



Strengthening Financial Resilience to Disasters in Asia

Rapid Financial Impact Assessment
After Floods Using Satellite Data:

A Focus on South and Southeast Asia

Full Technical Report

July 2017

**Disaster Risk Financing
& Insurance Program**



SUPPORTED BY
WORLD BANK GROUP



The
**ROCKEFELLER
FOUNDATION**



International Research Institute
for Climate and Society
EARTH INSTITUTE | COLUMBIA UNIVERSITY

Columbia Water Center
EARTH INSTITUTE | COLUMBIA UNIVERSITY

Disclaimer and Copyright

© 2017 The International Bank for Reconstruction and Development/The World Bank

1818 H Street NW
Washington DC 20433
Telephone: 202-473-1000
Internet: www.worldbank.org

All rights reserved

This publication is a product of the staff of the International Bank for Reconstruction and Development/The World Bank and the Columbia University Water Center and the International Research Institute for Climate and Society. The findings, interpretations, and conclusions expressed in this volume do not necessarily reflect the views of the Executive Directors of The World Bank or the governments they represent.

Columbia University accepts no responsibility or liability for any use that is made of this document or the results and information it includes other than by The World Bank for the purposes for which it was originally commissioned and prepared as stated in this document. Readers are cautioned against placing undue reliance upon any statements contained in it.

The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of The World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

Rights and Permissions

The material in this publication is copyrighted. Copying and/or transmitting portions or all of this work without permission may be a violation of applicable law. The International Bank for Reconstruction and Development/The World Bank encourages dissemination of its work and will normally grant permission to reproduce portions of the work promptly.

For permission to photocopy or reprint any part of this work, please send a request with complete information to the Copyright Clearance Center Inc., 222 Rosewood Drive, Danvers, MA 01923, USA; telephone: 978-750-8400; fax: 978-750-4470; Internet: www.copyright.com.

All other queries on rights and licenses, including subsidiary rights, should be addressed to the Office of the Publisher, The World Bank, 1818 H Street NW, Washington, DC 20433.

Acknowledgments

This technical report has been developed under a partnership between the Rockefeller Foundation and the World Bank Disaster Risk Financing and Insurance Program (DRFIP) to explore options for supporting the establishment of a regional disaster risk finance facility in Asia. DRFIP receives financial support from partners that include the Rockefeller Foundation, the Swiss State Secretariat for Economic Affairs, the UK Department for International Development, the European Union, the government of Japan, the government of Germany, and the Swedish International Development Cooperation Agency. DRFIP has provided advisory services on disaster risk financing and insurance to more than 60 countries worldwide.

The technical has been carried out by Columbia University under the guidance of a World Bank team comprised of Antoine Bavandi, Olivier Mahul, Benedikt Signer, and Emily White. The team from Columbia University was led by Upmanu Lall and Pietro Ceccato and comprised of Fabio Cian, Masahiko Haraguchi, and Maura Allaire, all from Columbia University.

Funding from the Rockefeller Foundation is gratefully acknowledged.

All the work related to this project can be found at www.financialprotectionforum.org/asiaregional

**Disaster Risk Financing
& Insurance Program**



Columbia Water Center
EARTH INSTITUTE | COLUMBIA UNIVERSITY



Contents

Overview	1
Objectives	5
Approach and Key Findings.....	6
References	8
Section 1. Economic Loss Estimation Using Remote Sensing	9
1.1 Summary	9
1.2 Introduction	9
1.3 Agriculture: Crop Loss Estimation.....	12
1.4 Illustrative Example: Rice Production Loss, Bangladesh 2015 Flood	13
1.5 Residential Sector: Loss Estimation	18
1.6 Conclusion.....	21
1.7 Future Work	21
1.8 References	27
Annex 1A. Flood Stage–Damage Curves for Buildings and Contents	30
Annex 1B. Flood Stage–Damage Curves for Agriculture.....	34
Section 2. Remote Sensing for Flood Area, Depth, and Exposure Mapping.....	41
2.1 Summary	41
2.2 Introduction	42
2.3 Identification of Relevant Case Studies in Southeast Asia and Corresponding Satellite Data	43
2.4 Flood Mapping by Means of Synthetic Aperture Radar and Optical Data.....	44
2.5 Estimation of Flood Occurrences	45
2.6 Population Exposure Assessment	53
2.7 Estimation of Surface Water Coverage from Optical Data from 1984 to 2015	56
2.8 Analysis of Vegetation Indices and Estimation of Anomalies in the Period of the Flood	57

2.9 Framework for Assessing Uncertainties	59
2.10 Conclusion.....	61
2.11 References	62
Annex 2A. SAR Missions Overview.....	62
Annex 2B. Data availability	63
Annex 2C. Flood Maps	71
Annex 2D. Flood Occurrences.....	75
Annex 2E. Flood Depth Estimation	76
Annex 2F. Airbus Defence - WorldDEM	79
Annex 2G. Flood Maps Cross-Comparison L8 – S1	81
Annex 2H. Improved Cross-Comparison of L8 and S1 Flood Maps.....	82
Annex 2I. Cross-Comparison of Flood Maps with NRT MODIS.....	83
Annex 2J. Cross-Comparison of GFMS Products with L8 and S1 Flood Maps.....	86
Annex 2K. Cross-Comparison of Population Data Sets	88
Annex 2L. Estimation of Yearly Surface Water Area	93
Section 3. Precipitation and Flood Return Period Analyses	96
3.1 Summary	96
3.2 Thailand.....	98
3.3 Bangladesh.....	101
3.4 Rainfall Analysis Summary	126
3.5 Flood Index Development.....	130
3.6 Future Steps: Machine Learning Applications to Storm Characteristic Identification, Event Return Period Estimation, and Automatic Infrastructure Identification	147
3.7 References	149
Annex 3A.	149

Overview

Global flood losses are expected to continue increasing and be highly concentrated in developing countries, especially in Asian megacities. Developing economies face significant challenges in responding to such disasters. The limited ability of the government to rapidly allocate funds for effective response immediately after flood disasters is a potential bottleneck of high consequence. In this report, we pilot one set of science informed strategies for overcoming such bottlenecks. The initial focus is on countries in Southeast Asia that are exposed to riverine floods, typhoons and tropical storms with significant consequences to the nation's economy and population. We sketch out and illustrate a methodology for risk and loss assessment in the context of planning a portfolio of financial instruments for rapid response at a national scale.

Rapid relief and response are essential in the aftermath of disasters, such as major floods, to reduce long term socio-economic impacts. This is especially critical in emerging market economies, such as those in South Asia and Southeast Asia, which have a chronic exposure to such events. Understanding how vulnerability and exposure to such events vary in time and space, and how the impact depends on socioeconomic, institutional and environmental factors is key for assessing societal risk at a national or regional scale. Economic development processes often lead to the concentration of assets and activity near traditional population centers and agricultural areas. These have historically evolved to be near waterways. The quality of the infrastructure varies, often being poorer in an area more prone to flooding. Consequently, the vulnerability and exposure to floods, especially to major flood events can be highly skewed by traditional development, and may come from direct losses of property and life, as well as impacts on health through increased incidence of water borne disease, the mobilization of pollution, and loss of access to water, energy, food, transportation, and health services. Losses in industrial and agricultural production and general business productivity can also contribute to reduced household and national income.

Recognizing the need for rapid response to such disasters, a variety of international organizations, including the World Bank, the Gates Foundation, the International Red Cross, and the World Food Program, have stimulated the development and testing of a variety of measures to make funds available, and also to invest in networks that facilitate the rapid identification of response and recovery needs. These include the deployment of sensor networks for emergency response, remote sensing capabilities to assess the evolution of the disaster, early warning systems, and the ability to rapidly mobilize water, food, transportation, evacuation, and financial resources. A variety of mechanisms, especially parametric mechanisms, such as index insurance, that do not require verification of loss, only the hydro-climatic magnitude of the event, have lately seen increasing use to help mobilize cash to support relief efforts.

Since the cost for physical risk mitigation efforts and disaster planning, as well as for the financial risk management mechanisms can be quite high, national or regional governments need to consider how best to make budgetary allocations for a portfolio that covers both types of instruments. This is especially important for growing economies with low per capita GDP, where national budgets may be limited and disaster related budgetary allocations as well as relief and recovery expenses compete with investments in social and economic development through infrastructure development and attraction of new industries and businesses.

South Asia and Southeast Asia provide examples of this situation. Thailand is of particular interest, since it is a relatively well developed economy in the region, and experienced an unprecedented, long-duration flood in 2011. Multiple tropical storms lashed the country over the period May-October, leading to regional flooding starting in Northern Thailand and eventually inundating the Bangkok Metropolitan Area. The World Bank estimated the losses due to flooding at approximately US\$46 billion. Most of these were related to the manufacturing industry located near Bangkok, where there was significant inundation. Supply chain impacts due to the loss of production were global, and were concentrated heavily in the auto and electronics industries. There were concerns that the multinational industrial presence in Thailand, which had accounted for a large share of the economic development in the country, may move out as a result. Despite the presence of a series of dams on the Chao Phraya River, the seasonal rainfall which was between 300% and 1,000% of the average annual rainfall, depending on the region, led to the long duration flood with significant impacts. The Thailand government has subsequently used a catastrophe bond mechanism to ensure access to finances if a similar event were to repeat, and has also had to invest in flood risk mitigation efforts. Relocation of facilities, private sector investment in structural risk mitigation, and adding redundancy to supply chains have also resulted. While other countries in the region may have a similar exposure to flood risk, they may not have the financial capacity to directly exercise similar options for risk mitigation or relief in the aftermath of the event.

Bangladesh is the second country considered in this pilot effort, given (1) its chronic propensity for flooding; (2) growing economy; (3) high exposure (70% of the land area flooded in 1998) and vulnerability to a major flood; and (4) the challenge posed to the development of a parametric risk product given the potential for flooding from rain over Bangladesh as well as from rivers that enter from India.

A *traditional approach* to flood risk and loss estimation as part of the development of products to address catastrophic risk, may consider an approach as indicated in Figure O.1 below:

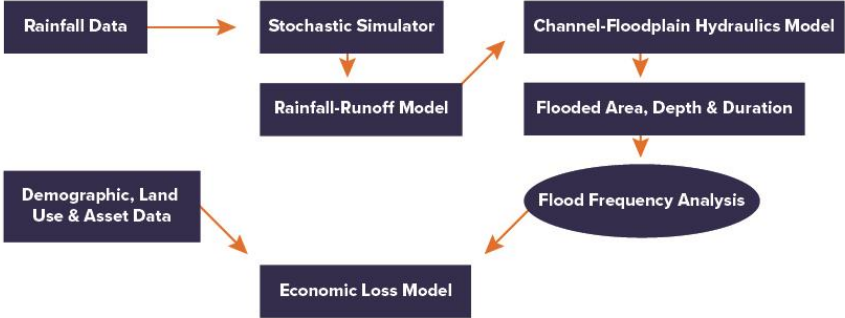


Figure O.1: Traditional approach to flood loss estimation through a sequential modelling process

Typically, 10-30 years of rainfall data from a set of rain gauging stations is used to represent daily rainfall over the river basin of interest. Recognizing that this is a short record, a stochastic model is developed to simulate a large number of years of rainfall. Using the historical rainfall and runoff data sets, typically at a few locations, a rainfall-runoff model is calibrated and used with the stochastic rainfall data. Sometimes, a detailed channel-floodplain hydraulics model is then used to move the flood wave through the system with presumably higher accuracy as to flood velocities and depths across the river basin and especially through areas considered to be at high risk. Finally, a frequency analysis is performed on the resulting peak annual flows to identify the T-year return period flow rate, depth, and velocity through the system. These are then integrated with asset and fragility data to develop economic loss-frequency or risk estimates. This is an intensive modelling effort, with significant attendant uncertainties and limitations, as discussed in (Delgado, Merz, & Apel, 2014) and also enumerated below.

1. The current understanding of climate variability and change suggests that there are significant epochal and quasi-periodic variations in rainfall, at inter-annual to decadal to multi-decadal scales (Jain & Lall, 2001). This means that any recent 10-30 years of data, even with stochastic modelling will not be representative of the long term odds of rainfall extremes and their variation. A rainfall record that spans at least a century may be more representative, and since these are not readily available in many developing country settings, proxy records such as climate reconstructions or re-analyses should be used to assess the long term risk and its variations.
2. The stochastic daily rainfall models typically used are typically Markovian and lack the ability to reproduce persistent rainfall from frontal or cyclonic systems that may lead to changing antecedent conditions for storms that may significantly change the flood potential in the region. This is even true when the output of these models is compared with the short, historical records they are fit to. Further, these models typically do not reproduce the upper tail of the probability distribution of either short duration (sub-daily) or persistent (multi-day) rainfall events very well, even though these are the types of events of primary concern for extreme flooding.
3. Data on streamflow are usually sparse and the records are usually short. Consequently, the calibration of the hydrologic and the hydraulic models is usually plagued by the curse of “equifinality” (Beven, 2006), i.e., many combinations of the model parameters give the same accuracy in model fitting, and that this redundancy is not easily resolved. There are two practical implications of this observation. The first is that there may be substantial error even after the best calibration, and given parameter uncertainty, it is important to consider these models as stochastic and not deterministic and generate simulations that span the range of potential parameter combinations to get a reasonable result. This is rarely done. The second follows from the fact that often there are very few, if any extreme floods of the kind that would generate a regional or national concern, in the data base used for model calibration. As a result, the “best” model parameters estimated are unlikely to be a good representation for the events of interest and the errors in inundation area and depth are hard to assess.

4. The flood frequency analysis performed with the simulated data is usually done point by point using the simulated flows at each location. The 100-year discharge event at each location of interest is selected and used for depth and loss estimation. However, the simultaneity of the occurrence of these putative 100-year return period events across the region is rarely assessed. Consequently, there is no reliable way to estimate the national loss associated with a 100-year event. In some cases, the loss for every flooding event is estimated and this regional or national loss series is then processed to identify the 100-year loss event. However, the uncertainty or representativeness of this loss amount is still difficult to assess given the earlier points. Finally, in the context of the design of a parametric financial product, even this assessment leaves the question as to what to use as the trigger for the payment associated with the parametric product open. If the loss magnitude itself is used, then a methodology for rapid, accurate and unbiased loss estimation, that is consistent with the way losses were derived for the modelling analysis needed. Moreover, this methodology has to be verifiable and trusted by both the purchaser and the provider of the parametric product.
5. Flood damage estimation has been receiving a significant amount of attention in the last decade as losses associated with catastrophic floods have increased. It is generally acknowledged that this is an area that has very high estimation uncertainty, especially in the developing country setting. (Merz et al., 2010) provide a comprehensive review of the state of the art, the information requirements and the process of structuring such estimates from asset inventories and other data sources, considering different types of short- and long-term impacts for the key sectors. They note that in general, damage data are rarely gathered, repair cost estimates are uncertain and are rarely updated, and most data available are heterogeneous, low quality, and often non-validated. Their comparative examples show that for the same setting flood damage estimates by different methods, just for building repair costs for two flooding events in Europe varied over one log cycle (2002 event) and two log cycles (1993 event).

Given these observations, it is natural to ask if especially for the problem context established for the current work, an approach that has much lower reliance on the traditional model chain is warranted.

Objectives

The objective of the research performed was to explore whether one could develop a strategy for a rapid financial response to a regional to national scale catastrophic flood event in a developing country setting (South and Southeast Asia). The strategy needs to consider the budgetary requirements for a disaster of consequence, i.e., one that requires the national or regional government to seek funds for rapid response. This necessitates an assessment of the probability distribution associated with the budgetary requirement. Second, one needs a readily measured trigger that would be a proxy for the event of consequence and can be rapidly assessed using data that is collected by a neutral party, i.e., neither the offeror of the contingent finance, nor the receiver. For the financial instrument to be priced, the probability of the trigger event needs to be established and related to the probability of the loss or the budgetary requirement.

The floods of interest are likely to be events that lead to a large inundated area, as has been the case for the most catastrophic floods in Bangladesh that affected agricultural and urban outcomes, and/or impact especially high-value economic activity, as was the case for the 2010 and 2011 floods in Thailand. Often, large-area flooding may also correspond to persistent and recurrent rainfall events and long duration of flooding. Thus, an appropriate index for a trigger event may include rainfall over an appropriate duration, the areal extent of flooded area, or the duration of flooding over a specific area. Note that the measure used by a hydrologist, e.g., river water level or discharge, would equivalently correspond to inundated area. However, while streamflow data are typically collected at a few points by regional or national governments, rainfall estimates or estimates of flooded area can be derived from international sources, and may hence provide better indices or triggers from the neutrality perspective. Even this observation takes us away from the traditional approach.

With this background the specific objectives pursued were to explore the following questions:

- Given a mapping of inundation, such as may be available from satellite information or other remote sensing, how can one rapidly assess at least the correct order of magnitude of direct and indirect flood damages across the key sectors?
- Given the variety of sources of satellite information, how best can one develop an accurate mapping of past and emerging flood events, covering both the area inundated, and the duration of inundation for each such event?
- Given a chronology of past flood events and their attributes, as identified from remote sensing, and/or archival sources, how can one effectively relate rainfall over different spatial domains, and time windows to the flood attributes (including damages where available)?
- How best can one integrate these three lines of exploratory inquiry into a strategy that identifies (1) an index that informs a trigger event; (2) the probability of occurrence of the trigger event; (3) the potential loss and payout associated with the trigger event, and (4) characterize the uncertainty across this process? *This strategy would inform both the country that is the target of the product and external parametric financial product designers for the design and pricing of appropriate instruments.*

Approach and Key Findings

The objectives indicated above were mapped into discrete tasks for economic loss identification, satellite based flood product development and index development, respectively. All analyses were conducted with Bangladesh and Thailand as case study targets for rapid prototyping.

Task 1: Economic Loss Estimation Using Remote Sensing

The approach developed and tested is described in Section 1 of the report. Standard methods from the literature for the estimation of agricultural and residential losses as a function of flooding depth and duration were tested with data calibrated to the conditions in Bangladesh and Thailand, assessed from satellite imagery for cropland and assets and estimated flood area. These were compared with reported estimates for selected events and estimates from the Government of Bangladesh. The rapid loss estimation method is shown to provide damage estimates that are of the same order of magnitude as the government and other reported estimates. However, agricultural losses in Bangladesh were overestimated by a factor of 3 to 5, possibly due to the assumption that the flooded area incurs a total loss of crop yield. Proposals for future research to better assess agricultural, residential, industrial, and infrastructure losses, as well as the economic value of remote sensing information for improving flood loss estimation are presented. From the perspective of the current project, the key finding is that given the high variability in event flood loss estimates using detailed methods, as identified in the literature, it is encouraging that the rapid loss estimation procedure does provide us with an estimate that would be in the range of the more detailed estimates. However, the use of in field surveys, information on the location of high-value assets from tax records, a longer history of concurrent remote sensing based flooded area and reported crop losses, and subsequent recovery expenditures by the public and private sector could substantially improve the accuracy of flood loss estimates and projections. Pooling data across countries in the region may offer the potential for making credible advances in this direction, to significantly advance the state of flood loss estimation.

Task 2: Remote Sensing for Flood Area, Depth and Exposure Mapping

The approach developed and tested is presented in Section 2 of the report. The main goals were to assess the feasibility of flooded area, depth and duration mapping using Synthetic Aperture Radar (SAR) and optical data from Landsat, Sentinel-1, and Sentinel 2; population and vegetation exposure assessment to support economic loss estimation; to develop an annual event time series from 1984 to 2015 for the Bangladesh and Thailand test cases; and to discuss and quantify the key sources of uncertainty. Based on data availability the 2015 flood in Bangladesh and the 2011 flood in Thailand were considered as targets for event analysis. Algorithms for flooded pixel classification using the satellite data were developed and implemented using the Google Earth Engine. Landsat and Sentinel-1 results compared favorably for the Bangladesh event. However, their agreement with areas reported by the Dartmouth Flood Observatory and with MODIS-based estimates was generally weaker, and this may be due to the processing algorithms used in each case, and the differences in the frequency and resolution of the products. Flood depth calculations are difficult since the freely available digital elevation models (DEMs) have rather coarse resolution. However, given that much of the terrain of concern in Bangladesh and Thailand is relatively flat, a median flood depth across the flooded area was computed as a proxy for the crop economic damage calculations. These estimates could be improved if a higher resolution DEM were available by making a fairly substantial payment, which was beyond the budget of our scoping grant. Three different population data sets were used to assess the population exposure in the flooded

areas identified. Substantial differences across these population products, each of which is built using different proxies and assumptions, were substantial, indicating that the uncertainty in these data sets may be at least as large as or larger than the intercomparison in the remote sensing estimates of flooded area. The normalized difference vegetation index (NDVI) estimates for vegetation were developed using traditional algorithms, and changes in the NDVI over the flooded pixels were assessed to develop an indicator for potential crop loss that was used in the economic loss assessment. The future plans include the development of an automatic algorithm for the processing of the satellite data to produce flood map time series, with an estimate of the probability of a pixel or region being flooded rather than just a binary classification, by integrating multiple sources of remote sensing information; and to improve the assessment of population exposure through better data sets.

