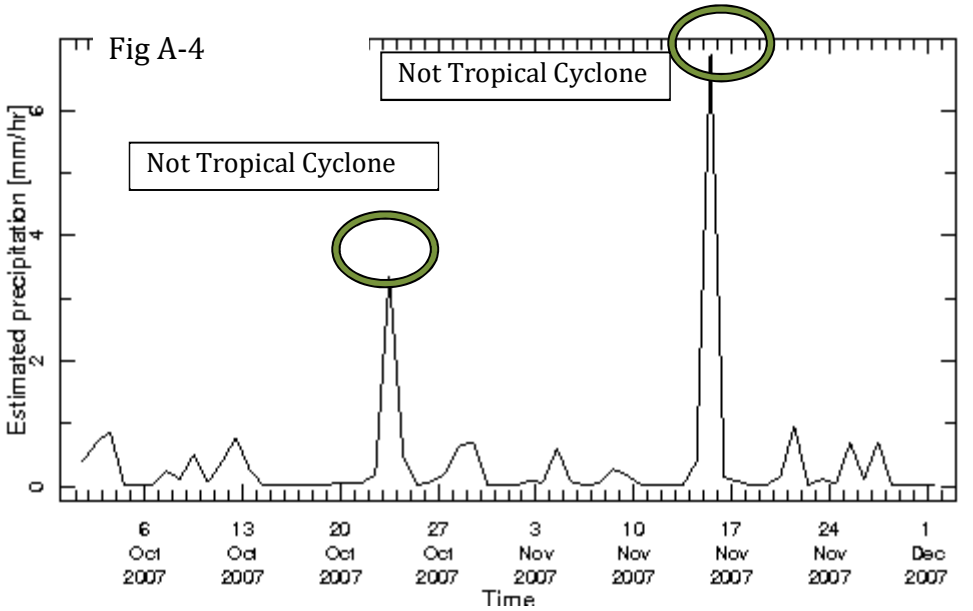

Not Tropical Cyclone
Super Typhoon
Not Tropical Cyclone
Estimated precipitation [mm/hr]
0
2
4
6
14 Jul 2007
28 Jul 2007
11 Aug 2007
25 Aug 2007
8 Sep 2007
22 Sep 2007
Time

Fig A-4


Not Tropical Cyclone
Not Tropical Cyclone
Estimated precipitation [mm/hr]
0
2
4
6
6 Oct 2007
13 Oct 2007
20 Oct 2007
27 Oct 2007
3 Nov 2007
10 Nov 2007
17 Nov 2007
24 Nov 2007
1 Dec 2007
Time