Task 3: Precipitation and Flood Return Period Analyses

The primary goal of this task was to explore whether long (~ century-long records) of daily precipitation were likely to be informative for the design of a parametric financial product for catastrophic flooding. This entails the ability to identify a trigger event based on some statistic of rainfall; the verification that it does correlate well with flooded area, duration, and/or economic loss, at least for catastrophic events; and the estimation of the return period of both the trigger event from the long precipitation record, and the potential flood statistic (area, duration, or loss). The strategy was to consider rainfall spatially averaged over an entire river basin or country, and over some duration as the potential index. The design challenge then is to identify the spatial domain to average over, and the time window of averaging that best informs the flood statistic of interest (i.e., loss > threshold or area > threshold). This idea was explored for Bangladesh and for Thailand, noting that Bangladesh could be flooded by Ganges or Brahmaputra floods coming from the Indian part of the catchment, in addition to rainfall events in Bangladesh, while Thailand floods were largely due to endogenous rainfall. A variety of statistical methods and selection criteria were explored, and reliable associations between 5-day rainfall windows and 30-day rainfall windows and large floods were established, using precipitation data available from NOAA, and from the European Centre for Medium Range Weather Forecasting – i.e., data sources external to the countries. The Generalized Extreme Value Distribution with covariates for its location and scale parameters, selected from the two rainfall windows, and in the case of Bangladesh also from observed river water level data from three stations representing the three major rivers, was then used to develop a return period view of the flood impacts given these indices, with the idea that a single parametric index or a composite index using multiple indicators could be used as the trigger. This idea was extended by considering a T-year flood loss event and relating it to the proposed covariates to design an index that is a function of one or more precipitation and water level indicators such that the probability of exceedance or non-exceedance of the target flood loss was the design probability accounting for the uncertainty. Note this means that in the case of two covariates, e.g., p_5 = the five-day average rainfall over Bangladesh, and r_G the river water level at the station on the Ganges river, develop a flood index as $f = a_0 + a_1 p_5 + a_2 r_G$, and solve for a_0 , a_1 , and a_2 , and a threshold f^* such that when $f > f^*$, the probability with which the flood area or flood loss with the target return period T is exceeded is 0.5. The procedure is demonstrated to work reliably for the Bangladesh application where a historical record of flood losses was available from the national government. In future work, we propose to automate the diagnostic and model building work, and to explore Bayesian methods for uncertainty reduction for the rare event estimates and index design. An additional possibility is to consider pooling loss and climate data across all of Southeast Asia to reduce the uncertainty associated with the design of such instruments.

References

Beven, K. (2006). *A manifesto for the equifinality thesis*. *Journal of Hydrology*, 320(1–2), 18–36. <https://doi.org/10.1016/j.jhydrol.2005.07.007>

Delgado, J. M., Merz, B., & Apel, H. (2014). *Projecting flood hazard under climate change: An alternative approach to model chains*. *Natural Hazards and Earth System Sciences*, 14(6), 1579–1589. <https://doi.org/10.5194/nhess-14-1579-2014>

Jain, S., & Lall, U. (2001). *Floods in a changing climate: Does the past represent the future?* *Water Resources Research*, 37(12), 3193–3205. <https://doi.org/10.1029/2001WR000495>

Merz, B., Kreibich, H., Schwarze, R., & Thieken, A. (2010). *Review article “assessment of economic flood damage.”* *Natural Hazards and Earth System Science*, 10(8), 1697–1724. <https://doi.org/10.5194/nhess-10-1697-2010>

Section 1. Economic Loss Estimation Using Remote Sensing

Maura Allaire, Columbia Water Center, Columbia University

1.1 Summary

This report presents a methodology for using remote sensing to estimate economic damages. Maura Allaire developed this methodology and performed the following activities from November 2016 to May 2017.

Economic loss estimates serve as the trigger for the proposed financial instruments to alleviate catastrophic flood risk. Assessing economic losses accurately is crucial to inform disaster finance decisions. The overall goal is to produce loss estimates that inform relief actions as well as financial planning tools.

We developed a rapid assessment methodology for economic loss across regions and economic sectors, with a test case for Bangladesh. Our methodology builds on Damage and Loss Assessment (DaLa) techniques and accounts for flood characteristics via remote sensing, exposure via secondary data, and vulnerability through flood stage–damage curves.

We find that this methodology produces flood loss estimates within the same order of magnitude when compared to national government estimates. Yet, our flood loss values tend to be overestimated when relying on flood stage–damage curves. This overestimation will be addressed in subsequent work. The analyses highlight the need to assess vulnerability to flood impacts based on a mixed approach of remote sensing, statistical analysis of previous disaster losses, and machine learning.

1.2 Introduction

A rapid assessment method has been developed that models the distribution of flood impacts over the test areas in Bangladesh and Thailand. This impact assessment incorporates the flood hazard index and is designed to be executed rapidly and easily after a disaster. Rapid analysis requires sufficient historical data to produce a return period view of impacts. Overall, the rapid assessment methodology provides a measure of the severity of a flooding event.

Our loss assessment methodology builds on DaLa techniques developed by the UN Economic Commission for Latin America and the Caribbean (UN-ECLAC). Furthermore, we use techniques applied in World Bank rapid assessment reports as a launching point. The rapid assessment method uses remotely sensed data on flood characteristics in order to develop a framework that evaluates damage to building and contents as well as production losses.

The main sectors of interest are the agricultural and residential sectors. Particularly during short duration events, these two sectors can account for a larger portion of total losses, especially in lower-income regions. During previous events in Bangladesh, housing and agriculture represented the bulk of total losses for a short-duration event (Government of Bangladesh, 2008). Industrial and business sector losses can be more challenging to estimate via remote sensing. For example, assessment of industrial losses cannot rely on generalized flood depth–damage curves and require an understanding of individual facilities (Smith, 1994).

Economic loss assessment requires an understanding of exposure as well as vulnerability. Exposure is the total value of elements with potential to be affected by hazards (UNDP, 2004). A related concept is vulnerability—the capacity to anticipate, cope with, and recover from hazard impacts (Wisner et al., 2003). Exposure is assessed by accounting for population and assets in the study areas. Meanwhile, vulnerability is considered in flood depth–damage curves, which characterize the relationship between inundation and direct damages.

Figure 1.1 depicts our framework for loss assessment, which includes analysis of flood characteristics, exposure, and vulnerability. In particular, it shows how remote sensing enables economic losses from major flood events to be estimated. The framework is divided between remote sensing and economic analysis. Flood characteristics and exposure information obtained through remote sensing informs the economic loss assessment. Processed satellite images from Sentinel-1 and LANDSAT characterize flood extent, depth, and duration. Flood depth is approximated by combining estimates of flooded area with a digital elevation model (DEM). Exposure is assessed both through information from remote sensing (e.g., land use, population) and government data (e.g., assets in all four sectors considered). The exposure assessment considers the location, value, and timing of productive activities.

Processed satellite images from Sentinel-1 and LANDSAT characterize flood extent, depth, and duration. Flood depth is approximated by combining estimates of flooded area with a DEM. Population affected by flood events can be assessed with gridded population data (e.g., Global Human Settlement (GHS); WorldPop, Global Urban Footprint (GUF)). In future work, remote sensing could also be utilized to characterize land use, such as identifying agricultural areas and assessing building and infrastructure stocks in urban areas. Land use, when combined with secondary data on population and assets, enables exposure to be better characterized.

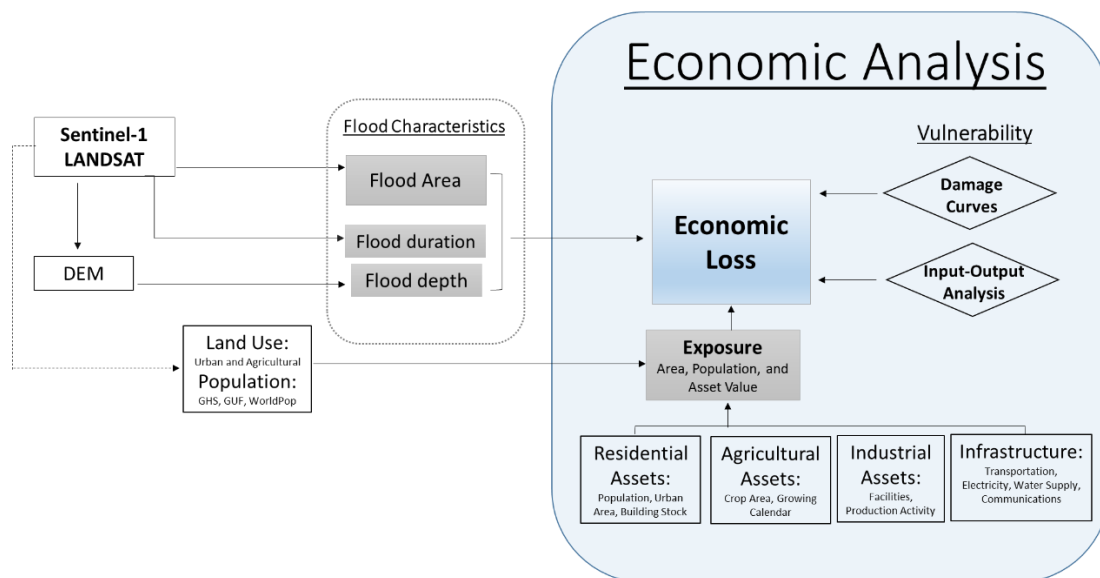


Figure 1.1. Conceptual framework for rapid financial response to natural hazards

More vulnerable assets will have a stronger response to increased flood stage. Depth-damage curves are a widely applied, standard method for estimating flood losses (Smith, 1994). In future work we propose to estimate indirect losses using input-output analysis, which accounts for vulnerabilities and quantifies the relationship between economic sectors.

Depth is partially important for assessing flood losses. A common approach for estimating flood losses is to rely on flood depth–damage curves. Stage-damage curves were originally developed in the U.S. by White (1945) and are now a widely applied, standard method (Smith, 1994). These damage functions relate hazard characteristics (e.g., flood depth, duration, flow velocity) to monetary damage for a particular asset type (Corti et al., 2009; Hammond et al., 2013; Parker et al., 1987; Smith, 1981; Wind et al., 1999). Vulnerability is considered in these relationships since more vulnerable assets will have a stronger response to increased flood stage than less vulnerable assets.

Damage functions rely on either empirical or synthetic data. Empirical data are actual damage values collected after a flood, such as via a survey or insurance claims records. Damage functions based on empirical data have been developed for many countries, including the U.S. (FEMA, 2003), Brazil (Nascimento et al., 2007), Japan (Dutta et al., 2003), and Germany (Merz et al., 2004). Synthetic data are derived from hypothetical analysis, which can include consideration of land use patterns, an inventory of assets, and/or survey questions to capture hypothetical mitigation behavior (Merz et al., 2010; Hammond et al., 2013). Synthetic analyses can include detailed surveys of building type and contents. An advantage of empirical data is that uncertainty and effects of mitigation measures can be measured (Kreibich et al., 2005; Thieken et al., 2008).

Guidelines for developing flood stage–damage curves tend to be available for industrialized countries (e.g., USACE, 1996; Penning-Rowsell et al., 2005; Messner et al., 2007). Yet, guidance and previous estimates for developing countries tends to be scarce due to a lack of quality damage data (Freni, Loggia, & Notaro, 2010). Flood stage–damage curves can be data-intensive since they require information on flood characteristics as well as damage. The relationships can either be in monetary or percent terms. In the case of percentage terms, damage is the percent of total depreciated value of the asset. Total damage can then be estimated by multiplying the percent damage by the total depreciated value. In some studies, replacement value is used instead of depreciated value; however, this can result in overestimating flood losses.

While flood stage–damage curves are a common and useful approach, they do have limitations. First, the useful lifespan of a given relationship is likely short. Asset values and susceptibility vary greatly over time and space (Merz et al., 2010). Several factors tend to be event-and location-specific, such as warning time, flood experience, and building type (Smith, 1994). Second, transferability of curves to other locations might be limited for similar reasons. For example, damage functions for a particular construction type are not similar across countries or time (Shah, 2011).

1.3 Agriculture: Crop Loss Estimation

Agricultural sector impacts are largely associated with production interruption. Interruption includes foregone and delayed production of agricultural goods. The initial assessment of agricultural impacts focuses on loss of rice crops. Past assessments in Bangladesh and Thailand have demonstrated that crops bear the overwhelming majority of sector impacts. For example, during the 2011 Thailand flood, rice comprised over 97% of crop losses (World Bank, 2012). Furthermore, rice is the dominant crop in Bangladesh, accounting for about 94% of total crop production (Kwak et al., 2015a).

Loss in rice production is estimated as the value of reduced yield attributable to the flood event. The extent of loss is particularly dependent on the growth stage during which the crop is inundated. Crop losses are estimated using the following information: crop area that is inundated, plant growth stage, typical yield per area, and crop price. Crop loss for each geographic unit i and crop type k is estimated as:

$$\text{Crop Loss}_{ik} = \text{Price}_k A_{ik} \text{Yield_noflood}_{ik} D_{ikt}, \quad (1)$$

where Price = Price of crop k per unit weight;

A = cultivated area of crop k that is inundated in geographic area i ;

Yield_no flood = yield per area of crop k in situation without a flood; D = damage (as a percent of yield) for crop k in time period t ; as estimated via a

flood stage–damage curve.

The crop price is the country-specific, monthly producer price reported in the FAOSTAT data set of the Food and Agricultural Organization (FAO).¹ Changes in price attributable to the flood event are not considered during this phase of the project. Past studies have treated crop prices as unaffected by flooding (e.g., Dutta et al., 2003; NRCS, 1986). Cultivated area and yield of a given geographic unit are obtained from county-specific data sets.^{2,3} Damage is estimated using flood stage–damage curves that are specific to each crop.

The yield in the situation without flooding is estimated based on regression analysis of time series of rice yields at the sub-country level.⁴ Past time series data make it possible to estimate what production would have been had the disaster not occurred. It is important to evaluate agricultural production

¹ Alternatively, province-level weekly prices are available for Thailand: Office of Agricultural Economics (OAE), "Thailand Agricultural Prices." Annual prices are available for Bangladesh; see BBS (2016). Generally, international prices are appropriate for exports (ECLAC, 2003).

² In Thailand, crop area and yield are available for years 2003 and 2012 from the Thailand Agricultural Census. Additionally, production costs are available from OAE, "Thailand Cost of Production Survey."

³ In Bangladesh, crop area and yield are available for years 2008 and 2012–2015 from the Bangladesh Bureau of Statistics (BBS). An agricultural census was conducted in 2008 (BBS, 2008), while statistical yearbooks are available for other years (e.g., BBS 2016).

⁴ Care must be taken to identify drought and flood years. For example, Thailand had drought in 2008 and late 2014 to 2016.

impacts based on divergence from the scenario if no disaster occurred. Comparing conditions with and without the disaster event is not equivalent to measuring production before and after the event. Pre-flooding patterns of crop production and prices are assessed.

Considering the time period and crop growth stages is important for agricultural flood stage–damage curves. Seasonal timing of a flooding event will considerably influence agricultural impacts. If flood occurs when planting has just begun, then there is the possibility to plant a second crop and avoid a complete loss. However, if a flood occurs when a crop is ready for harvesting and it is no longer feasible to sow a new crop, then a total loss is possible.⁵

1.4 Illustrative Example: Rice Production Loss, Bangladesh 2015 Flood

As an illustrative example, the methodology presented above is implemented for the 2015 flood in Bangladesh. Flooding lasted from May to September 2015 and affected parts of Bangladesh, India, and Pakistan. Bangladesh regularly experiences flooding, driven by heavy monsoon rainfall and poor drainage. Even in years with relatively normal rainfall, about 20% of the country is flooded (Mirza, 2002).

The 2015 flood was similar in magnitude to the 2004 and 2007 floods. About 42% of total land area was inundated, which damaged an estimated 1.3 million hectares of crop land (Table 1.1). The 2004 and 2007 floods have estimated return periods of 12 years and 14 years, respectively. Yet, the largest flood event to occur in the past 40 years was in 1998, when 69% of total land area was inundated (Figure 1.2). This massive event is estimated to have a return period of 90 years.

We focus on production loss for rice, given that rice accounts for 94% of the total crop production in Bangladesh (Kwak et al. 2015a). Furthermore, agriculture is the largest economic sector in Bangladesh, accounting for more than 16% of GDP, and employs 45% of the labor force (BBS, 2016).

Data

Secondary data and processed information from remote sensing were assembled at the district (zila) level.

Agricultural data on crop area by district, yield by crop type, crop calendar, meteorological data, and losses associated with flooding are available through the Bangladesh Bureau of Statistics (BBS) Yearbook of Agricultural Statistics. Remote sensing allows us to estimate flood extent, depth, and duration at the district level.

⁵ If a crop is completely destroyed, the costs must be estimated according to the crop growth stage. If destruction or damage is partial, estimates must be adjusted.

Year	Affected area ('000 km ²)	% total land area	Affected Pop. (millions)	Crops damaged (million ha)	Return period (years)
1974	53	37%	30	.	9
1987	57	40%	30	.	13
1988	90	63%	47	2.1	55
1998	100	69%	31	1.7	90
2004	56	39%	33	1.3	12
2007	62	42%	14	2.1	14
2015	61	42%		1.3	

Table 1.1. Bangladesh Floods 1974-2015: Flood Damages and Return Period

Data Source: BBS, 2016; Dasgupta et al. 2010

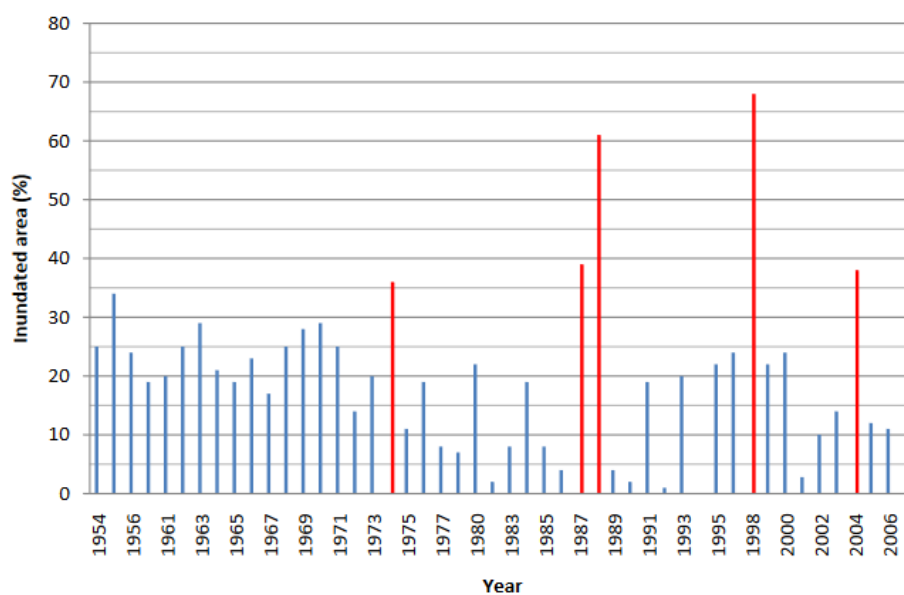


Figure 1.2. Bangladesh floods 1954-2007: Inundated area

Methods

Flooded crop area is estimated based on remote sensing output and cultivated area from Bangladesh Bureau of Statistics (BBS, 2016). Remote sensing provides information on the portion of each district that is inundated. Meanwhile, agricultural statistics provide information on cultivated area by district and rice variety for the year 2008, the most recent agricultural census year. We assume that the portion of crop area within each district devoted to a given rice variety remains fixed. Therefore, we calculate the area under cultivation for each district and rice variety using these assumed proportions and national-level rice cultivation data for growing season 2014-2015. To obtain flooded crop area for each district and rice variety, we multiply the percent inundated area of the district by area under cultivation.⁶

Then, crop loss is estimated using Equation 1. Percent reduction in yield is based on stage-damage functions. Rice varieties in Bangladesh are assumed to have similar flood stage–damage functions, which are based on Kwak et al. (2015b), displayed in Annex 1B. These flood stage–damage curves consider flood depth and duration.

Rice Varieties and Growing Season

We also must consider seasonal timing of the flood event, since extent of crop loss will depend on growth stage. The three main varieties of rice are Aman, Aus, and Boro. Each of these types have local and high yield sub-varieties. Each of the main crops is associated with a growing season. Boro rice grows from late November to early May, during the dry season, and depends on irrigation. Boro is the highest-yielding rice variety, but its production can be disrupted by pre-monsoon flash floods in April and May, which are infrequent. Boro accounts for about 42% of total rice yield and 55% of total rice area (BBS, 2016).

The Aus season lasts from March to August, which means that late harvesting can be vulnerable to monsoon floods, since its harvesting season is July–August (BBS, 2016). Aus rice only accounts for a small portion of total rice production, and it’s the lowest-yielding variety grown in Bangladesh. Meanwhile, Aman rice grows from June to December and relies on monsoon rains. Aman accounts for 48% of total rice yield and 38% of total rice area (BBS, 2016). Table 1.2 depicts the crop calendar and shows how the three growing seasons overlap.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Aus Rice			Sowing				Harvest					
Aman Rice					Sowing							
Boro Rice	Sowing			Harvest								

⁶ This assumes that cultivated areas are evenly distributed in a given district. In future work, land use information obtained via remote sensing will allow us to conduct the loss assessment at the sub-district level in order to more accurately estimate flooded crop area.