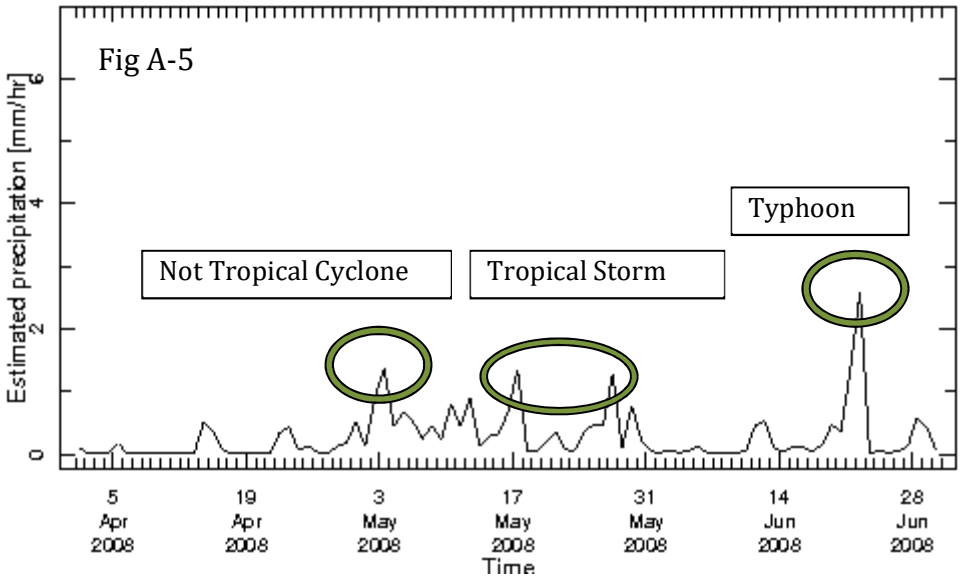


Fig A-5
Estimated precipitation [mm/hr]
0
2
4
6
Not Tropical Cyclone
Tropical Storm
Typhoon
5 Apr 2008
19 Apr 2008
3 May 2008
17 May 2008
31 May 2008
14 Jun 2008
28 Jun 2008
Time

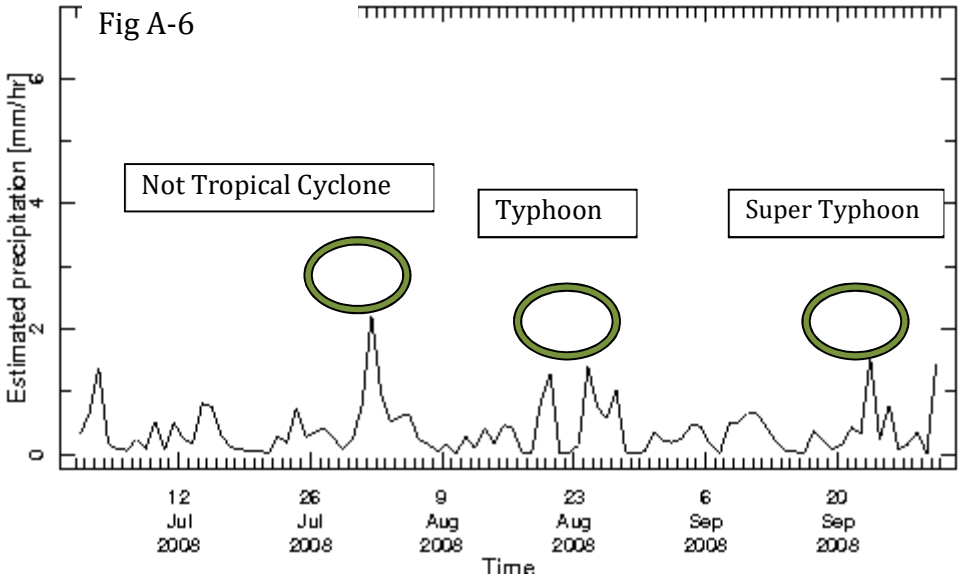


Fig A-6
Estimated precipitation [mm/hr]
0
2
4
6
Not Tropical Cyclone
Typhoon
Super Typhoon
12 Jul 2008
26 Jul 2008
9 Aug 2008
23 Aug 2008
6 Sep 2008
20 Sep 2008
Time