Table 1.2. Crop Calenda, Bangladesh Rice Varieties
Data source: Nelson et al. 2014

Given that the 2015 Bangladesh flood occurred in mid-May to September, only Aus and Aman rice should have been affected. Boro rice harvesting ends by mid-May. Aus rice was inundated during the entire post-sowing season and harvest period. Meanwhile, Aman rice was flooded during the typical sowing period and most of the growing period. Farmers would not have been able to plant a second crop of either Aus or Aman that could be harvest before the end of the year. Therefore, both Aus and Aman are included in the production loss assessment, which is based on stage–damage functions.

Results

Our demonstration of the crop loss methodology achieves estimates that are the same order of magnitude as government assessments. Yet, our estimates are approximately two to four times higher than values estimated by the Government of Bangladesh. The flood stage–damage curves tended to overestimate flood losses in the agricultural sector. This can be partially attributed to the coarse resolution of cultivated land area information available from the national government. In addition, inundation does not necessarily cause loss, yet flood stage–damage curves tend to make this assumption.

We find that crop production losses due to the 2015 Bangladesh flood are modest. Results are presented for flooded crop area, yield reduction, and value of production loss. In terms of flooded crop area, only a small portion of total crop area in Bangladesh was inundated. Three divisions out of seven total in the country were estimated to have some portion of crop area flooded. Dhaka was the division with most crop area inundated (1,668 km² of rice area), while Sylhet had 713 km². The relative magnitude of production loss is similar to that of flooded crop area, since mean flood depth is about 1 m in most districts. Mean flood depth does not vary considerably across districts, since publicly available DEMs can only provide depth with an accuracy of one meter. In some districts, flood depth and duration are not of large enough magnitude to be considered to cause crop damage. Mean flood depth must be at least one meter, due to the low resolution of the DEM. Furthermore, flood duration of two days or less is not considered to cause significant crop damage.

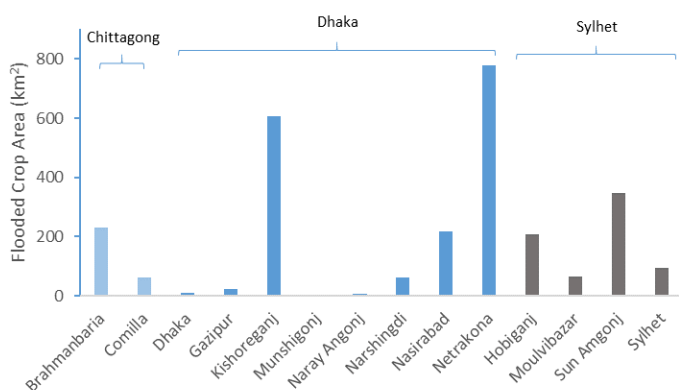


Figure 1.3. Flooded crop area, by division (province) and zila (district)

Division	Flooded Crop Area		Production Loss ('000 metric tons)
	km ²	% total crop area	
Chittagong	232	1.2%	47
Dhaka	1,668	6.1%	352
Sylhet	713	8.2%	114

Table 1.3. Flooded Crop Area and Production Loss, by Division

The value of production loss is estimated to be modest at the national level (US\$140 million). Yet, two districts bore two-thirds of this total. These two districts are in Dhaka division: Netrakona (US\$53 million) and Kishoreganj (US\$49 million). At the national level, Aman rice was the most affected, with yields reduced by 468,000 metric tons (4% of total production) (Table 1.4). The value of Aman production loss is estimated to be US\$132 million, assuming a price of US\$0.28 per kg (BBS, 2016). Meanwhile, lost Aus rice production is estimated to be 45 thousand metric tons (2% of total production), with a value of US\$8 million, assuming a price of US\$0.18 per kg (BBS, 2016). The value of total lost rice production is estimated to be 0.1% of GDP.

	Production Loss (‘000 M.tons)	% Production Loss	Value of Production Loss (million USD)	% GDP
Aus Rice	45	1.9%	\$8	0.0%
Aman Rice	468	3.6%	\$132	0.1%
Boro Rice	-	-	-	-
Total	513	1.5%	\$140	0.1%

Table 1.4. National Production Loss, by Rice Variety

Comparison with Government of Bangladesh

When comparing our estimates to those of the Government of Bangladesh, we find that the crop loss estimates have the same order of magnitude. Our estimate of lost rice production is 513 million metric tons, while the Government of Bangladesh calculates 134 million metric tons (Table 1.5). Yet, our estimates are substantially higher. Our estimates for Aus rice production loss are approximately four times greater, while our Aman rice estimates are more than twice as high. Overestimation will be addressed through improvements in methodology in Phase 2. Improvements are especially needed regarding estimation of flood depth (through higher-resolution DEM) and estimation of flooded crop area below the district level.

	Production Loss (‘000 M.tons)	% Production Loss
Aus Rice	8	0.4%
Aman Rice	125	0.9%
Boro Rice	-	-
Total	134	0.4%

Table 1.5. Government of Bangladesh: Estimate of Crop Loss due to 2015 Flood

Data source: BBS, 2016

1.5 Residential Sector: Loss Estimation

Household impacts focus on building and content damage, for our initial methodology development. The magnitude of loss is influenced by flood characteristics (e.g., flood depth) as well as dwelling characteristics (e.g., location, building quality). This approach of evaluating building and contents damage relies on flood stage–damage functions. A damage function relates hazard characteristics (e.g., flood depth, flow velocity) to monetary damage for a particular asset type. In order to empirically estimate building and content damage, we must understand the pre-disaster housing stock and expected losses. Useful housing stock characteristics include number of buildings in a given area, quality of buildings (e.g., material, age), building size, construction costs, and typical contents (e.g., furnishings and equipment).

Structural damage in the residential sector to building type k in each geographic unit i would be estimated as:

$$Res\ Loss_Structural_{ki} = ND_{ki} FA_{ki} Value_per\ area_{ki} DS_{ki} \quad (2)$$

where ND = number of residential dwellings of building type k in geographic unit i ;

FA = floor area of building type k in geographic unit i ;

$Value_per\ area$ = average replacement value of a residential dwelling per unit area;

DS = structural damage (as a percent of replacement value) for building type k in geographic unit i ; as estimated via a flood stage–damage curve

Content damage can be calculated separately, if information on building height and content value per household are available:

$$Res\ Loss_Content_{ki} = ND_{ki} Value_{ki} DC_{ki} \quad (3)$$

where ND= number of residential dwellings;

Value = average replacement value of a residential contents;

DC = content damage (as a percent of replacement value) for building type k in geographic unit i ; as estimated via a flood stage–damage curve

DaLa methods recommend first collecting information on numbers of buildings in a given area and grouping into categories (e.g., single-family houses, apartments). Then, number of dwellings and floor area can be obtained through remote sensing or government data. Floor area data are available for Thailand in Shah (2011). Where floor area is not available through secondary data, it can be approximated via satellite imagery. For example, Dutta et al. (2003) approximated floor area as the portion of urban land use multiplied by fraction of urban land occupied by buildings. In order to value buildings, replacement cost is recommended in DaLa methods, rather than depreciated value, since it is more readily obtainable (ECLAC, 2003). A common metric is average construction costs per floor area.

Structural damage as a percent of replacement value is estimated via flood stage–damage curves. In this way, our method goes beyond DaLa methodology since we consider how flood characteristics (such as floodwater depth) will influence the extent of damage. In contrast, DaLa methodology defines damage categories (e.g., full, partial) as a portion of replacement cost. These categories do not consider how variations in flood characteristics caused different magnitudes of loss. For informal dwellings, replacement values tend to be treated as equal to the cost of the most basic units in government housing programs (ECLAC, 2003).

Flood depth is considered through use of flood stage–damage curves. These curves can be developed for specific building types (Thieken et al., 2008; Penning-Rowsell et al., 2010). Moreover, recent efforts have demonstrated that flood loss prediction can slightly improve (by about 5% for building damage and 12% for content damage) when considering household income and grouping households accordingly (e.g., Wahab and Tiong 2016). Remote sensing allows us to estimate flood depth and to estimate the number of buildings inundated at a certain flood depth.

Data

Provincial-level population is available for both Thailand and Bangladesh. In the next round of study, we could develop a district-level or sub-district level data set. We rely on a variety of sources for population estimates, including international data sets (e.g., nationmaster.com) as well as national sources (e.g., Thailand National Statistical Office).⁷ Accurate estimates of population are important for assessing flood losses in the residential sector. For example, population may need to be used to approximate number of households and dwellings since Thailand does not have a housing census. Number of residential dwellings can be estimated as $ND_i = \frac{Pop_i}{HH_size_i}$ since this information is not provided in the Shah (2011) GIS data set.⁸

Building, population, and infrastructure inventory information for Bangkok is available from Shah (2011), based on data obtained from the Bangkok Metropolitan Administration.⁹ This GIS data set (at the 500m grid level) includes buildings, floor area, and replacement values for years 2010 and 2030.

Flood Stage–Damage Curves

In the case of Thailand, we have detailed data on household losses.¹⁰ In Bangladesh, we rely on previous stage-damage curves that have been developed for several localities. For example, as shown in Annex 1A, stage-damage curves have been developed for Khulna City for buildings and contents in five sectors: (1) residential; (2) commercial; (3) industrial; (4) manufacturing; and (5) roads (ADB, 2010).

In Thailand, detailed data were collected during two rounds of household surveys after the massive 2011 flood event (Nabangchang et al., 2015). These data are used to develop flood damage curves that account for both flood depth and duration. This survey work was conducted with 469 households in the Bangkok Metropolitan Area and collected information on economic costs incurred due to the 2011 flood, housing characteristics, and socioeconomic status. This study finds that most property damage during the 2011 Thailand flood was borne in the form of contents damage. Households tended to incur little structural damage, which concurs with the findings of the World Bank's rapid assessment of this event (World Bank, 2012). This might be attributable to slow-moving floodwater and/or the types of construction materials in Thailand (e.g., concrete; simple wooden frames).

⁷ Some local officials in Thailand believe that NSO estimates are too high (World Bank, 2012).

⁸ This approach to approximating number of dwellings was also used in the World Bank rapid assessment of the 2011 Thailand flood (World Bank, 2012). In 2010, the average size of a Thai household was 3.2, while in 2000 it was 3.8.

⁹ This data set was previously used by the World Bank to estimate a hazard index for 30-year and 2-year events. Our study differs from this previous effort in that we address flood events larger in magnitude than the 30-year flood and we utilize flood stage–damage curves.

¹⁰ In addition, province-specific household information is available in a GIS data set from Michael Bauer Research, www.arcgis.com/home/item.html?id=8658f09cf1ae4a44aaa813fbb687b59f.

Moreover, regression analysis finds that flood depth can explain a significant amount in the variation in losses of non-poor households (Annex 1A, Table 1A.2). The regression analysis results in the following association between depth (in cm) and damages (in US\$) for households in neighborhoods not classified as poor:

$$\ln(\text{Structural and Content Damage}) = 6.6 + 0.01\text{Depth} + 0.001\text{Duration} \quad (3)$$

Duration was found to not be significantly associated with household damages, which concurs with past studies from Southeast Asia (e.g., Wijayanti et al., 2015; Sahasakmontri, 1989; Tang et al., 1992). This might be due to the fact that the 2011 flood was a long-duration event or because housing structures tended to be constructed of resilient materials such as concrete or simple wooden frames.

1.6 Conclusion

Economic loss estimates serve as the trigger for the proposed financial instruments to alleviate catastrophic flood risk. Assessing economic losses accurately is crucial to inform disaster finance decisions. The overall goal is to produce loss estimates that inform relief actions as well as financial planning tools.

We developed a rapid assessment methodology for economic loss across regions and economic sectors, with a test case for Bangladesh. Our methodology builds on Damage and Loss Assessment (DaLa) techniques and accounts for flood characteristics via remote sensing, exposure via secondary data, and vulnerability through flood stage–damage curves.

We find that this methodology produces flood loss estimates within the same order of magnitude when compared to national government estimates. Yet, our flood loss values tend to be overestimated when relying on flood stage–damage curves. This overestimation will be addressed in subsequent work. The analyses highlight the need to assess vulnerability to flood impacts based on a mixed approach of remote sensing, statistical analysis of previous disaster losses, and machine learning.

1.7 Future Work

We found that the flood stage–damage curves used tended to overestimate reported flood losses in the agricultural sector, by a factor of two to three. This may be partially attributed to the coarse resolution of asset data available from national governments. In addition, inundation does not necessarily cause loss, yet flood stage–damage curves tend to make this assumption. The use of remote sensing for asset identification at sub-district level should help reduce the uncertainty associated with losses. Remote sensing will allow for more detailed crop mapping, to understand the distribution of crop area within a district, and to account for the crop stage during the growth season. Dwelling and building assets can be identified via remote sensing and zoning maps and/or property tax maps will be used in order to identify areas of high value and low value.

The scope of the loss assessment will include:

- The development of methodology for direct losses, based on statistical analysis of historical flood losses and stage-damage curves;
- Methodology for indirect losses, based on input-output analysis;
- Vulnerability profile for each country to identify susceptible sectors and regions;
- Analysis of the value of this remote sensing application for disaster assessment.

Surprisingly, the methodology for estimating the economic loss of flood events has not advanced much over the last several decades. Although there have been major theoretical and methodological advances in nonmarket valuation techniques in the environmental and resource economics field, these have not been incorporated into assessments of flood loss (Allaire, 2015; Nabangchang et al. 2015). This is partially due to engineering studies dominating the flood loss assessment literature. Engineering studies tend to make a distinction between tangible and intangible components based on the extent to which the consequences of the flood can be expressed in monetary terms (Thieken et al., 2005; Dutta et al., 2003). Yet, nonmarket valuation techniques can quantify many loss categories in welfare-theoretic terms (Smith, 2004).

Our loss methodology will produce monetized estimates of direct and indirect flood loss, across multiple sectors. Direct losses result from physical contact of floodwater with people and assets. Meanwhile, indirect losses are incurred by individuals and assets that are not inundated. Sectors included are agriculture, residential, industry, and infrastructure. The full impacts of flooding include damage to buildings and other assets, production interruption, and public service interruption.

Task 1. Economic Loss Methodology: Direct Losses

A rapid assessment method will be developed to model the distribution of flood impacts across regions and economic sectors. During Phase 1, we began development of an economic loss assessment methodology that can estimate the distribution of flood impacts over selected countries in Southeast Asia. This impact assessment is designed to be executed rapidly and provides a measure of the severity of a flooding event.

Improvements for Agricultural Sector

In future model development, we will develop an alternative crop loss assessment method that primarily relies on remote sensing output. This serves as an alternative to the current method that uses flood stage–damage curves. To do so, we will assess the statistical relationship between reductions in yield and normalized difference vegetation index (NDVI) anomalies. A tobit regression model will be used to model the percent reductions in yield for geographic area i and crop k in month t (D_{ikt}). This model will include all years of the observable record, both flood and non-flood years. Yield reduction will be represented as a function of NDVI anomalies, cumulative precipitation deficit, excess rainfall, and previous season crop prices. Developing a relationship between NDVI anomalies and reduced yield would allow the rapid assessment methodology to mostly rely on remote sensing information when assessing future flood events, rather than government estimates of crop production. In addition, as an alternative to NDVI, changes in fluorescence will be assessed, where available. Yield reduction estimates from past government assessments, flood stage–damage curves, and the statistical techniques relying on NDVI anomalies will be compared to determine which approach is preferable.

Additional improvements for crop loss assessment include land use maps to identify crops. Furthermore, the cost of interruption will be estimated as the reduction in value added, which is equal to the market value of output, less the value of inputs. In the case of crop production, this is the market value of the crop, less the cost of inputs that include labor, capital, and materials. Alternatively, value added can be the accounting profits and depreciation expenses, wages, and taxes.

Improvements for Residential Sector

In future work, residential structural damage for a variety of building types will be estimated as a function of number of dwellings of a given type, floor area, average replacement value per unit area, and structural damage estimated via a flood stage–damage curve (as a percent of replacement value).

Zoning maps and/or property tax maps will be used in order to identify areas of high-value and low-value properties. In addition, accounting for future land use change and value of assets will be critical. Future population affected will be estimated based on expected growth rates from CIESIN (2013). Meanwhile, future building stocks and values will be based on expected scenarios developed by national governments and/or the World Bank. With rapid development in Southeast Asia, the value of exposed assets is growing. For example, between 2010 and 2030, Bangkok is anticipated to undergo a 41% increase in population and building replacement value, while infrastructure replacement value will more than double (Shah, 2011).

Poverty mapping can identify particularly vulnerable households. Through machine learning we will develop relationships between flood loss and construction material as well as socioeconomic indicators based on a unique dwelling-level data set we created for Bangkok. Through in-person surveys with 469 Bangkok households, we have dwelling-level information on construction material, socioeconomic status, and flood losses during the 2011 Thailand flood. This data set will be used to develop relationships between construction material, socioeconomic indicators, and flood losses. In this way, we can characterize dwellings that tend to be inhabited by lower-income households and building susceptibility to flood damage.

In addition, since flooding does not necessarily cause loss, a probabilistic representation of impacts within a variety of sectors, based on historical data from national ministries and/or Munich Re, is needed. Emerging machine learning methods may also improve the identification of flood loss as it

relates to socioeconomic indicators, along with image recognition of physical assets and a unique dwelling-level survey data set we created for Bangkok.

Development of Industrial Sector

Both property and production losses are incurred as a result of flooding in the commercial and industrial sector. Key industries in the study countries include garment and tourism. For property damage, insurance claims from Munich Re and/or national ministries can be used for firms with insurance coverage. Flood stage–damage curves can be developed based on these empirical data. These relationships will be a contribution since previous estimates for developing countries tend to be scarce due to a lack of quality damage data (Freni, La Loggia, and Notaro, 2010). Flood stage–damage curves can be data-intensive since they require information on flood characteristics as well as damage.

Historical loss data will also be crucial for developing a return period view of flood impacts. Based on cross-sectional analysis, we'll be able to test the performance of the index against historical event occurrence and modify the index accordingly.

Development of Infrastructure Sector

Key infrastructure will be identified based on remotely sensed data and secondary information from national ministries. Infrastructure will include transportation (e.g., highways, bridges, railways, airports), electric power plants and grid, water supply, and communication systems. Economic losses associated with infrastructure include damage to physical structures as well as interruption of public services. Public service interruptions arise from damaged infrastructure and can create considerable disruptions in daily life. For example, transportation interruptions can result in increased travel costs, while electricity outages could necessitate expenses of running a generator. Losses caused by public service interruption have the potential to be much greater than physical infrastructure damage (Brozović et al., 2007).

A framework for economic analysis for impacts resulting from temporary suspension of infrastructure services or damage to infrastructure will be developed. This work represents a significant contribution since limited past work has addressed links between structural damage and economic losses of service interruptions across numerous infrastructure types (Kelly, 2015). Since inundation does not necessarily cause structural damage or service interruption, we will develop a probabilistic representation of infrastructure impacts, based on historical data from national ministries and/or Munich Re.

Task 2. Economic Loss Methodology: Indirect Losses

Indirect losses are incurred by individuals and assets that are not inundated. These are especially relevant for the commercial and industrial sector. The direct loss assessment accounts for both property damage and production interruption due to direct inundation. Higher-order effects are one type of indirect loss that can influence production due to changes in demand or supply. Demand for products can fluctuate due to lost income and flood expenses, while supply will vary with costs of production. Accounting for higher-order impacts is crucial, since as property damages increase, secondary economic impacts can grow exponentially (Hallegatte, 2008).

Input-output (IO) analysis will account for higher-order impacts on intermediate and final consumption. By linking production in each sector with intermediate and final demand, we will account for inter-industry relationships. National IO tables from the Organisation for Economic Co-operation and Development (OECD) and/or national government will be used. With this analysis, we will gain insight into how a shock to one sector can propagate to others, due to changes in final demands and availability of intermediate goods and services.

By addressing both direct and indirect losses, this project will develop a framework for comprehensive decision support to inform disaster interventions. Understanding of efficient policies will be advanced in two ways: (1) flood loss assessment that includes production losses and indirect losses, and (2) evaluation of how intervention decisions change when more comprehensive flood impacts are considered. Currently, disaster management decisions do not account for the full, societal costs of flooding disasters (Allaire, 2015). Therefore, benefits are limited to avoided property damage. This raises several policy-relevant questions. To what extent will decision outcomes change when broader impacts are considered? Should avoided losses that extend beyond property damage be incorporated into World Bank and member government decision-making methodology?

Task 3. Vulnerability Profile Analysis

Based on direct and indirect loss estimates, a vulnerability profile will be created for each of the countries included in the project. The vulnerability profile will identify which sectors and regions are especially susceptible to flooding impacts. In addition, it can characterize the main drivers of loss and uncertainty. Therefore, this analysis can identify which assets are most critical and should be prioritized for emergency response. Sensitivity analysis will be conducted in order to simulate how economic losses change based on the extent of flood severity for a given sector and asset class.

In addition, poverty mapping can identify particularly vulnerable households. We can develop flood vulnerability maps that consider poverty, based on a unique data set we created for Bangkok. Through in-person surveys with nearly 500 Bangkok households, we have dwelling-level information on construction material, socioeconomic status, and flood losses during the 2011 Thailand flood. This data set will be used to develop relationships between construction material, socioeconomic indicators, and flood losses. In this way, we can characterize dwellings that tend to be inhabited by lower-income households and buildings susceptible to flood damage.

Task 4. Examine the Value of Remote Sensing Applications for Disaster Assessment

This project offers an important opportunity to examine how remote sensing tools can inform decisions regarding disaster response. Furthermore, improvements in decision-making can be valued. Therefore, we will be able to estimate the societal benefits of this rapid assessment tool. The value of information is an emerging topic in the earth science field and is critically important for informing investment in observational instruments and programs.

One way to assess the value of remote sensing applications is to compare interventions that would have occurred with and without remote sensing information. For example, in the context of flood aid, we can compare disbursements of flood compensation with and without information obtained from our rapid assessment tool. Disaster aid programs face both errors of inclusion (Type I error) and exclusion (Type II error). Type I errors represent a loss to government budgets since aid is disbursed to those who are not affected by the flood event. Meanwhile, Type II errors are a loss to households who are affected by flooding, but do not receive aid.¹¹ Our rapid assessment tool will also allow for more timely information regarding flood severity and economic losses across sectors. This could enable more immediate emergency response actions, which could reduce further losses.

Through retrospective analysis of past flooding disasters, we will assess how flood aid disbursements might change with additional information from the rapid assessment tool. Are there reductions in errors of inclusion and exclusion? If so, to what extent?

Furthermore, we can value different observing technologies and data resolutions, which is helpful to determine appropriate expenditures on data for governments, international organizations, and the private firms. For example, Sentinel-1 satellite observations offer an advantage over LANDSAT due to its capability to make observations regardless of cloud cover. These differences in information between observing programs are especially crucial for major flood events, which tend to coincide with precipitation events and cloud cover. Subsequent differences in flood characterization will be accounted for as well as differences in emergency response decisions.

In addition, we will value different resolutions of remote sensing information. High-resolution data, such as digital elevation models, can be costly. Therefore, it is important to consider which data resolutions are justifiable, considering the resulting improvements in disaster response efforts. At different data resolutions, we will assess the change in accuracy of flood loss estimates as well as changes in errors of inclusion and exclusion for flood compensation programs.

¹¹ For these households, the compensation amount represents a lower-bound of their loss because denial of aid could cause households to incur additional impacts (e.g., health effects, greater expenditure on emergency supplies).

1.8 References

- Asian Development Bank (ADB). (2010). *Strengthening the Resilience of Water Sector in Khulna to Climate Change*. Manila: ADB.
- Asian Development Bank (ADB). (2015). *Bangladesh: Capacity Building for Disaster Risk Finance*. www.adb.org/sites/default/files/project-document/173391/42249-013-tacr-01.pdf;
- www.adb.org/projects/documents/capacity-building-disaster-risk-finance-final-report-tacr.
- Bangladesh Bureau of Statistics (BBS). (2008). *Structure of agricultural holdings and livestock population, Volume 1*.
- Bangladesh Bureau of Statistics (BBS). (2012). *The Statistical Year Book of Bangladesh*. Statistics Division. Ministry of Planning.
- Bangladesh Bureau of Statistics (BBS). (2016). *Yearbook of Agricultural Statistics: 2015*.
http://203.112.218.65/WebTestApplication/userfiles/Image/LatestReports/Yearbook_2015.pdf.
- BWDB (Bangladesh Water Development Board). (2014). *Observed river data and reports of 1988, 1998, and 2004 floods*. <http://www.bwdb.gov.bd>.
- CIESIN (Center for International Earth Science Information Network)-Columbia University. (2013). *Low Elevation Coastal Zone (LECZ) Urban-Rural Population and Land Area Estimates, Version 2*. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).
- Crow, H. (2014). *Assessment of the FEMA HAZUS-MH 2.0 Crop Loss Tool: Fremont County, Iowa 2011*. M.S. Thesis. University of Southern California.
- Economic Commission for Latin America and the Caribbean (ECLAC). (2003). *Handbook for estimating the socio-economic and environmental effects of disasters*. Mexico City, Mexico.
- Foster, E. E. (1941). *Evaluation of flood losses and benefits, Proceedings: American Society of Civil Engineers*.
- Government of Bangladesh. (2008). *Cyclone Sidr in Bangladesh: Damage, Loss, and Needs Assessment for Disaster Recovery and Reconstruction*. Dhaka, Bangladesh.
- Islam, A. S., Haque, A., & Bala, S. K. (2010). *Hydrologic characteristics of floods in Ganges–Brahmaputra–Meghna (GBM) delta*. *Natural Hazards*, 54:797–811.
- Japanese International Cooperation Agency (JICA). (1985). *Master plan on flood protection/drainage project in eastern suburban Bangkok, Main report conducted for Bangkok Metropolitan Administration, Department of Drainage and Sewerage*.
- Kwak, Y., Arifuzzaman, B., & Iwami, Y. (2015a). *Prompt Proxy Mapping of Flood Damaged Rice Fields Using MODIS-Derived Indices*. *Remote Sensing* 7, 15969–15988.
- Kwak, Y., Gusyev, M., Arifuzzaman, B., Khairul, I., Iwami, Y., & Takeuchi, K. (2015b). *Effectiveness of Water Infrastructure for River Flood Management: Part 2–Flood Risk Assessment and Its Changes in Bangladesh*. *Proceedings of the IAHS*, 370: 83–87.

Merz, B., Kreibich, H., Theiken, A., & Schmidtke, R. (2004). Estimation uncertainty of direct monetary flood damage to buildings. *Natural Hazards and Earth System Sciences*, 4, 153-163.

Messner F., Penning-Rowsell, E., Green, C., Meyer, V., Tunstall, S., & van der Veen, A. (2007). *Evaluating flood damages: guidance and recommendations on principles and methods*. UFZ, European Community's Sixth Framework Programme. Leipzig, Germany.

Mirza, M. M. Q. (2002). Global warming and changes in the probability of occurrence of floods in Bangladesh and implications. *Global Environmental Change*, 12:127–138.

Nabangchang, O., Allaire, M., Leangcharoen, P., Jarungrattanapong, R., & Whittington, D. (2015). Economic Costs Incurred by Households in the 2011 Bangkok Flood. *Water Resources Research*, 51(1)58-77.

Nelson, A., Boschetti, M., Manfron, G., Holecz, F., Collivignarelli, F., Gatti, L., Barbieri, M., Villano, L., Chandna, P., & Setiyono, T. (2014). Combining Moderate-Resolution Time-Series RS Data from SAR and Optical Sources for Rice Crop Characterization: Examples from Bangladesh. In Francesco Holecz, Paolo

Pasquali, Nada Milisavljevic, and Damien Closson (Eds.), *Land Applications of Radar Remote Sensing*. DOI: 10.5772/57443.

Penning-Rowsell, E., and Chatterton, J. (1979). *The Benefits of Flood Alleviation: A Manual of Assessment Techniques*. Gower Technical, Aldershot.

Penning-Rowsell, E., Johnson, C., Tunstall, S., Tapsell, S., Morris, J., Chatterton, J., & Green, C. (2005). *The benefits of flood and coastal risk management: a handbook of assessment techniques*. Middlesex University Press, London.

Penning-Rowsell, E., Viavattene, C., Pardoe, J., Chatterton, J. B., Parker, D. J., & Morris, J. (2010). *The benefits of flood and coastal risk management: a handbook of assessment techniques 2010*. Middlesex University, Centre for Higher Education Research, London.

Sahasakmontri, K. (1989). *Estimation of flood damage functions for Bangkok and vicinity*. Master's thesis. Asian Institute of Technology, Bangkok.

Scawthorn, C. et al. (2006). "HAZUS-MH Flood Loss Estimation Methodology II, Damage and Loss Assessment."

Shah, F. (2011). *Methodology report: calculating multi hazard city risk*. Working Paper. World Bank. Washington, DC. <http://documents.worldbank.org/curated/en/720221468247885684/Methodology-report-calculating-multi-hazard-city-risk>.

Siebert, S., Henrich, V., Frenken, K., & Burke, J. (2013). *Global Map of Irrigation Areas*. Rheinische Friedrich-Wilhelms-University: Bonn, Germany; Food and Agriculture Organization (FAO): Rome, Italy. www.fao.org/nr/water/aquastat/irrigationmap/index10.stm.

Smith, D. I. (1994). Flood damage estimation—A review of urban stage-damage curves and loss functions. *Water SA*, 20(3), 231–238.

Tang, J. C. S., Vongvisessomjai, S., & Sahasakmontri, K. (1992). Estimation of flood damage cost for Bangkok. *Water Resources Management* 6,(1), 47–56.

Tateishi, N., Hoan, N., Kobayashi, T., Alsaaidh, B., Tana, G., & Phong, D. (2014). Production of Global Land Cover Data – GLCNMO2008. *Journal of Geography and Geology* 6, 99–122, doi:10.5539/jgg.v6n3p99, 2014.

Thieken, A. H., Olschewski, A., Kreibich, H., Kobsch, S., & Merz, B. (2008). Development and evaluation of FLE-MOps—a new Flood Loss Estimation Model for the private sector. In Proverbs, D., Brebbia, C. A., & Penning-Rowse, E. (Eds.), *Flood recovery, innovation and response*. Southampton: WIT Press, 315–324.

UNDP (United Nations Development Program). (2004). *Reducing disaster risk: A challenge for development*. John S. Swift Co.

U. S. Army Corps of Engineers (USACE). (1996.) *Risk-based analysis for flood damage reduction studies: engineering manual*. Washington, DC.

Wahab, R., and Tiong, R. (2016). *Multi-variate residential flood loss estimation model for Jakarta: an approach based on a combination of statistical techniques*. *Natural Hazards*.

Wijayanti, P. et al. (2015.) *Estimation of River Flood Damage in Jakarta: The Case of Pesanggrahan River*. *Natural Hazards*.

Wathier, C.-M. (2014). *Probabilistic Evaluation of Flood Damage in Buildings*. MS Thesis. MIT

Wisner, B., Blaikie, P., Cannon, T., & Davis, I. (2003). *At risk: Natural hazards, people's vulnerability and disasters*. London: Routledge.

WFP (World Food Program). (2010). *Bangladesh Food Security Monitoring Bulletin Issue No. 2, August–October*.

Yang, C. R., and Tsai, C. T. (2000). Development of a GIS-based flood information system for floodplain modeling and damage calculation. *Journal of the American Water Resources Association*, 36(3), 567–577.

Annex 1A. Flood Stage–Damage Curves for Buildings and Contents

Khulna City, Bangladesh (ADB, 2010)

Depth damage curves were developed for five land use categories. The categories were (1) residential; (2) commercial; (3) industrial; (4) manufacturing; and (5) roads. Figure 1A.1 depicts curves for the residential and manufacturing sectors.

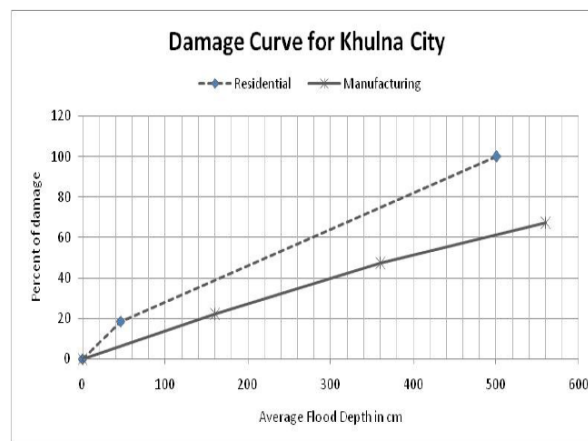


Figure 1A.1. Flood Stage–Damage Curves for Khulna City, Bangladesh

Source: ADB, 2010

Bangkok, Thailand (Sahasakmontri, 1989; Tang et al., 1992)

Damage curves were developed as a function of depth and duration, based on a 1983 flood event. This event caused 6.6 billion baht in damage (National Statistical Office; cited in Lekuthai and Vongvisessomjai, 2001). In this study, data were obtained from surveys with 1,041 households (described in Sahasakmontri, 1989) and cost functions are used from JICA 1985. Contents in buildings were the major component of damage.

$$Damage_{Residential} = -300 + 45.4Depth + 33.8L$$

Where Damage per HH in Thai baht; Depth= max. flood depth in cm; L= duration of flood event (Sahasakmontri, 1989; Tang et al., 1992)

Based on this method, 4.48 billion baht in damage (where US\$1 = THB 25.5) were estimated for the residential sector in Bangkok (Tang et al., 1992).

ESTIMATION OF FLOOD DAMAGE COST FOR BANGKOK

Table III. Estimated flood damage in 1983, eastern and western Bangkok

Sector (1)	Damage in eastern Bangkok (2)	Damage in western Bangkok (3)
Residential	4 479 634 000	51 340 370
Commercial	703 963 600	27 871 030
Industrial	643 100 800	38 386 660
Agricultural	1 658 872 000	279 449 700
Total	7 485 570 400	397 047 760
Grand total	7 882 618 160	

US\$ 1.00 ≈ 25.5 baht.
For details, see Anusinha (1989).

Table 1A.1. Total Estimated Damage for 1983 Flood

Source: Tang et al., 1992

Jakarta, Indonesia (Wijayanti et al., 2015)

Residential damage estimates were calculated for a three-day flood event in January 2013. Flood loss estimates included direct cost (structural; content) and indirect cost (loss of income; evacuation; labor cost; cost of illness; clean-up). Note: the majority of average flood damage per household was contents: structure (14%), contents (60%), indirect (26%). Regressions by income group indicated that wealthier households have greater variance of flood damage explained by covariates, which agrees with the flood stage–damage analysis for the residential sector presented in our report.

Residential flood damage (per household, in thousand Indonesian rupees):

$$\text{Damage} = -14.2 + 13 \text{ FD} + 4.5 \text{ DUR} + 0.197 \text{ INC} + 8 \text{ FA} + 2.7 \text{ DIS} + e$$

where FD= max. flood depth (cm); DUR= duration (hrs); INC= HH income (IDR/month); FA= house area (m²); DIS=distance from house to the nearest river (m)

FIMA Residential Depth-Damage Curves (FEMA, 2003; Wathier, 2014)

Federal Insurance and Mitigation Administration (FIMA) has developed flood depth-damage curves for the U.S. In the residential sector, losses include both structure and contents losses, and are determined relative to actual cash value (depreciated replacement cost). Structural and contents damage are estimated separately, then summed to obtain total residential damage. Structural damage is based on total flood area for a certain census block, while contents damage is based on number of dwellings. Housing types include one-story, two or more floors (with or without basement), split level (with or without basement), and mobile home.

The HAZUS manual does not provide the depth-damage functions. Wathier (2014) interpolates the equations for the FIA credibility-weighted building depth-damage curves issued by the NFIP Actuarial Information System (1998).

Structural Damage:

One-story building, no basement:

If depth=0, damage=0; If depth >9 feet: damage= 0.5

If depth>0 & depth<9 feet:

$$y = -1E-06*d^5 + 7E-05*d^4 - 0.0013*d^3 + 0.0091*d^2 + 0.0138*d + 0.1669$$

Two-story (or more) building, no basement:

If depth=0, damage=0; If depth >14 feet: damage= 0.5

If depth>0 & depth<14 feet:

$$y = 0.0005*d^2 + 0.0219*d + 0.0954$$

Contents Damage:

Content value is divided between the first and second floor in the HAZUS model (60% on first floor; 40% on second). A floor height of 10 feet is assumed by Wathier (2014). The general model is Contents Damage=fxn (d_c, V_c, story, c), where f_{story}= contents loss (L_c) for each floor and a difference value (V_c) can be assigned to each content Loss (L_c).

First, calculate how many floors are affected by flood (assuming each story is 10ft):

$$L_c = \text{fxn} (\text{depth of flood on floor } c; \text{ value of contents on floor } c)$$

Then, total content loss:

$$\text{SUM across floors } (L_c * V_c)$$

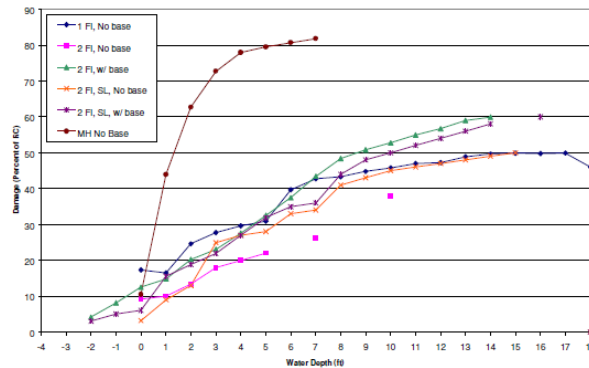


Figure 5.2 FIA Credibility-Weighted Building Depth-Damage Curves as of 12/31/1998

Figure 1A.2. Residential loss curves (structural and contents damage): Flood depth (ft) vs. % depreciated replacement cost

Source: FEMA (2003)

Bangladesh: Cyclone Sidr in 2007 (World Bank, 2008)

Data for housing sector losses are based on DaLa methods.

Housing type: Household Income and Expenditure Survey (2005) classified housing into four categories based on quality: pucca (permanent quality): 8%; semi-pucca: 84%; kutcha (mud brick): 1%; and shanty (hay/bamboo): 7%.

Bangkok: 2011 Flood

Columbia Water Center's regression analysis of residential sector flood damage.

For households in neighborhoods not classified as poor:

$$\ln(\text{Structural and Content Damage}) = 6.6 + 0.01\text{Depth} + 0.001\text{Duration}$$

where Damage is in terms of US\$; Depth in cm; Duration in days

Results for non-poor:

```
. reg ln_damage_tot_usd depth_road duration_soi if poor==0
```

Source	SS	df	MS	Number of obs	=	220
Model	84.6637084	2	42.3318542	F(2, 217)	=	37.82
Residual	242.864214	217	1.11918993	Prob > F	=	0.0000
				R-squared	=	0.2585
				Adj R-squared	=	0.2517
Total	327.527922	219	1.49556129	Root MSE	=	1.0579

ln_damage_~d	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
depth_road	.0108656	.0012543	8.66	0.000	.0083934 .0133378
duration_soi	.0014896	.0072406	0.21	0.837	-.0127813 .0157606
_cons	6.609876	.476654	13.87	0.000	5.670412 7.54934

Results for non-poor:

```
. reg ln_damage_tot_usd depth_road duration_soi if poor==1
```

Source	SS	df	MS	Number of obs	=	246
Model	1.04599402	2	.522997012	F(2, 243)	=	0.29
Residual	433.314085	243	1.78318553	Prob > F	=	0.7461
				R-squared	=	0.0024
				Adj R-squared	=	-0.0058
Total	434.360079	245	1.77289828	Root MSE	=	1.3354

ln_damage_~d	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
depth_road	.0008169	.0017553	0.47	0.642	-.0026407 .0042744
duration_soi	.0027956	.005373	0.52	0.603	-.007788 .0133793
_cons	6.110433	.4042206	15.12	0.000	5.31421 6.906656

Table 1A.2. Flood Stage–Damage Curves for Bangkok

Annex 1B. Flood Stage–Damage Curves for Agriculture

Flood Stage–Damage Curves for Rice in Bangladesh (Kwak et al., 2015b)

Damage curves were estimated for a flood event that occurred in July to September in 2007 in the Ganges, Brahmaputra, and Meghna basin. Rice damage curves were developed based on crop calendar (WFP, 2010), crop areas estimated from satellite images from the Global Land Cover data set (Tateishi et al., 2014), and the 2014 household survey of Bangladesh Water Development Board (BWDB, 2014).

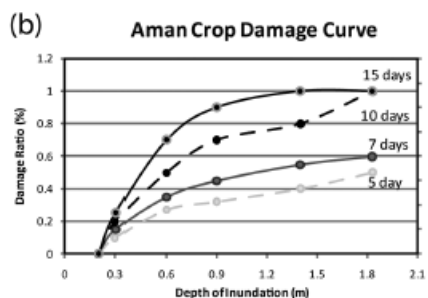


Figure 1B.1. Flood stage–damage curves for Aman rice

The crop damage curve considers flood depth and duration. Rice damage begins once floodwater is 0.3m deep and increases in damage begin to flatten once floodwater depth reaches 1m. Aman rice is the most relevant rice variety for this 2007 flood event because it is harvested in July and August, which is when the flood event occurred.

We approximate the flood stage–damage curves for Aman rice as:

<u>Flood duration</u>	<u>Stage-Damage Function</u>
5 days	% Damage = 0.2300*ln(Depth) + 0.3492
7 days	% Damage = 0.2828*ln(Depth) + 0.4576
10 days	% Damage = 0.4643*ln(Depth) + 0.7083
15 days	% Damage = 0.5099*ln(Depth) + 0.8306

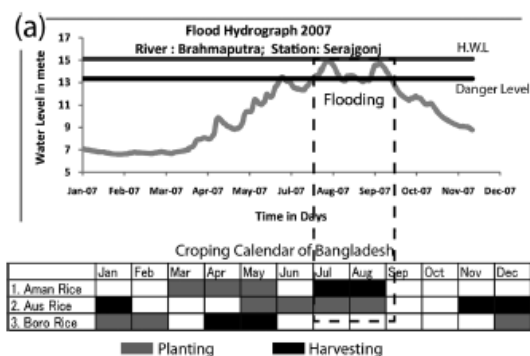


Figure 1B.2. Hydrograph in 2007 and crop calendar for Bangladesh rice varieties

Data source: WFP, 2010

Used data	Raster resolution	Source
River discharge	20 km (10 min)	BTOP model (2014)
Topographic data	90 m to 0.45 km (15 arcsec)	HydroSHEDS (2008)
Land cover	0.45 km (15 arcsec)	GLCNMO (2008)
Population data	1 km (30 arcsec)	LandScan (2009)
Water level & Stage-damage curves	84 gauge stations	BWDB (2014)

Table 1B.1. Data Sources

Bangkok, Thailand (Sahasakmontri, 1989; Tang et al., 1992)

Damage curves were developed as a function of depth and duration, based on a 1983 flood event. In this study, data were obtained from surveys with 1,018 farmers (described in Sahasakmontri 1989) and cost functions are used from JICA (1985).

$$Damage_{Agriculture} = -1047.2 + 553.5Depth$$

where Damage is per establishment (in Thai baht); Depth= max. flood depth in cm

Sahasakmontri (1989) finds that duration of flood event does not have a significant association with agricultural damage.