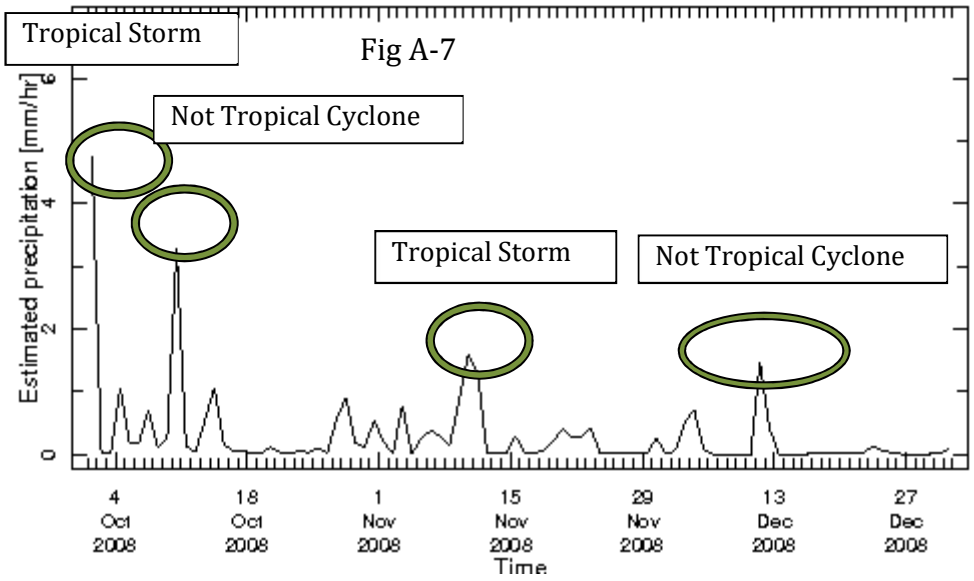

Tropical Storm
Not Tropical Cyclone
Tropical Storm
Not Tropical Cyclone
Estimated precipitation [mm/hr]
0
2
4
6
4 Oct 2008
18 Oct 2008
1 Nov 2008
15 Nov 2008
29 Nov 2008
13 Dec 2008
27 Dec 2008
Time

Fig A-7

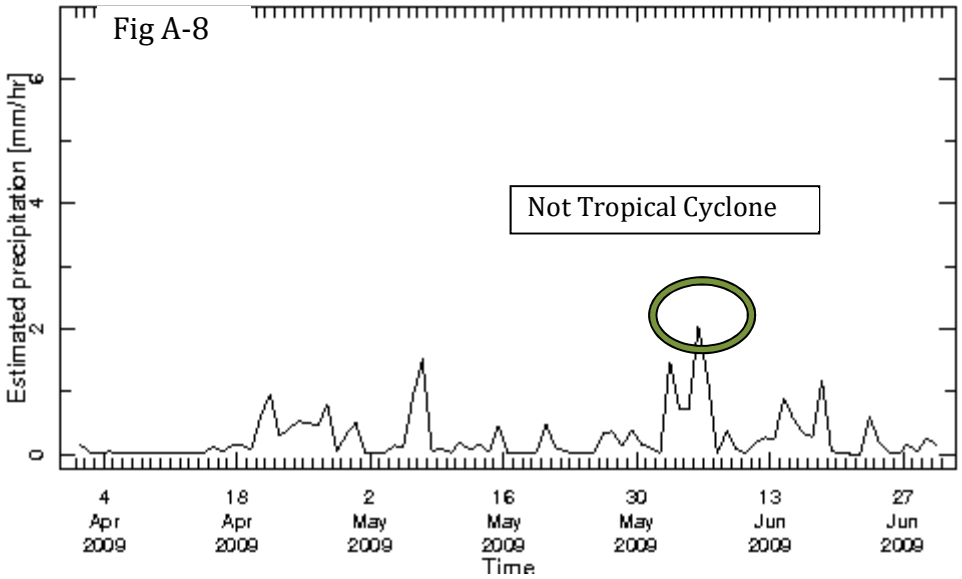

Not Tropical Cyclone
Estimated precipitation [mm/hr]
0
2
4
6
4 Apr 2009
18 Apr 2009
2 May 2009
16 May 2009
30 May 2009
13 Jun 2009
27 Jun 2009
Time

Fig A-8

Fig A-9

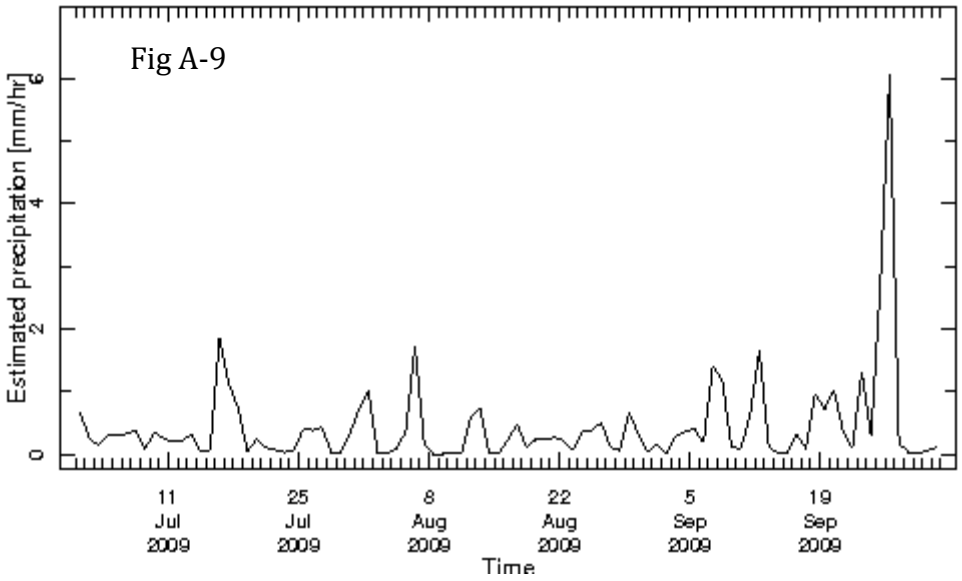


Fig A-10

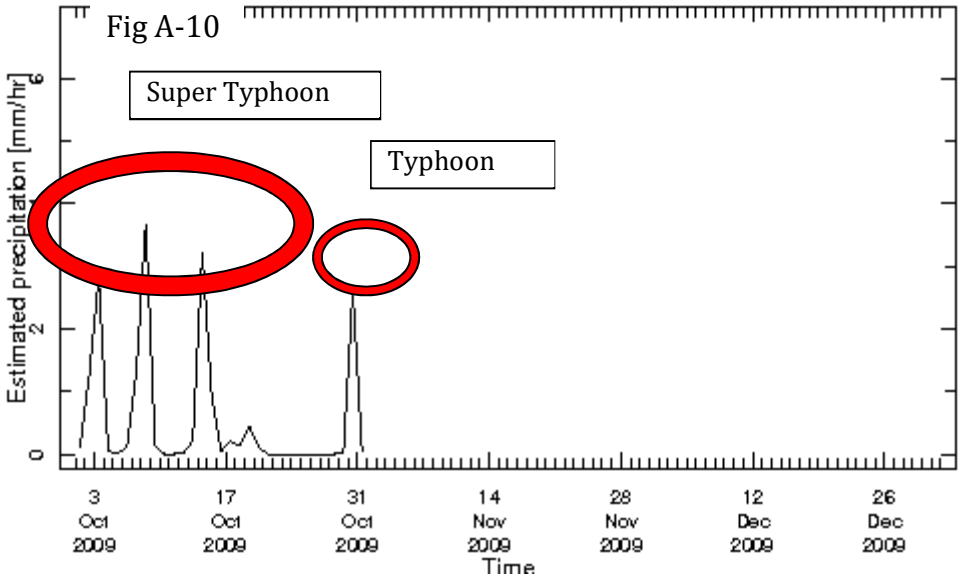


Figure A-11: NDVI (Jan 2006 – March 2011)