Bangladesh: Relationship between Flooded Area and Depth (Yang et al., 2015)

Yang et al. (2015) estimated relationship between inundated area and flood depth in Bangladesh, using a data set from 1950s-2006. The authors do not find evidence that find that multivariate flood damage function has an advantage over a single variable. Flood-affected area (FAA) was estimated as a function of flood characteristics (FC) such as maximum daily water level (MDL) and duration (DAL)). DAL is calculated as the number of days that water level is above its dangerous water level line.

$$FAA = \frac{Max.Flood Area}{(1+MinFC \cdot e^{-G(FC-CFC)^{1/\epsilon}})}$$

Ichinomiya River Basin, Japan (Dutta et al., 2003)

Damage curves were developed for a four-day flooding event in Japan in September 1996. Both depth and duration are taken into account. Data include cultivated area, growing season, yield per area, and market value per unit weight. Agricultural losses were relatively small during this event because the flood mostly affected urban areas and harvesting of rice was completed several weeks before flooding occurred. The damage curves developed for rice are shown below:

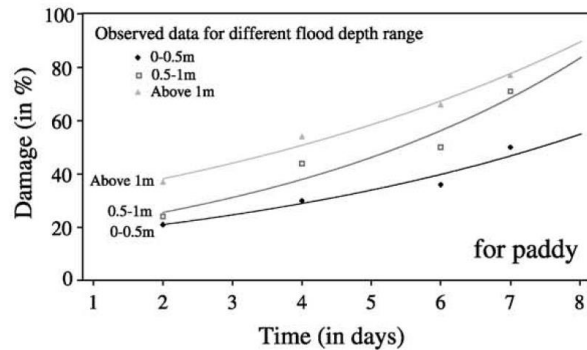


Figure 1B.3. Paddy rice damage curves for Ichinomiya River Basin, Japan

Source: Dutta et al., 2003

Within the time period considered (eight days), damage does not reach 100%. In order to estimate total crop losses, Dutta et al. (2003) use the following equation:

$$Crop\ Loss_i = \sum_k Value_k A_{ik} Yield_{ik} D_{ik} m_k$$

where Value= Value of crop k per unit weight;

A = cultivated area of crop k in geographic area i ;

Yield = typical yield per area of crop k ;

D = damage (as a percent of yield) for crop k ;

m = loss factor, which depends on period of growing season

Thailand 2011 Flood, Crop Loss Estimates (World Bank, 2012)

The World Bank conducted a rapid assessment of the 2011 Thailand flood, which included impacts on crop production. Since this flood event persisted for several months, most crops were completely destroyed. Losses for each crop type k in province i were estimated as:

$$\text{Crop Loss}_{ik} = \text{Value_Production}_{2011_k} A_{ik} \text{Yield}_{2011_{ik}} \% \text{Loss}_{ik}$$

Data on previous year yields and cultivated area (A) were obtained from the Thailand Office of Agricultural Economics (OAE). Crop losses are displayed in Table 1B.2.

Agricultural land was mostly affected in the Northeast and Central regions. Affected crops included rice, sweet corn, vegetables, and fruit trees. Rice was especially affected since wet season rice is planted in August and harvested in November and December. Therefore, fields were flooded during the entire growing season. Although in the past, farmers in flood areas grew floating rice, now higher-yielding varieties are preferred. Affected crop area, by region, is provided in Table 1B.3. Furthermore, two crops each year are now planted, whereas in the past there tended to be only one harvest. For example, farmers in flood-prone areas in the Central Region (e.g., Ayutthaya) grow rice earlier so that they can harvest before the flood arrives. These quick-growing varieties can be harvested in late August. However, the 2011 monsoon season (and subsequent flooding) arrived much earlier than normal.

Table 8:
Agriculture sector – Estimated losses in 2011–2014 by commodity, in Thai baht, millions

Of the 26 assessed provinces, the total losses to the fisheries sub-sector are estimated to be roughly THB 809 million in 2011. Table 8 summarizes the total losses, disaggregated by commodity for each year from 2011 to 2014.

	2011	2012	2013	2014
Crop	27,522	1,927	1,204	482
Rice	26,645	0	0	0
Sugar cane	288	0	0	0
Fruit trees	227	1,927	1,204	482
Flowers	144	0	0	0
Maize	16	0	0	0
Other	202	0	0	0
Livestock	1,997	0	0	0
Poultry	1,159	0	0	0
Swine	46	0	0	0
Cattle	722	774	0	0
Other	70	0	0	0
Fisheries	809	0	0	0
Tilapia	319	0	0	0
Catfish	256	0	0	0
Shrimp	234	0	0	0
Striped snakehead fish	1	0	0	0
Total	30,328	2,701	1,204	482

Source: Estimates based on data from MOAG and field survey
 *Note: Total Losses in 2011 consists of three components: (1) Total production losses (2) Higher production cost and (3) Losses due to forced early sale. Total losses in 2012–2014 consist of losses of production of permanent trees, and losses of meat and egg production due to death of animals.

Table 1B.2. Agricultural Sector Losses: Thailand 2011 Flood

Source: World Bank, 2012

Region	Affected provinces	Affected crops (rai)	Affected Livestock (heads)	Affected Fishery (rai)
Central	14	1,844,327	2,148,271	142,039
North	4	3,249,466	3,202,310	16,668
Northeast	8	2,220,912	846,330	18,418
Total	26	7,314,705	6,196,911	177,124.82

Source: Ministry of Agriculture and Cooperatives, Thailand. www.moac.go.th, accessed on November 21, 2011

Table 1B.3. Affected Crop Area, Livestock, and Fisheries: Thailand 2011 Flood

Source: World Bank, 2012

Table 9 summarizes the total damage and losses in the agriculture sector covering all three sub-sectors in 26 provinces.

Table 9:
Agriculture sector – Damage
and losses by province in
Thai baht, millions

Province	2011			2012	2013	2014
	Damage	Losses	Total	Losses	Losses	Losses
Ang Thong	53	297	350	55	0	0
Ayuthaya	552	3,906	4,458	97	4	2
Bangkok	81	788	869	10	0	0
Chachoengsao	248	1,962	2,210	99	34	14
Chai Nat	21	166	187	25	4	1
Kalasin	9	55	64	26	8	3
Khon Kaen	759	2,612	3,370	229	125	50
Lop Buri	89	883	972	123	6	2
Maha Sarakham	22	86	107	3	0	0
Nakhon Nayok	19	172	191	3	0	0
Nakhon Pathom	71	593	664	37	8	3
Nakhon Sawan	102	461	563	217	92	37
Nonthaburi	37	362	399	4	0	0
Pathum Thani	189	1,645	1,834	47	22	9
Phitsanulok	299	1,258	1,557	53	28	11
Phichit	77	323	401	42	8	3
Prachinburi	54	476	530	27	6	3
Roi Et	1,383	6,335	7,718	248	128	51
Samut Sakhon	31	344	374	300	187	75
Saraburi	108	1,184	1,292	57	9	4
Singburi	8	112	120	20	0	0
Si Sa Ket	864	3,621	4,485	682	421	168
Suphan Buri	111	792	902	71	1	0
Surin	249	1,124	1,373	17	3	1
Ubon Ratchathani	189	609	798	135	69	28
Uthai Thani	41	163	204	72	41	16
Total	5,666	30,328	35,994	2,701	1,204	482

Source: Estimates based on data from MOAC and field surveys

Table 1B.4. Agricultural Sector Losses by Province (Source: World Bank, 2012.)

	Total Loss (USD, millions)	% Damage
Manufacturing	32,852	71%
Finance & Banking	3,760	8%
Tourism	3,093	7%
Housing	2,733	6%
Agriculture	1,317	3%
Transport	994	2%

Table 1B.5. Comparison of Agricultural Losses with Other Sectors, Thailand 2011

Section 2. Remote Sensing for Flood Area, Depth, and Exposure Mapping

Pietro Ceccato, Fabio Cian

2.1 Summary

This report summarizes the activities performed by the remote sensing team (Pietro Ceccato and Fabio Cian) from November 2016 to May 2017. The team performed the following activities:

Activities

- Identification of relevant case studies in Southeast Asia and corresponding satellite data
- Flood mapping by means of Synthetic Aperture Radar (SAR) and optical data
- Estimation of flood occurrences
- Estimation of flood depth and statistic report at administrative level
- Statistical comparison of flood maps derived from SAR and optical data
- Comparison of flood map with other products, both satellite and model-based
- Population exposure assessment
- Estimation of surface water coverage from optical data from 1984 to 2015
- Analysis of vegetation indices and estimation of anomalies in the period of the flood
- Framework for assessing uncertainties

The team identified as useful case studies the floods in Thailand in October/November 2011 and the one in Bangladesh in the period June-September 2015. The two case studies are different in terms of hydrology, environment, and data coverage. The main difference is in terms of data availability, with the flood in Bangladesh covered by the new Sentinel-1 (S1) SAR mission (launched in late 2014), which made it possible to have a large amount of data covering the event in comparison to the past.

2.2 Introduction

The aim of the remote sensing team was to show the potential of new Earth Observation (EO) data, both optical and radar, for flood risk assessment, regarding: (1) flood mapping (extent, occurrence, and depth), and (2) people exposed to disasters and assets impacted. The results obtained were then provided to the rest of the team for the assessment of economic impacts and for the analysis of the return period of the floods analyzed.

Earth Observation (EO) data, in particular Synthetic Aperture Radar (SAR), have been used for decades for flood mapping. Many reliable and well-consolidated methodologies have been developed to extract floods from these data. For instance, the European Copernicus Emergency Management Service¹² and the International Charter on Space and Major Disaster¹³, two of the major operative emergency response centers, are regularly providing flood maps in response to emergencies using SAR data (Martinis, Kersten, & Twele, 2015).

SAR data are particularly useful in case of floods. SAR microwaves can penetrate clouds, allowing the monitoring of land in almost any weather conditions. Having its own source of illumination, SAR can also acquire useful data during nights and without being affected by the relative position of the sun (O'Grady, Leblanc, & Gillieson, 2011; Waisurasingha et al., 2007; Wilson & Rashid, 2005). In case of flood, the smooth surface of the water (compared to the wavelength) reflects the signal to a specular direction. Therefore, the return to the satellite will be theoretically null, letting the pixel representing water be very dark and easily classifiable.

One of the biggest limitations in flood mapping using EO data is the availability of images over the area hit by the event and acquired during the days it happened. Satellite revisiting time of older missions can be up to 30 days or more, limiting the probability of having useful images over the area of interest.

In 2013, with the launch of the Landsat-8 satellite, followed by the launch of Sentinel-1 (S1) (in 2014) and Sentinel-2 (in 2015), the era of EO big data started. The probability of having useful data for monitoring disasters has increased enormously, and the accessibility of data has become easier, given that many data sets, such as the ones cited, are freely available.

In particular, the European Space Agency (ESA) S1 radar satellite constellation is able to map the entire Earth every six days (with interferometric capabilities), giving an unprecedented opportunity to access a large number of archived scenes, which is of key importance to detect changes and assess economic impacts in case of disasters. The constellation is made of two satellites: Sentinel-1A (S1A), launched in April 2014 and operational since October of the same year, and Sentinel-1B (S1B), launched in April 2016. The large volume of data provided by the constellation opens new frontiers in image processing, making it possible to perform statistical analysis on long time-series of data and to develop new approaches in change detection analysis.

¹² <http://emergency.copernicus.eu/mapping/>

¹³ <https://www.disasterscharter.org>

Based on an innovative flood mapping technique developed by Fabio Cian (Cian et al., 2017a under review), we made use of a normalized index based on a statistical analysis of SAR multi-temporal series. The index is computed by means of a comparison of a large amount of reference scenes to those acquired during the investigated flood. The index easily allows the categorization of “flooded” areas: either areas temporarily covered solely by water or areas with mixed water and vegetation (shallow water in short vegetation). The methodology has been developed specifically to exploit S1 data but can be applied to any other sensor. The index exhibits very good performances and improved simplicity in flood mapping compared to other methods, which are usually more complex and strongly user-dependent.

From the precise flood extent maps, by means of a high resolution digital elevation model (DEM), water level was derived looking at the elevation along the contours of the flooded areas (Cian et al., 2017b, submitted). Elevation values along the contours of each flooded area were extracted and statistically analyzed. The statistical analysis estimated the elevation of the water plane. Elevation along the contour of a flooded area should be the same. In practice, different values of elevation can be found due to imprecisions in the flood map or in the DEM, and a non-perfect alignment of the two data sets. The higher the resolution of the available DEM, the more precise is the water depth that can be derived. For the area analyzed, we disposed only of the Shuttle Radar Topography Mission (SRTM) DEM at 30-meter horizontal resolution and 1-meter vertical resolution. The suboptimal data allow a coarse estimation of flood depth, nevertheless useful for impact assessment.

2.3 Identification of Relevant Case Studies in Southeast Asia and Corresponding Satellite Data

Freely available SAR and multi-spectral data covering past flood events were searched during this project for many countries in Southeast Asia, among others Malaysia, Thailand, and Bangladesh. In particular, data were acquired from ESA’s Sentinels and ENVISAT missions and United States Geological Survey’s Landsat (5, 7, and 8). Commercial SAR data were explored, such as German Aerospace Center’s TerraSAR-X (TSX) and the Italian Space Agency’s COSMO-SkyMED (CSK). Table 2A.1. in Annex 2A shows an overview of the SAR missions considered.

As presented in Annex 2B, based on the available data, the flood in **Bangladesh in 2015** and **Thailand in 2011** have been selected as the two most representative case studies.

Regarding **Bangladesh**, several events were explored. We found that **the flood of June/August 2015 is the event best covered with S1 data**. This case study is ideal since the high frequency of acquisition of S1 (one image every six days) can allow a very good statistical analysis of reference images. Table 2B.1. in Annex 2B shows the list of images available in the S1 archive. **Error! Reference source not found.** shows the relative orbit’s footprint (ascending and descending) covering the area of interest.

Figure 2B.2. shows in detail the portion of each footprint considered in our analysis. Figure 2B.3. shows quick looks of the data used.

Regarding **Thailand**, the big flood of late October 2010, which would have been of great interest, is covered mainly by ALOS-PaISAR data (JAXA), which are not freely accessible. Similarly, the flood in Chanthaburi Province of July 2013 is covered by TSX data, also not freely accessible.

The flood of late October/November 2011 in Bangkok is covered mainly by TSX (see Table in Annex 2B). The data are not free. The images are ScanSAR with a resolution of 18.5 m. These images are similar to the new S1 data and could have been a good test bench. Nevertheless, the non-open access policy of these data prevented their use in this scope; the same is true for the Canadian Radarsat-2 data, which cover the event but are not openly accessible.

ENVISAT-ASAR partially covers the event, as indicated by Table in Annex 2B. These data, freely and openly accessible, are the precursor of S1, which are the data we intend to test and use mostly for future events (since they are openly accessible). Figure 2B.4. shows a quick look of these data. Therefore, this event presents itself as a suitable case study.

Even though we knew that **Malaysia** suffered from flood events in recent years, no free SAR data have been found covering major flood events. Optical data have also been explored (Landsat and Sentinel-2) but no useful data have been found. While Southeast Asia is well covered by S1 data, the mission is not acquiring systematically over Malaysia. The acquisition plans show that Malaysia is excluded with a frequency of two weeks, i.e., acquisitions for two weeks, no acquisitions for the next two weeks, and so on.¹⁴ This decreases the probability to have images that cover flood events and it explains why there are no data for the most recent floods.

Table 4 in Annex 2B shows an overview of flood events for which EO data have been searched, for Bangladesh, Thailand, and Malaysia.

2.4 Flood Mapping by Means of Synthetic Aperture Radar and Optical Data

The aim of this task is to use remote sensing data to map flood events in terms of extent, occurrence, and depth in Bangladesh and Thailand. The mapping of floods is designed to be executed rapidly and easily after a disaster, with sufficient relevant historic hazard data to produce a return period for an impact assessment methodology.

The remote sensing team mapped the flood events based on:

1. SAR using S1 and ENVISAT-ASAR using the method developed by Cian et al. (2017a)
2. LANDSAT satellite images using the method developed by Pietro in collaboration with Jean-Francois Pekel (J. F. Pekel et al., 2011, 2014).

The algorithms have been transferred into Google Earth Engine. Monthly flood maps have been created for Bangladesh (May to October 2015) and Thailand (October 2011). The products have been converted to shape files for further analysis in GIS packages, such as for computing the percentage of flooded area at different administrative levels. In Annex 2C we can see the results for Bangladesh obtained with S1 data

¹⁴ <https://sentinel.esa.int/web/sentinel/missions/sentinel-1/observation-scenario>

(Figure 2C.1.1) and with Landsat 8 data (Figure 2C.2.), and the results for Thailand obtained with ASAR data (

Figure 2C.3).

Figure 2C.4 shows the percentage of flooded areas for the Thailand flood, at administrative level 3 (left) and level 2 (right).

2.5 Estimation of Flood Occurrences

Given the high number of observations provided by S1, we counted the number of times that water was detected for each pixel, on a monthly basis and for the whole period of observation (May to September 2015) in Bangladesh. Occurrences are affected by the number of observations (i.e., repeating cycle and acquisition plan of the mission). Nevertheless, they can be a useful indicator of flood severity. In fact, occurrences can relate to flood duration. Given the overlap of the different orbits, estimating the duration is not straightforward. Moreover, the fact that a pixel maintained the status “water” in two consecutive observations does not actually mean that the flood was present for all the time in between. This measure of severity, useful for a correct economic assessment of impacts, has been left for further development in Phase 2 of this project. In

Figure 2D.1 in Annex 2D, we can see occurrences for Bangladesh 2015.

Estimation of Flood Depth and Statistic Report at Administrative Level

Flood depth has been computed following the methodology proposed by Cian et al. (2017b). The elevation of the water surface is estimated through a statistical analysis of the elevation values along the contour of each flooded area, which in theory should be all the same. In fact, water should rest in a flat surface and the elevation at the transition line “water-no water” should be constant. In practice, due to different sources of error (misalignment between SAR and DEM images, different resolution of SAR and DEM, errors in DEM, errors in the flood map, etc.), different elevation values are found and a statistical analysis is needed to estimate the correct elevation of the water surface. With an adaptive threshold, we looked for the most appropriate water elevation starting from the maximum elevation value found along the contour.

The resulting flood depth for each pixel is simply the difference between the estimated water elevation and the terrain elevation. The result is strongly influenced by the resolution and accuracy of the DEM; the higher the resolution, the more accurate the result. For both case studies, we disposed of the SRTM DEM at 30-meter horizontal resolution and 1 m vertical. The coarse resolution of this data set does not allow achieving high precision in flood depth. **Error! Reference source not found.** in Annex 2E shows the result for Bangladesh, giving the flood depth estimation for the month of July 2015. In the lower side of the image, we can see a zoom showing the “noise” affecting the result. These fluctuations of values are due to the resolution of the DEM. To get rid of this problem, we made an attempt using a statistical analysis, i.e., the median flood depth value computed for each single flooded area. Figure 2E.2. shows the result of this technique.

This approach may introduce extra noise locally, but on larger areas the median depth values seem to be a reasonable compromise between DEM resolution and depth values for economic impact assessment. In Phase 2 of this project, we foresee a development for this methodology for improving the accuracy of

the results and a better quantification of the uncertainty connected to this information. In fact, a small variation in flood depth may result in a big variance in economic impacts. Finally,

Figure 2E.3. in Annex 2E shows the results for Thailand, October 22, 2011.

Having a higher resolution DEM would consistently improve the precision of this product. One data set available at the global scale is the **WorldDEM** of Airbus Defence, with **12-meter horizontal resolution** and **2-meter vertical accuracy**. This data set is not free, and very expensive considering the extent of the flood events considered in this scope. Table 2F.1. in Annex 2F shows an overview of DEM products with respective prices. For a subset of our case study in Bangladesh (

Figure 2F.1. in Annex 2F), the purchase of such a product would be about €200,000 .

Finally, median flood depth has been computed, for both Bangladesh and Thailand, at administrative level 2 and 3 as a proxy of economic damage for rice crop production.

Statistical Comparison of Flood Maps Derived from SAR and Optical Data

The flood maps derived from Landsat 8 and S1 data have been compared to one another using the statistical analysis proposed by Lessel & Ceccato (2016), taking Landsat 8 as a reference. The statistics were computed using (1) Agreement accuracy, (2) Probability of Detection (POD), (3) False Alarm Ratio (FAR), (4) Critical Success Index (CSI), and (5) Kappa Index, computed as follows:

Flood Map to Assess	Reference Flood Map	
	No Flood	Flood
No Flood	correct negative	Miss
Flood	false alarms	hit

Table 2.1. Statistics Used to Cross-Compare Different Flood Maps

Name	Formula	Perfect Score (100% Agreement)
Accuracy	$(\text{hits} + \text{correct negatives}) / \text{total}$	1
Probability of Detection (POD)	$\text{hits} / (\text{hits} + \text{misses})$	1
False Alarm Ratio (FAR)	$\text{false alarms} / (\text{hits} + \text{false alarms})$	0
Critical Success Index (CSI)	$\text{hits} / (\text{hits} + \text{misses} + \text{false alarms})$	1

As previously mentioned, the comparison of different products is affected by several factors:

- Landsat is affected by cloud cover; therefore the difference with S1 could increase in case of significant cloud cover.
- Acquisitions, even though averaged on the month, could refer to different stages of the flood; therefore one satellite could observe a much lesser extent than the other if the acquisitions are far apart from each other.
- Revisiting time is also affecting the detection; Landsat 8 does not have as frequent a revisit time (16 days) as S1 (6 days) and MODIS (daily).
- Products have different spatial resolution. In order to compare Landsat with MODIS, products need to be degraded to the lower resolution. Therefore, the precision of S1 (20-meter resolution) had to be degraded to 30 meter (Landsat 8) and 250 meter (MODIS).

Figure in Annex 2G shows the results of the comparison between Landsat 8 and S1. In order to perform the cross-comparison, a resampling step was performed. In particular, S1 flood maps have been resampled at 30-meter spatial resolution in order to match Landsat 8 flood maps.

The results show a very good match between the products, when cloud cover is not affecting the optical observation and when both satellites acquired data at close to the same time. The month of June is an example of a good match ($K = 0.58$, $CSI = 0.47$), as well as the month of September ($K = 0.66$, $CSI = 0.56$). The months of July ($K = 0.24$, $CSI = 0.17$) and August ($K = 0.06$, $CSI = 0.04$) are examples of a bad match between the two products. The main reason is the cloud cover affecting the Landsat 8 images. Figure ¹⁵ reports the Landsat 8 acquisition from January to September 2015 over the area of interest in Bangladesh. For each image, the percentage of cloud cover is reported. We can see how the observations of July and August over the area of interest are affected almost completely by cloud cover, preventing the observation of the flood. Therefore, the flood was observable only with SAR data, which are not affected by cloud cover. This explains the low agreement between the two products.

¹⁵ The image was generated from this webpage: <https://remotepixel.ca/projects/satellitesearch.html>.

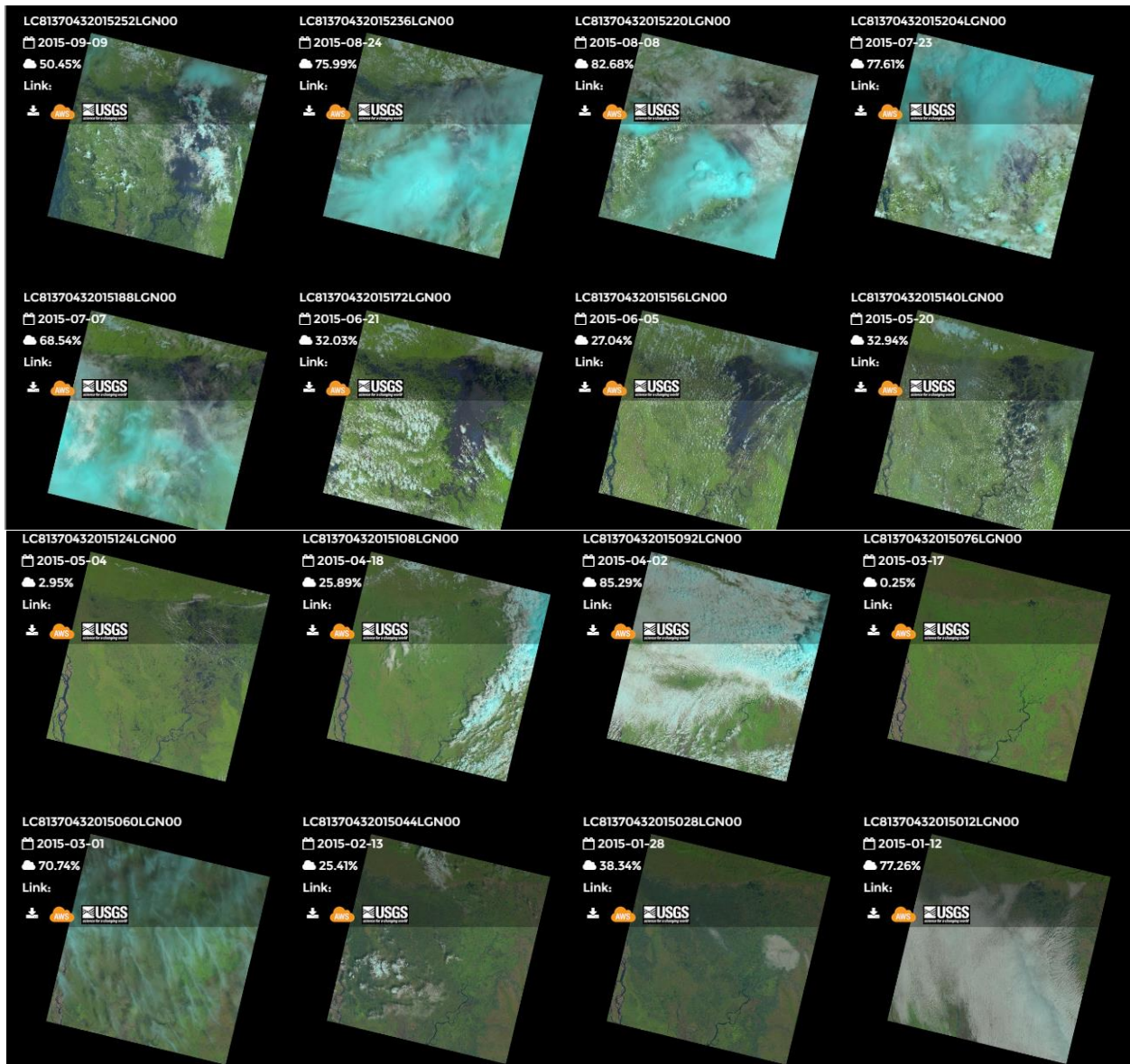


Figure 2.1. Cloud cover for Landsat 8 acquisition from January 1, 2015, to September 9, 2015

Additionally, we recomputed the statistics avoiding the sources of difference between the two products. The flood maps presented previously were obtained from images on different orbits, either for S1 or L8 data. In certain cases, due to missing data for one of the orbits involved or cloud coverage, the products derived from S1 and L8 were showing differences, which then had influenced the statistic relative to their comparison. In order to show a more consistent statistic, an area where both S1 and L8 had good data coverage was selected and analyzed, as shown in Figure (top) in Annex 2H. The example refers to the maps of June 2015. The statistics were considerably improved. Despite a decrease of the overall accuracy (due to a lower number of agreements for non-flooded pixels, due to the smaller size of the image), Kappa increased by 6% (from 0.58 to 0.64), POD by 10% (from 0.6 to 0.7), and CSI by 14% (from 0.47 to 0.61). FAR decreased by 15% (from 0.33 to 0.18). This shows how the two products have a higher agreement than previously presented. However, some

limitations of the two products remain (cloud cover for Landsat 8; shadowing, layover, etc. for S1) and are the main sources of discrepancies.

The methodology used for retrieving flooded areas from Landsat 8 did not exclude permanent water bodies. These “false” alarms were affecting the statistics when comparing Landsat 8 flood maps with the flood maps retrieved from S1. Therefore, permanent water bodies were removed and statistics re-computed for the flood of June 2015.

Figure (middle) in Annex 2H shows how POD has further increased by 3% (now 0.73) and CSI by 2% (now 0.63).

Finally, knowing that cloud cover is still affecting the statistics, the S1 flood map was cropped using the cloud mask computed from the L8 observations of the month of June (Figure , bottom). In Annex 2H, the results show how Kappa has further increased by 8% (now 0.72), POD by 5% (now 0.78), and CSI by 6% (now 0.69).

The two products therefore show a high agreement when we get rid of some of the sources of difference. A perfect simultaneous acquisition of optical and SAR data over a flooded area would result in an almost identical flood extent. Nevertheless, there will be always some intrinsic differences due to the different technologies and algorithms employed. For example, SAR is affected by shadow and layover effect (geometry of acquisition), which is a difference with optical observation that can cause discrepancies. Moreover, in urban areas both SAR and optical data will have difficulties in detecting water surface.

We can affirm that Sentinel-1 is more reliable than Landsat when there is a lot of cloud cover. As shown, when excluding the clouds, the two products are very similar. Therefore, users should first trust Sentinel-1-based products when cloud coverage is high. Sentinel-1 can be complemented with Landsat-based products for improved information about the extent and duration of the event. In case of availability of field measurements, these could be used for validation and fine-tuning the satellite-based products.

Comparison of Flood Maps with Other Products, Both Satellite and Model-based

Flood maps derived from S1 and L8 data were further compared with:

- NRT Global Flood Mapping from MODIS images at 250-meter resolution¹⁶
 - NRT MODIS versus L8
 - NRT MODIS versus S1
 - NRT MODIS versus L8 and S1 combined
- The Global Surface Water layer from the Joint Research Centre derived from Landsat data¹⁷
- Global Flood Monitoring System (GFMS)¹⁸ at 1 km resolution

In order to perform the cross-comparison with the NRT MODIS maps, a resampling step had to be performed. In particular, S1 and L8 flood maps have been resampled at 250 meter in order to match NRT flood maps' resolution.

Annex 2I shows the results of this cross-comparison.

Cross-comparison of Landsat 8-based flood maps and NRT MODIS is shown in

Figure 2I.1.. In June 2015 we can see a good comparison between Landsat 8 and the MODIS product (CSI = 0.36), with some mismatch due to cloud cover on Landsat and probably due to the coarse resolution of MODIS on areas detected only by Landsat 8. July and August do not show a good agreement (CSI = 0.19 and 0.04 respectively) between the two products, mainly due to the cloud cover on Landsat images.

The comparison with Sentinel-1 (Figure 2I.2.) shows a much better agreement but still with differences due to different timing of acquisition and coarser resolution of MODIS. CSI is equal to 0.38, 0.53, and 0.52, respectively for June, July, and August.

Combining Landsat 8 and Sentinel-1 together (

Figure), we obtain a lower agreement with MODIS product in June (CSI = 0.36), mainly due to the fact that Landsat 8 and Sentinel-1 together detect more floods than MODIS. In July we note a higher agreement, with CSI equal to 0.55, and in August the same agreement, CSI = 0.52.

Concerning the Global Surface Water layer from the Joint Research Centre, only a visual inspection was conducted. The comparison shows a very good correspondence with the L8-based flood maps. This data set is a valuable source of flood data but it presents the same problems as Landsat: the frequency of observation and the cloud coverage prevent the detection of certain flood events. In fact, as an example, this data set does not detect the November 2010 flood which occurred in northeastern Italy around the city of Vicenza.

¹⁶ <http://oas.gsfc.nasa.gov/floodmap/getTile.php?location=090E030N&day=167&year=2015&product=14>

¹⁷ <https://global-surface-water.appspot.com/download>

¹⁸ <http://eagle1.umd.edu/flood/download/>

GFMS products were compared visually, and an overview of the results is shown in Annex 2J. GFMS flood estimation products are based on a hydrological runoff and routing model, which takes as an input real-time TRMM Multi-Satellite Precipitation Analysis (TMPA) precipitation information. Given that this is a model-based product, we don't expect a perfect timing correspondence between the satellite observation and the model outcome. Therefore, we looked for the maximum correspondence of GFMS products in a time window of two days centered on the date of the satellite image acquisition.

Landsat 8 and Sentinel-1 flood extent maps have been compared to GFMS products for 2015 (Figure 2J.1.1) and 2016 (

Figure 2J.2.2). Data have been selected in order to see a correspondence between the flood extent in the GFMS data and the flooded area appearing in the LANDSAT 8 and Sentinel-1 data. That means that there is no perfect correspondence between the dates of the different products.

We noticed that in many instances, the GFMS was showing the presence of flooded areas, while there was nothing reported in the LANDSAT 8 images, as we can see in Figure 2.2.

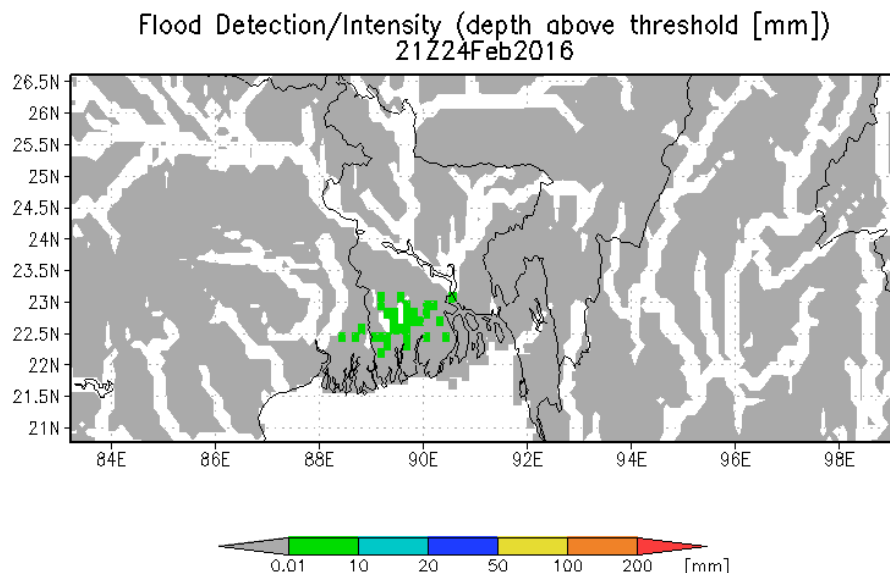


Figure 2.2. GFMS flood estimation over Bangladesh for February 24, 2016, 21 UTC, where flooded areas are reported, while LANDSAT 8 did not show any water bodies

This could be an indication of false alarms from the GFMS, but we cannot be sure without accessing field measurements to verify the GFMS products.

For the periods where Landsat 8 showed flood events, we noticed that the area estimated from Landsat 8 was smaller than the areas presented by the GFMS, as we can see in Annex 2J. We can see a good correspondence in June 2015, but almost total disagreement in July and August 2015 (

Figure 2J.1.), mainly due to cloud cover in the Landsat data. Almost total disagreement can also be noticed in 2016 (

Figure 2J.2.), with Landsat affected by cloud cover in June, July, and partially in August.

Concerning Sentinel-1, we can see a better agreement, though still not satisfactory. In 2015 (

Figure 2J.1.), Sentinel-1 is showing flood in the northeast side of the country as GFMS products do, but the latter seem to indicate as flooded a much bigger area. In fact, in July and August GFMS indicates that almost the entire country is flooded. For 2016 (

Figure 2J.2.), the results are similar, i.e., GFMS seems to overestimate. In this case we have a good correspondence in August, while in June and July the flooded areas do not match.

The comparison between GFMS and Landsat 8/Sentinel-1 products is also not easy due to the different resolution, much coarser in the case of GFMS.

Finally, we also attempted to assess the differences for the products delivered by the Dartmouth Flood Observatory,¹⁹ but for the Bangladesh 2015 event no products were available.

2.6 Population Exposure Assessment

Population exposure to floods was assessed using different population data sets. We considered the following data sets:

- Global Human Settlement Layer–Population produced by the Joint Research Centre, European Commission:²⁰ The data report population density (population per pixel) and are available at 250 m and 1 km spatial resolutions in a World Mollweide projection, which creates some difficulties for comparison and analysis.
- WorldPop:²¹ This is a project of the Flowminder foundation,²² which provides open access products derived from mobile operator data, satellite data, and household survey data. Besides population density data, it produces data about population displacement, estimate of poverty, social indicators, and prediction of infectious diseases spreading. The project started in 2013 and provides an open access archive of spatial demographic data sets. We used the Prediction weighting layer used in population redistribution at ~100m pixel resolution, as estimated using the random forest (RF) model described in Stevens et al. (2015). The data report population density (population per pixel). More information on the methodology used is available in the metadata file provided by WorldPop. This data set excludes the area regularly flooded in Bangladesh, as can be seen in Figure 2.3.

¹⁹ <http://floodobservatory.colorado.edu/Archives/index.html>

²⁰ <http://ghsl.jrc.ec.europa.eu/datasets.php>

²¹ <http://www.worldpop.org.uk/>

²² <http://www.flowminder.org>

- Global Urban Footprint (built-up area footprint only) produced by the DLR:²³ Released in 2016, it is a global map of human settlements at 0.4 arcsec resolution (~12 m). The high resolution and accuracy of the data set make it useful for exposure analysis. The data report urban area as a binary map. The data can be accessed after ordering them at DLR.²⁴ The data set represents urban footprint and not population based on TerraSAR-X and TanDEM-X scenes.

Annex 2K shows the results of the comparison. Figure (top) shows the JRC GHSL-POP layer (people per pixel) superimposed to the flood detected in June 2015. In this data set, population was distributed over urban areas detected from remote sensing data, i.e., only the pixels reporting human settlement bear a certain population density. Figure 2K.1 (center) shows the WorldPop layer (people per pixel). In this case, the methodology used to disaggregate population data is more complex (see Stevens et al., 2015) and the population density looks very different from the previous data set. Moreover, it is evident that over the regularly flooded area analyzed, there is a population density of zero, as can be seen in Figure 2.3, where only the WorldPop layer is shown.

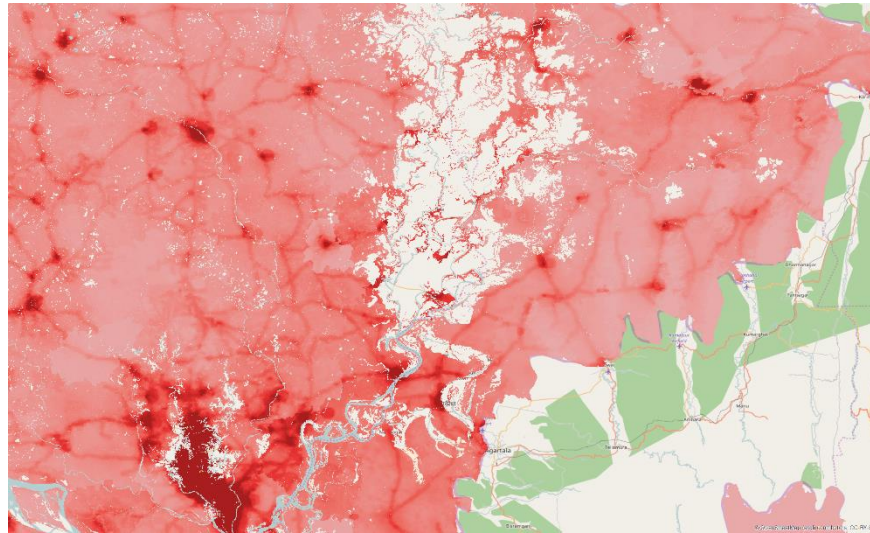


Figure 2.3. WorldPop layer. It can be noted that over the area of the flood there are no data about the population density.

Figure 2K.1 (bottom) shows the Global Urban Footprint layer, which is simply a binary map showing urban areas without providing information about the population.

Figure 2K.2 shows a zoom in two different areas affected by the flood of June 2015. The differences between the three data sets can be appreciated, as well as the higher resolution of the GUF layer with respect to the JRC GHSL.

A quantitative comparison of the JRC GHSL and WorldPop data sets was performed after converting them to the same projection (Mollweide has been chosen to be consistent with the JRC–GHSL layer) and

²³ http://www.dlr.de/eoc/en/desktopdefault.aspx/tabid-9628/16557_read-40454/

²⁴ Email to Manfred Keil at: Manfred.Keil@dlr.de.

resampling them to the coarser resolution (250 m, like the JRC–GHSL). The resampling step has worsened the quality of the GUF and WorldPop data sets, impeding a perfect comparison of the three. Nevertheless, the results provide insight into the differences between the data sets, as shown in Figure 2K.3, where JRC GHSL has been compared to WorldPop. From the difference of the two layers, it appears that the JRC GHSL layer reports a higher density over bigger built-up areas, while WorldPop also seems to distribute population outside of built-up areas. As shown in Figure 2K.3, the difference in the two data sets can be really high (up to thousands of people) and therefore provides considerably different exposures.

Superimposing population data to the flood maps, we estimated the affected people (in the case of the GHSL- and WorldPop and the area affected the case GUF Error!

	GUF [Km ²]	WorldPop [million People]	JRC – GHSL-POP [million People]
S1 only	6.93	0.246	1.470
L8 Only	37.81	0.272	2.478
S1 and L8	0.96	0.132	0.311
S1 Tot	7.89	0.378	1.781
L8 Tot	38.77	0.403	2.789
S1+L8	45.7	0.649	4.259

POP
layer)
urban
(in
of the
layer).

Reference source not found.2K.1 and

Table 2K.2 in Annex 2K show the results for the months of June 2015 and July 2015 respectively.