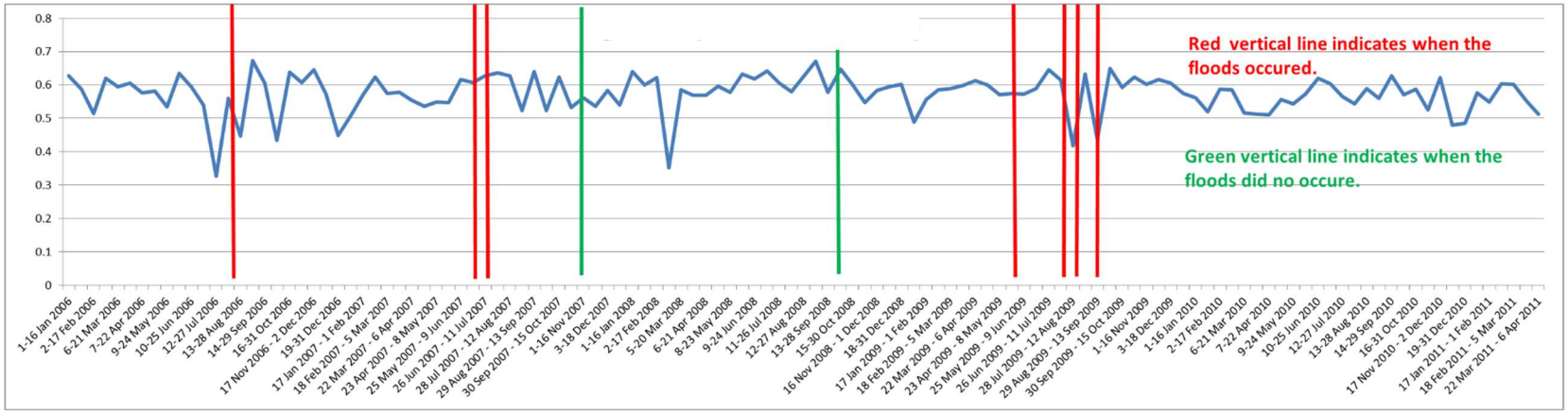


Figure A-12: EVI (Jan 2006 – March 2011)

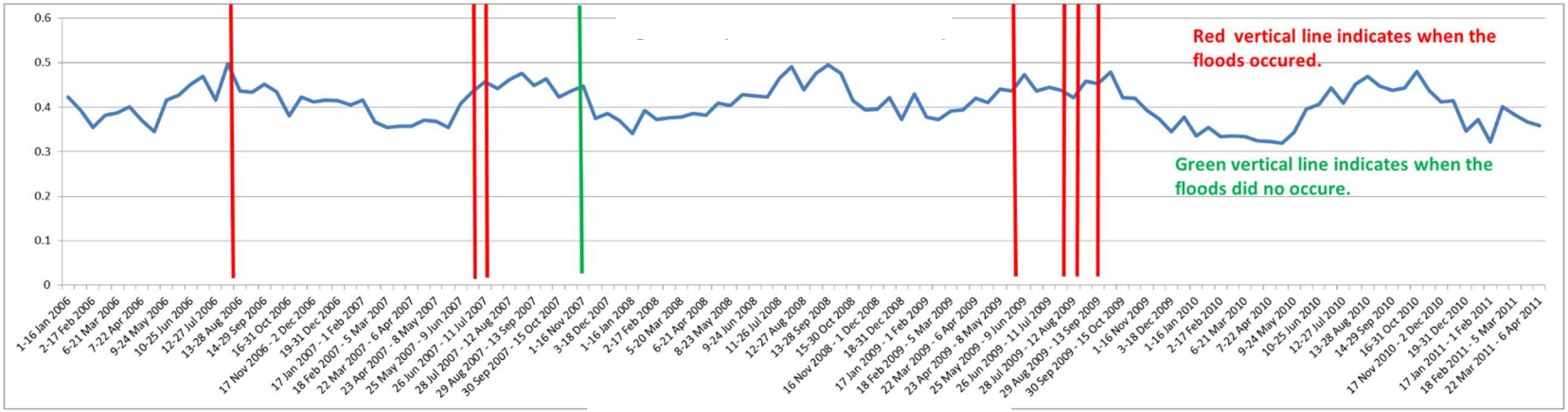


Figure A-13: NDWI (Jan 2006 – March 2011)