“S1 ONLY” and “L8 ONLY” refer to the flood extent detected only by S1 or L8. “S1 AND L8” refers to the flood detected by both sensors. “S1 TOT” and “L8 TOT” refer to the total flood extent detected respectively by S1 and L8. “S1+L8” refers to the combined flood extent of S1 and L8, i.e., the sum of “S1 ONLY” and “S1 AND L8,” and “L8 ONLY” and “S1 AND L8.”

Depending on the area affected (if densely populated or not), the discrepancies between the two population data sets can be quite high. The JRC GHSL-POP layer provides almost in all cases a much higher number of people affected. For the flood of July, the difference in exposure is very high, higher than 3.5 million: 0.65 million based on the WorldPop data, 4.26 million in the case of the JRC GHSL-POP.

For the month of July, the cloud cover on Landsat images did not allow detection of most of the flooded area. Despite the big difference in flood extent, the difference in the estimation of affected people is of a smaller magnitude. The reason is the low population density within the area affected by the flood.

Figure shows two details of the area analyzed. It is clear how the three data sets provide different information, which leads to different exposure assessment. On area A, the JRC GHSL-POP layer reports high population density overlapping a flooded area. In the same area, OpenStreetMap does not report a settlement. The WorldPop layer provides a much smaller number of people (lower population density) affected, and the DLR GUF presents a further different distribution of settlements, which results in a different surface of built-up areas affected. Similarly, on area B, the population density reported by the

JRC GHSL-POP layer is much higher than that reported by WorldPop, while DLR GUF reports a smaller built-up area. Also in this case, the exposure differs depending on the data set employed.

New products will be made available in the near future, which may become useful for this study. Facebook Connectivity Lab is releasing a new population density layer, High Resolution Settlement Layer (HRSL).²⁵ At the time of writing this report, the data set was not available for a few locations, with Bangladesh and Thailand probably available around May 2017. The data set is promising, and in Phase 2 of the project it will be considered and compared with the other data set available.

In addition, the Land Surface department at DLR is working on a new release of the GUF. The new data set is supposed to have an increased accuracy and it refers to year 2015. The new data set is going to be freely accessible, unlike the precedent version, which is accessible after purchase or special agreement. In the near future, DLR should also release the data set for past years until 1990, allowing further studies like change detection analysis. DLR uses the Thematic Exploitation Platforms (TEP) of the European Space Agency to process the data, and the GUF data set is already available inside these new platforms. It is likely that ESA will put some efforts into developing these data sets, for example for providing information about population density, which is valuable information for disaster impact assessment, similar to what the project WorldPop is currently providing.

2.7 Estimation of Surface Water Coverage from Optical Data from 1984 to 2015

As support to the statistical analysis of precipitation data, we extracted the area of seasonal surface water detected by Landsat 8 every year in the whole country. The data set containing this information is the “JRC Yearly Water Classification History v1.0” available in Google Earth Engine and produced by the European Commission Joint Research Centre. The data set contains maps of the location and temporal distribution of surface water from 1984 to 2015, from which we extracted the information for Bangladesh and Thailand as presented respectively in Table 2L.1.2L.1 and Table 2L.1. Estimation of Yearly Surface Water Area in Bangladesh Detected by Landsat from 1984 to 2015 in Annex 2L. The data set refers to seasonal water, which we used for this statistical analysis. Seasonal water is water surfaces appearing in the image as non-permanent water surfaces, which therefore can be considered as flooded areas.

The limitations of optical data have to be kept in mind. Landsat data can be affected by cloud coverage, which prevents the detection of surface water. Moreover, especially for older data, revisit time of the satellite and capability of continuous data acquisition can substantially limit data coverage over the areas of interest and therefore limit the capacity of detecting surface water. The statistic provided by this analysis is biased by the limitations of the particular satellite technology employed and the algorithm used to detect water surfaces.

²⁵ <http://www.ciesin.columbia.edu/data/hrsl/>

2.8 Analysis of Vegetation Indices and Estimation of Anomalies in the Period of the Flood

In order to assess the impact of floods on agriculture, we analyzed time series of monthly normalized difference vegetation index (NDVI) over the districts of Netrokona, Mymensingh, Kishoregani, Gazipur, and Tangail in Bangladesh.

The NDVI provides an idea of the vegetation status. In this analysis, we computed monthly NDVI anomalies in order to detect a decrease of vegetation activity compared to the mean value. The term “anomaly” means the difference between the value of a quantity and its climatological mean value. The “monthly anomaly” we computed is the difference between the original monthly value of NDVI in a given month and the monthly climatological value for that month of the year (Equation 1).

$$r'_{ij} = r_{ij} - \frac{1}{N_i} \sum_{j=1}^N r_{ij} \quad (1)$$

Here, for months i and years j , r'_{ij} is the monthly anomaly, r_{ij} is the original monthly NDVI value, and the remainder of the equation is the calculation of the monthly climatology (which is subtracted from the original monthly values). We used the time period May 2000 to January 2017 to compute the monthly climatology.

Monthly anomalies of NDVI indicate the difference (positive or negative) between a monthly NDVI value and its “normal” value for that month of the year. Calculating anomalies is one way to remove the annual cycle from a time series. The graph below shows the monthly NDVI anomalies. We can observe that negative anomalies in September 2004 and September 2007 correspond to flood events in all the six districts. However, the negative anomalies in July 2002 correspond to a drought. Analysis of NDVI values as impact of floods will have to be analyzed carefully since two opposite disasters (flood/drought) might have the same impact on the vegetation and crops.

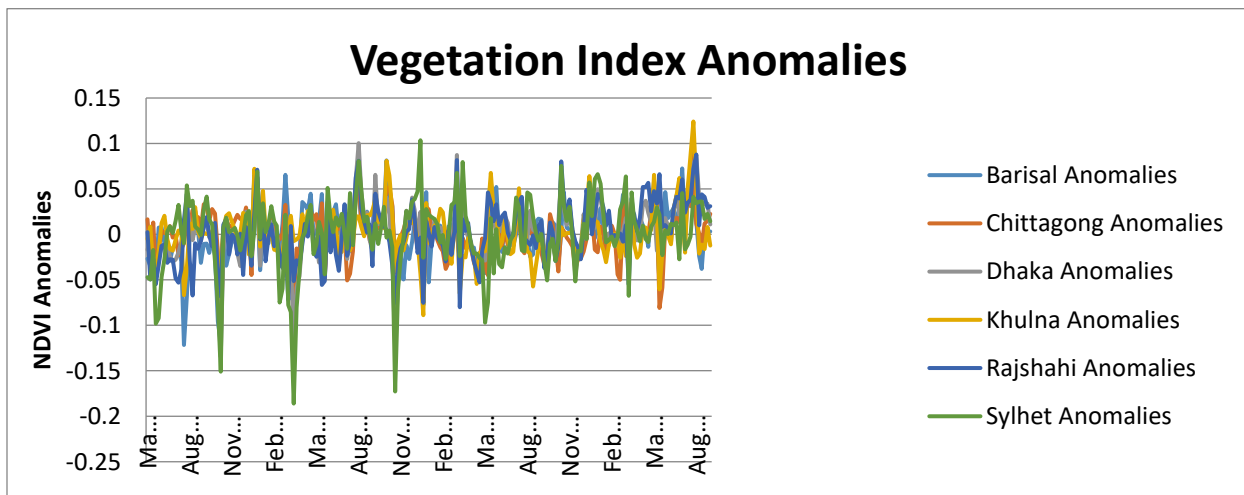


Figure 2.4. Vegetation index anomalies

2.9 Framework for Assessing Uncertainties

We propose here below a structure to assess the uncertainties linked to (1) flood detection based on different satellites, and (2) population data set used.

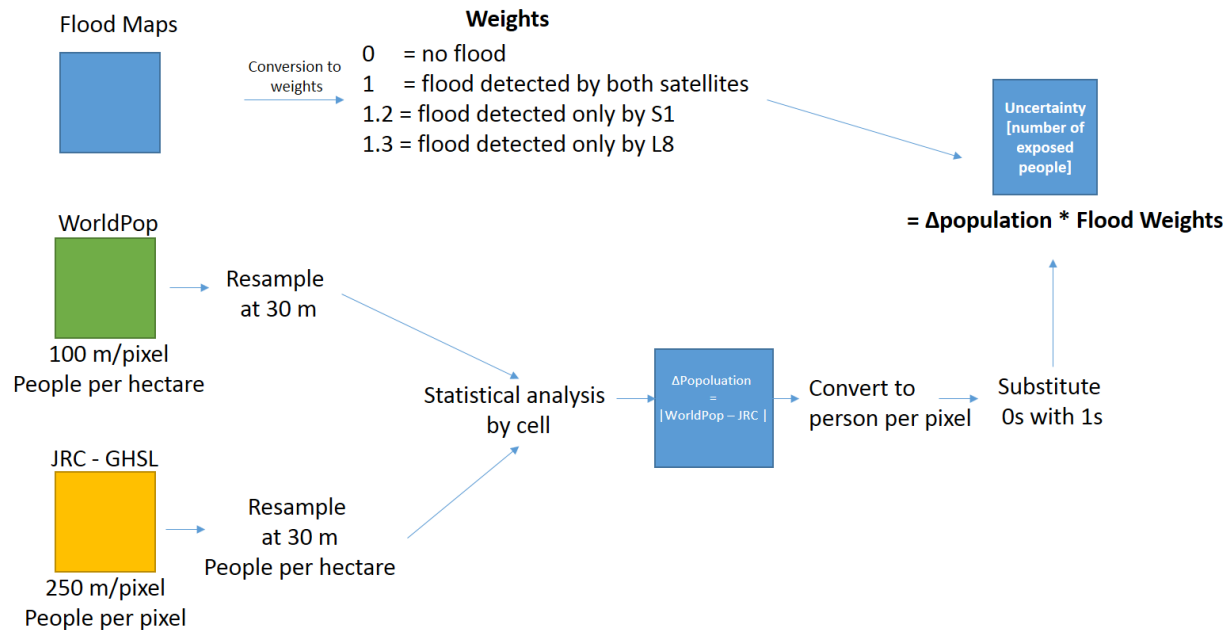


Figure 2.5. Uncertainty estimation framework

We propose to weight the uncertainty linked to the flood maps by providing a score to each pixel detected either as flood or no flood from the different sensors. As an example we propose weights as follows:

0 = no flood

[W = 0 → uncertainty is 0 (when no flood is detected, we can't infer anything about the flood)]

1 = flood detected by both satellites

[W = 1 → the weight in maintaining constant the uncertainty given by the population data sets]

1.2 = flood detected only by S1

[W = 1.2 → the uncertainty is amplified by 20% since it was detected only by S1. S1 receives a smaller weight in order to highlight its independence by cloud cover as in the case of L8]

1.3 = flood detected only by L8

[W = 1.3 → the uncertainty is amplified by 30% since it was detected only by L8]

The flood weights are then multiplied by the difference between the population data sets, in this case the difference between the population data set from WorldPop and the population data set from the JRC-GHSL (noted Δ population).

When both population data sets are equal (Δ population = 0) the uncertainty is simply given by the flood weight.

The results of the multiplication between the **flood weight** and **Δ population** provide a value of uncertainty expressed as a number of exposed people uncertainty-weighted by the flood uncertainty. The uncertainty numbers can be displayed as a map where a color legend can be defined by means of thresholds (low, medium, high).

Wherever the uncertainty goes beyond high value, it will be displayed in red allowing the users to easily identify the areas where there are high uncertainties based both on flood uncertainty and population uncertainty, and where ground truth may be required.

As an example, the map below represents an uncertainty color scale, where blue has low uncertainty (below 10); yellow has medium uncertainty (11 < uncertainty < 20), and red (above 20) has high uncertainty. The map can be created at pixel level or summarized at different geographical units such as the district or provinces.

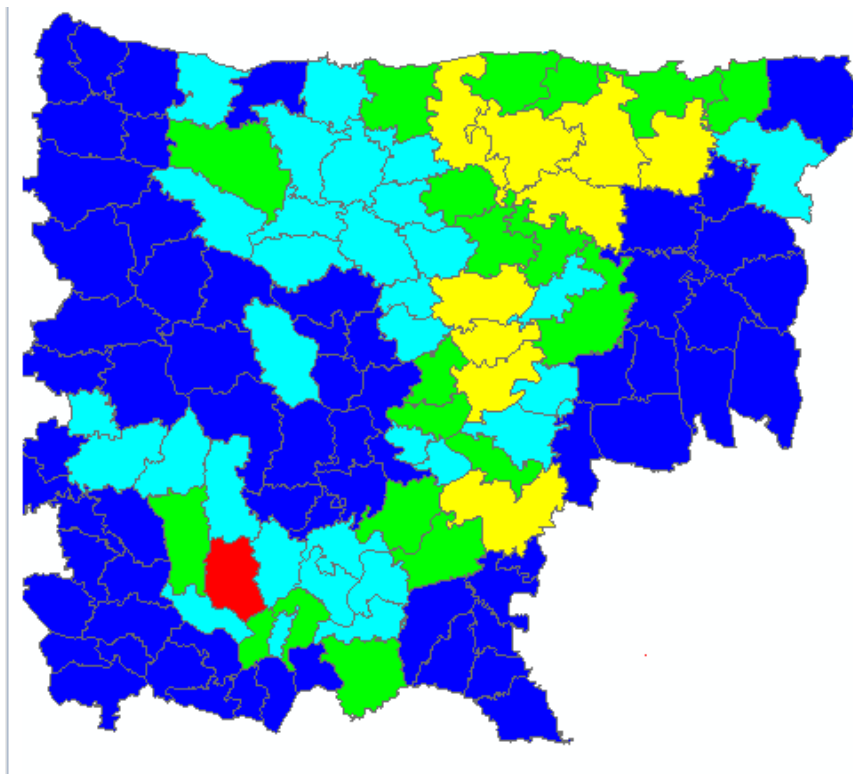


Figure 2.6. Example of uncertainty classification

2.10 Conclusion

We have developed a methodology to map floods' extent, occurrence, and depth using Sentinel-1 and Landsat 8 images and compared them with statistical indices. We have also developed the procedure to assess the impact on population based on three different data sets. We showed that different results can be obtained depending on the data set selected.

The mapping of floods has been developed on the Google Earth Engine platform, which allows the processing of large amounts of data. However, the comparison of different products was done using ArcGIS and required a lot of manual manipulation of the data to compute the statistics.

Our objectives for future work are:

- Develop an automatic procedure to map and analyze floods' extent, occurrence, and depth and compare the results with minimal intervention from the user
- Detect the areas where flood has been reported by more methodologies → indicating a sort of “probability” of flood
- Provide information on uncertainty (which will require some interactions with the users to define the best methodology to use)
- Provide guidance to the user on how to make a decision based on different data sources
- Add additional products based on MODIS data, International Charter, EMS Copernicus, ZKI DLR, and GFMS in order to have more comparisons and provide statistics on their agreement (like model ensemble approach)
- Integrate field measurement, providing points using smartphone or GPS
- Assess different data sources of population and poverty
- Provide the possibility to automatically estimate the number of people affected
- Provide the possibility to assess the accuracy of flood and population data sets combined

2.11 References

- Cian, F., Marconcini, M., & Ceccato, P. (2017a). Normalized Difference Flood Index for rapid flood mapping: taking advantage of EO big data. Submitted to *Remote Sensing of Environment*.
- Cian, F., Marconcini, M., & Giupponi, C. (2017b). Flood depth estimation by means of high-resolution SAR images and LiDAR data. Submitted to *Natural Hazards and Earth Science Systems*.
- Beven, K. (2006). A manifesto for the equifinality thesis. *Journal of Hydrology*, 320(1–2), 18–36. <https://doi.org/10.1016/j.jhydrol.2005.07.007>.
- Delgado, J. M., Merz, B., & Apel, H. (2014). Projecting flood hazard under climate change: An alternative approach to model chains. *Natural Hazards and Earth System Sciences*, 14(6), 1579–1589. <https://doi.org/10.5194/nhess-14-1579-2014>.
- Jain, S., & Lall, U. (2001). Floods in a changing climate: Does the past represent the future? *Water Resources Research*, 37(12), 3193–3205. <https://doi.org/10.1029/2001WR000495>.
- Lessel, J., & Ceccato, P. (2016). Creating a basic customizable framework for crop detection using Landsat imagery. *International Journal of Remote Sensing*, 37(24), 6097–6107. <https://doi.org/10.1080/2150704X.2016.1252471>.
- 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. <https://doi.org/10.1016/j.isprsjprs.2014.07.014>.
- Merz, B., Kreibich, H., Schwarze, R., & Thielen, A. (2010). Review article: “Assessment of economic flood damage.” *Natural Hazards and Earth System Science*, 10(8), 1697–1724. <https://doi.org/10.5194/nhess-10-1697-2010>.
- O’Grady, D., Leblanc, M., & Gillieson, D. (2011). Use of ENVISAT ASAR Global Monitoring Mode to complement optical data in the mapping of rapid broad-scale flooding in Pakistan. *Hydrology and Earth System Sciences*, 15(11), 3475–3494. <https://doi.org/10.5194/hess-15-3475-2011>.
- Pekel, J. F., Ceccato, P., Vancutsem, C., Cressman, K., Vanbogaert, E., & Defourny, P. (2011). Development and application of multi-temporal colorimetric transformation to monitor vegetation in the Desert Locust habitat. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 4(2), 318–326. <https://doi.org/10.1109/JSTARS.2010.2052591>.
- Pekel, J. F., Vancutsem, C., Bastin, L., Clerici, M., Vanbogaert, E., Bartholomé, E., & Defourny, P. (2014). A near real-time water surface detection method based on HSV transformation of MODIS multi-Spectral time series data. *Remote Sensing of Environment*, 140, 704–716. <https://doi.org/10.1016/j.rse.2013.10.008>.
- Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data, 1–22. <https://doi.org/10.1371/journal.pone.0107042>.
- 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*.
- Wilson, B. A., & Rashid, H. (2005). Monitoring the 1997 flood in the Red River valley. *Canadian Geographer*, 49(1), 100–109.

Annex 2A. SAR Missions Overview

Name	Agency	Start/End	Free
ERS-1	ESA	1991 – 2000	Yes
ERS-2	ESA	1995 – 2003	Yes
ENVISAT-ASAR	ESA	2002 – 8 April 2012	Yes
TerraSAR-X	DLR (Germany)	2007 – ongoing	No
TanDEM-X	DLR (Germany)	2010 – ongoing	No
COSMO-SkyMED	ASI (Italy)	2007 – ongoing	No
ALOS-PALSAR	JAXA (Japan)	2006 – ongoing	No
ALOS-PALSAR 2	JAXA (Japan)	2014 – ongoing	No
Radarsat-1	CSA (Canada)	1995 – 2013	No
Radarsat-2	CSA (Canada)	2007 – ongoing	No
Sentinel-1	ESA	September 2014 – ongoing	Yes

Table 2A.1. SAR Missions Overview

Annex 2B. Data availability

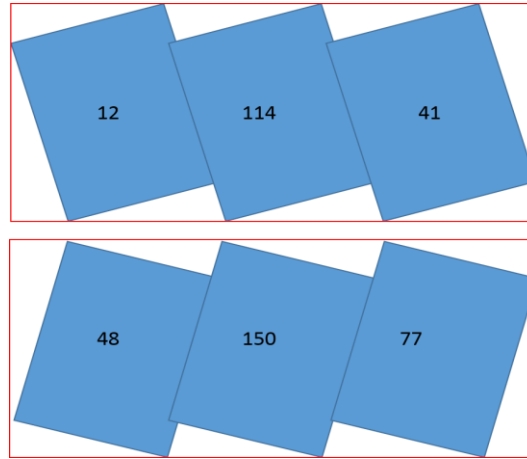
Bangladesh

Ascending 12	Ascending	Ascending 41	Descending 48	Descending 150	Descending 77
23/11/2014	18/11/2014	01/11/2014	14/11/2014	08/11/2014	15/11/2014
17/12/2014	12/12/2014	25/11/2014	08/12/2014	02/12/2014	09/12/2014
10/01/2015	05/01/2015	19/12/2014	01/01/2015	26/12/2014	02/01/2015
03/02/2015	29/01/2015	12/01/2015	18/02/2015	19/01/2015	03/03/2015
11/03/2015	22/02/2015	05/02/2015	02/03/2015	12/02/2015	27/03/2015
04/04/2015	18/03/2015	13/03/2015	26/03/2015	08/03/2015	20/04/2015
28/04/2015	11/04/2015	06/04/2015	19/04/2015	01/04/2015	14/05/2015
22/05/2015	05/05/2015	30/04/2015	01/05/2015	25/04/2015	07/06/2015
03/06/2015	29/05/2015	24/05/2015	13/05/2015	19/05/2015	01/07/2015
15/06/2015	22/06/2015	17/06/2015	25/05/2015	12/06/2015	18/08/2015
27/06/2015	16/07/2015	11/07/2015	06/06/2015	06/07/2015	
09/07/2015	09/08/2015	28/08/2015	18/06/2015	23/08/2015	
21/07/2015	02/09/2015	09/09/2015	30/06/2015	16/09/2015	
02/08/2015	14/09/2015	21/09/2015	12/07/2015		
26/08/2015	26/09/2015		24/07/2015		
			17/08/2015		

Flood	Reference	Presence of water (excluded from reference)	Negligible presence of water
-------	-----------	---------------------------------------------	------------------------------

Table 2B.1. Sentinel-1 Data Availability over Bangladesh in the Period May-September 2015

Descending Orbits



Ascending Orbits

Figure 2B.1. Relative orbit footprints

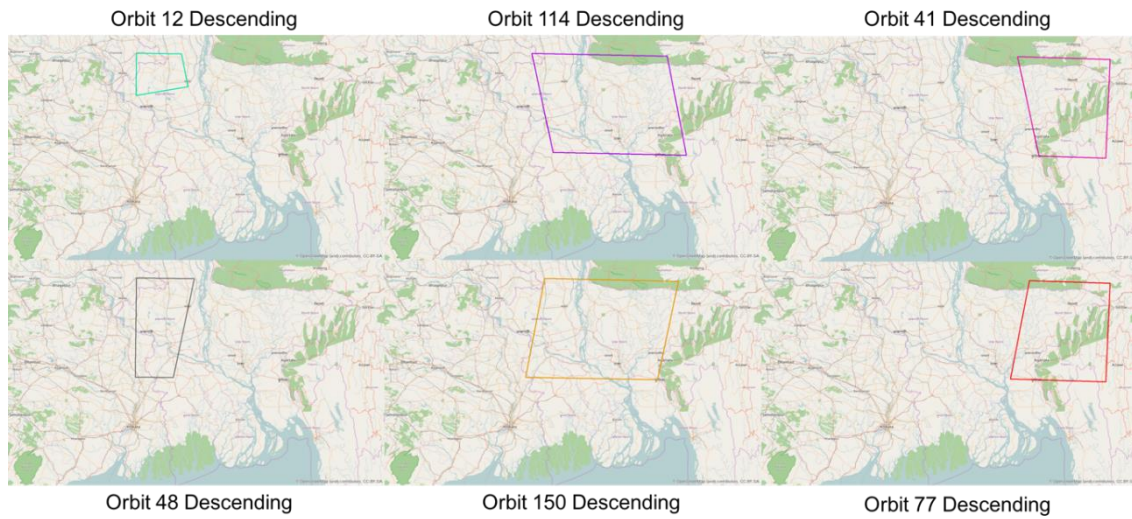


Figure 2B.2. Image footprint for different relative orbits employed in this study. *S1* Images of the six different relative orbits considered have been clipped based on the area of interest.

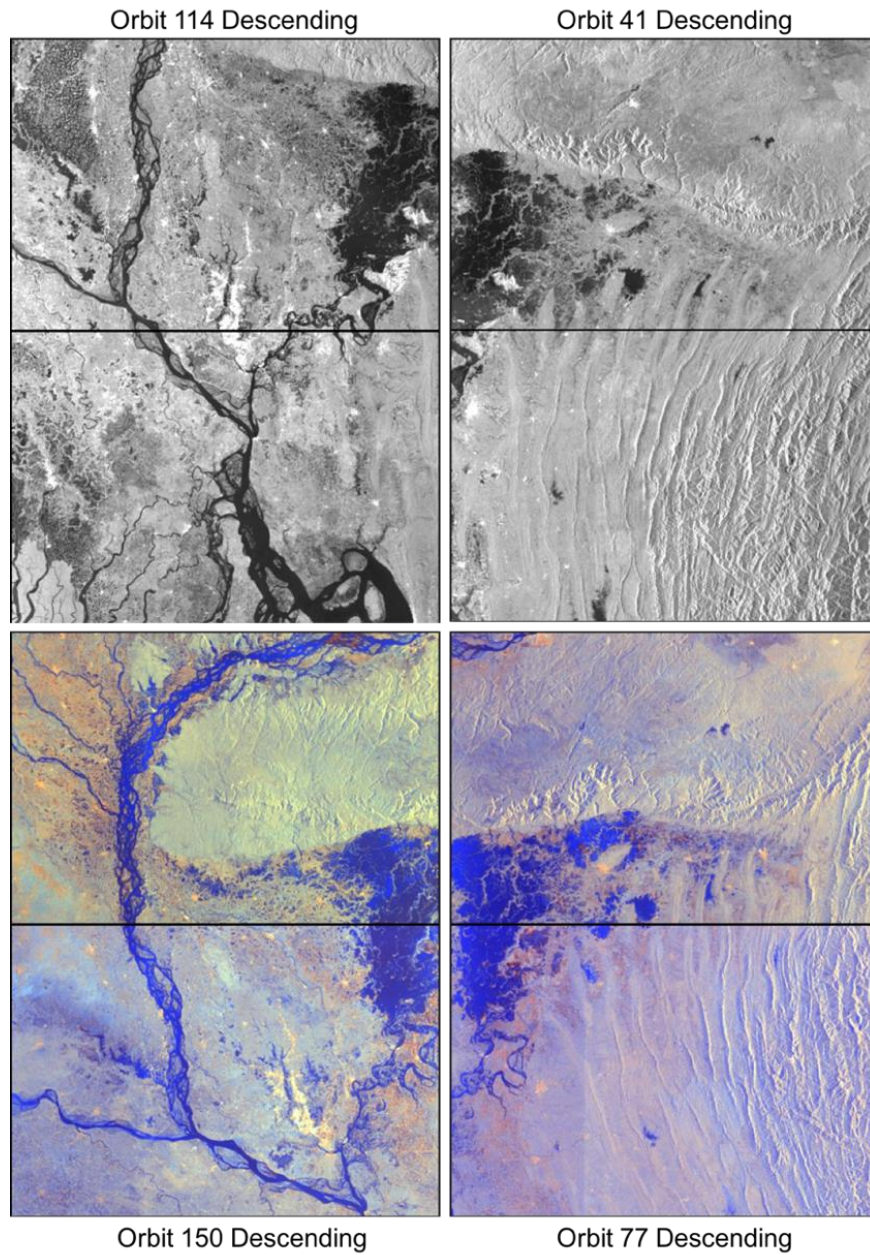


Figure 2B.3. Sentinel-1 data sample over Bangladesh for the relative orbits, descending 114 and 41, and ascending 150 and 77

Thailand

Date	Mode	Orbit
2011-10-26	Scan	Ascending
2011-11-06	Scan	Ascending
2011-11-12	Scan	Ascending
2011-11-17	Scan	Ascending
2011-11-28	Scan	Ascending

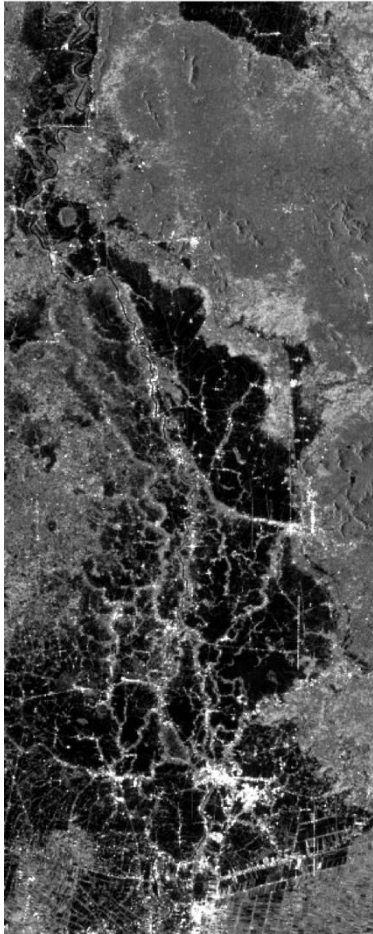
Table 2B.2. TanDEM-X Data Availability for the Flood of October/November 2011

Only One reference image: TanDEM-X, 2011-02-10

Date	Mode	Orbit
2011-10-19	Image Mode	Ascending
2011-10-22	Image Mode	Ascending

Table 2B.3. Envisat-ASAR Data Availability over Thailand for the Flood of October/November 2011

2011 – 10 - 19



2011 – 10 - 22

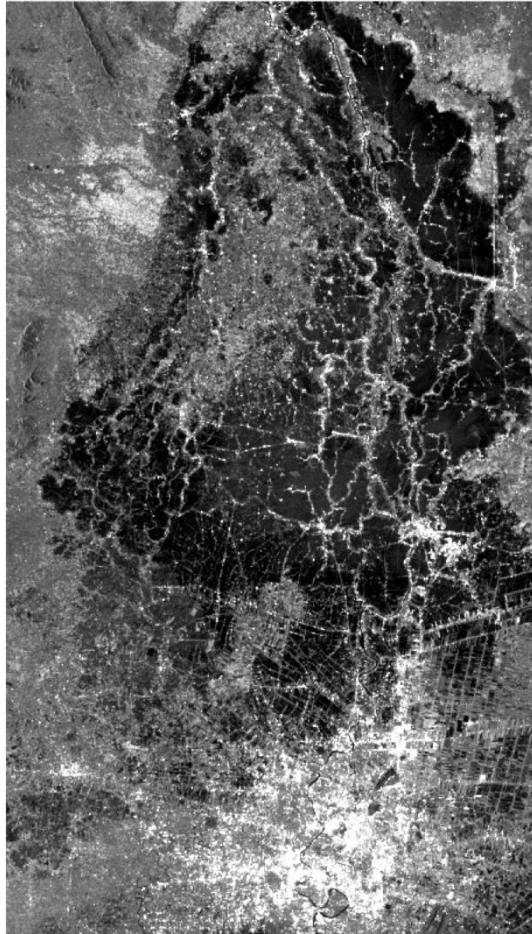


Figure 2B.4. Data samples of the two ASAR images available for the flood of October 2011

List of Flood Events for Which EO Data Have Been Searched

Flood Events in Bangladesh			
Year	Date	Location	Data
1998	July/Sept	Dhaka	No Data
2004	July 8 - August		ASAR
2005		N-W Bangladesh	No Data
2007	July 21 - 15	Dhaka, Khulna, Rajshahi, Sylhet	ASAR
2010	June	Cox's Bazar	ASAR
Flood Events in Thailand			
Year	Date	Location	Data
2010	30-October	Center/North of the country	ALOS & ASAR
2013	25 July	Chanthaburi Province	TSX
Flood Events in Malaysia			
Year	Date	Location	Data
2000	30-Apr	Kuala Lumpur	No Data
2001	26-Apr	Kuala Lumpur	No Data

2001	29-Oct	Kuala Lumpur	No Data
2002	11-Jun	Kuala Lumpur	No Data
2003	10-Jun	Kuala Lumpur	No Data
2007	10-Jun	Kuala Lumpur	No Data
2010	Nov	Kedah and Perlis	No Data
2014	15-Dec – 3 Jan 2015	East Coast and northern regions (Johor, Kedah, Kelantan, Negeri Sembilan, Pahang, Perak, Perlis, Sabah, Sarawak, Selangor and Terengganu)	No Data
2016	4-9-Feb	Sarawak, Johor, Melaka, Negri Sembilan	No Data
2016	19-22 Feb	Sarawak State, Malaysia	No Data
2016	18-19 Jul	Penang and Kedah, Malaysia	No Data

Table 2B.4. List of Flood Events for Which EO Data Have Been Searched

Annex 2C. Flood Maps

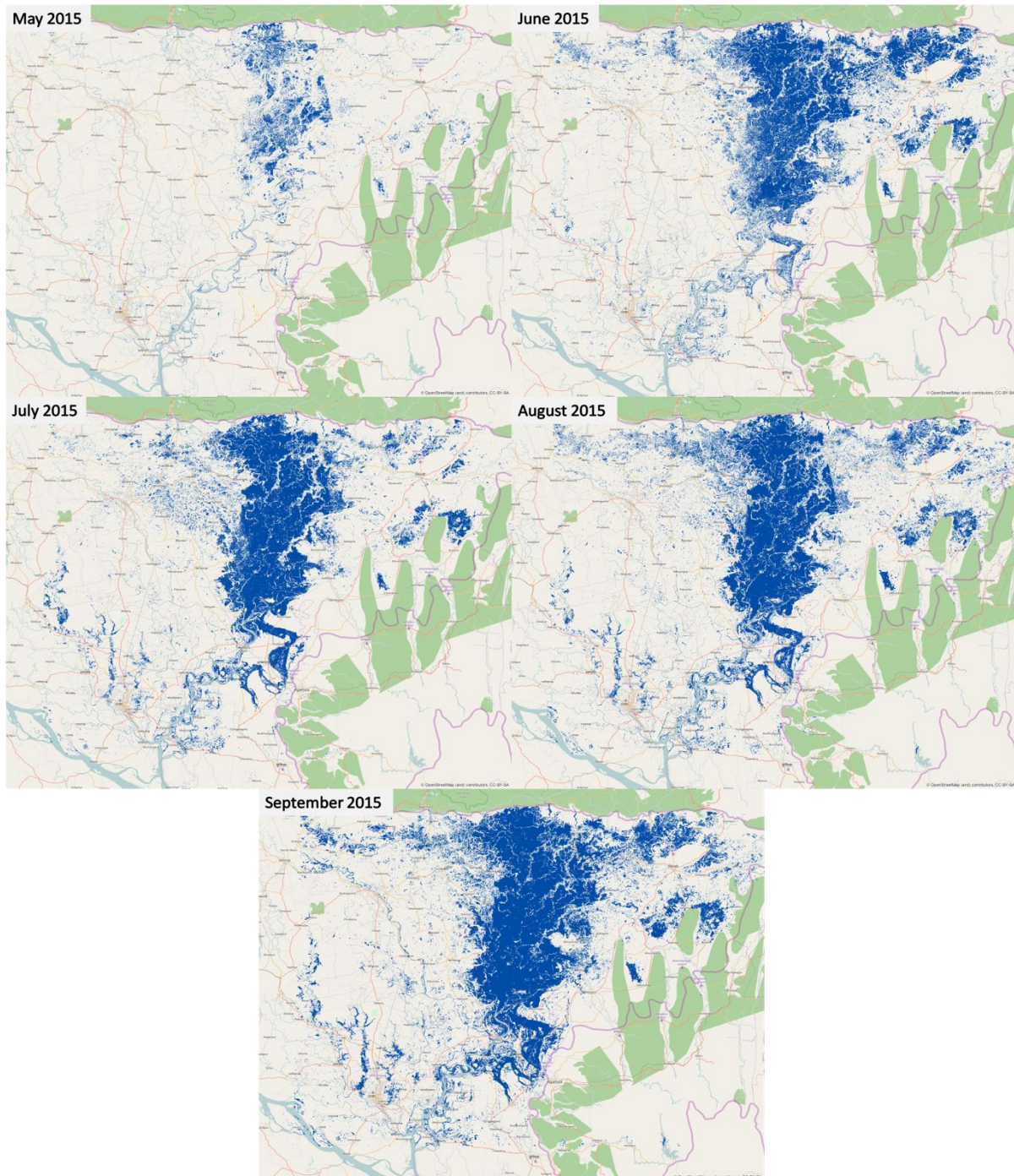


Figure 2C.1. Flood maps for **Bangladesh 2015** flood, from May to September, derived from **S1** images

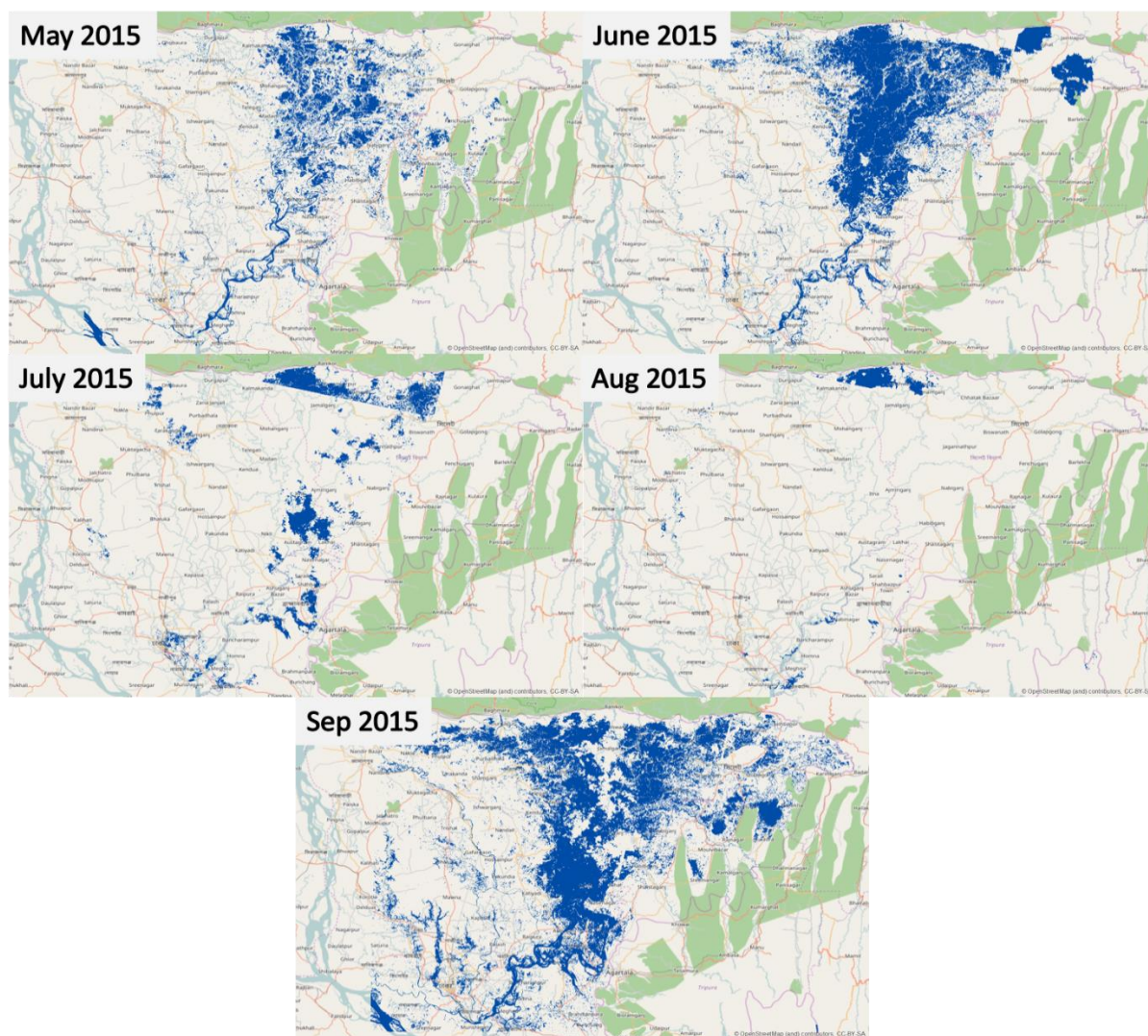


Figure 2C.2. Flood maps for **Bangladesh 2015** flood, from May to September, derived **from Landsat 8 images**. The difference from the maps based on S1 is evident, especially for July and August. In fact, the differences depend mainly on two factors: (1) observation frequency/duty cycle of the instrument, and (2) cloud cover affecting optical data. The two sensors, having different frequency of acquisition and different duty cycle (acquisition time during an orbit) may or may not capture the same flood extent. Moreover, the cloud cover affecting optical data may prevent the observation of the flood, totally in case of 100% of cloud coverage. For instance, in August it is not possible to observe the flood with L8 due to intense cloud cover.

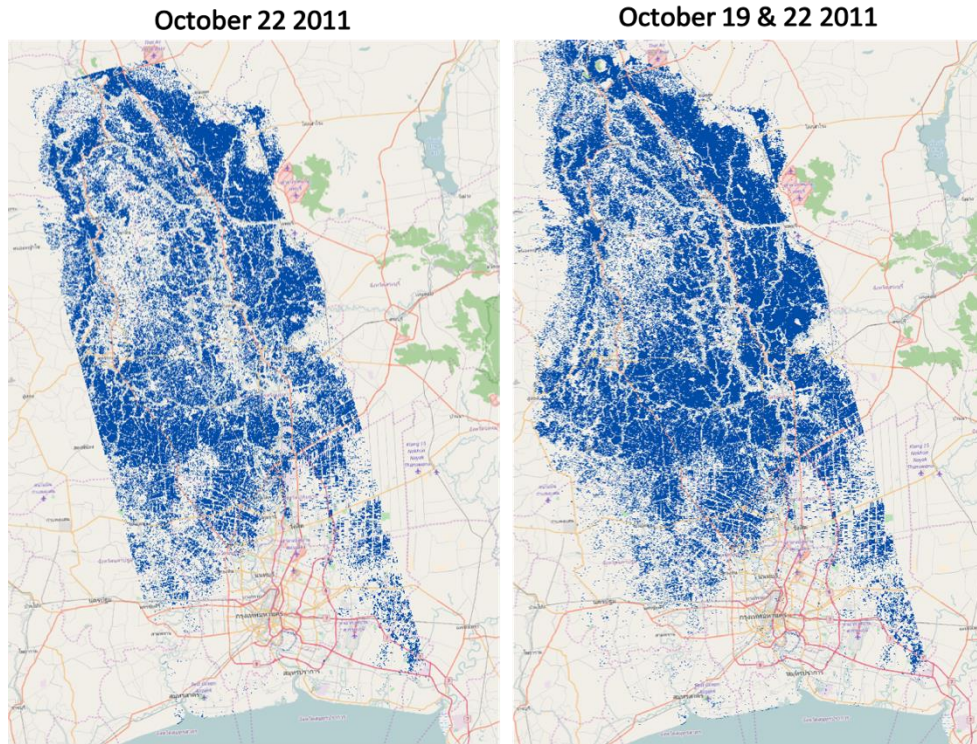


Figure 2C.3. Flood maps for Thailand October 2011 flood, derived from ASAR images

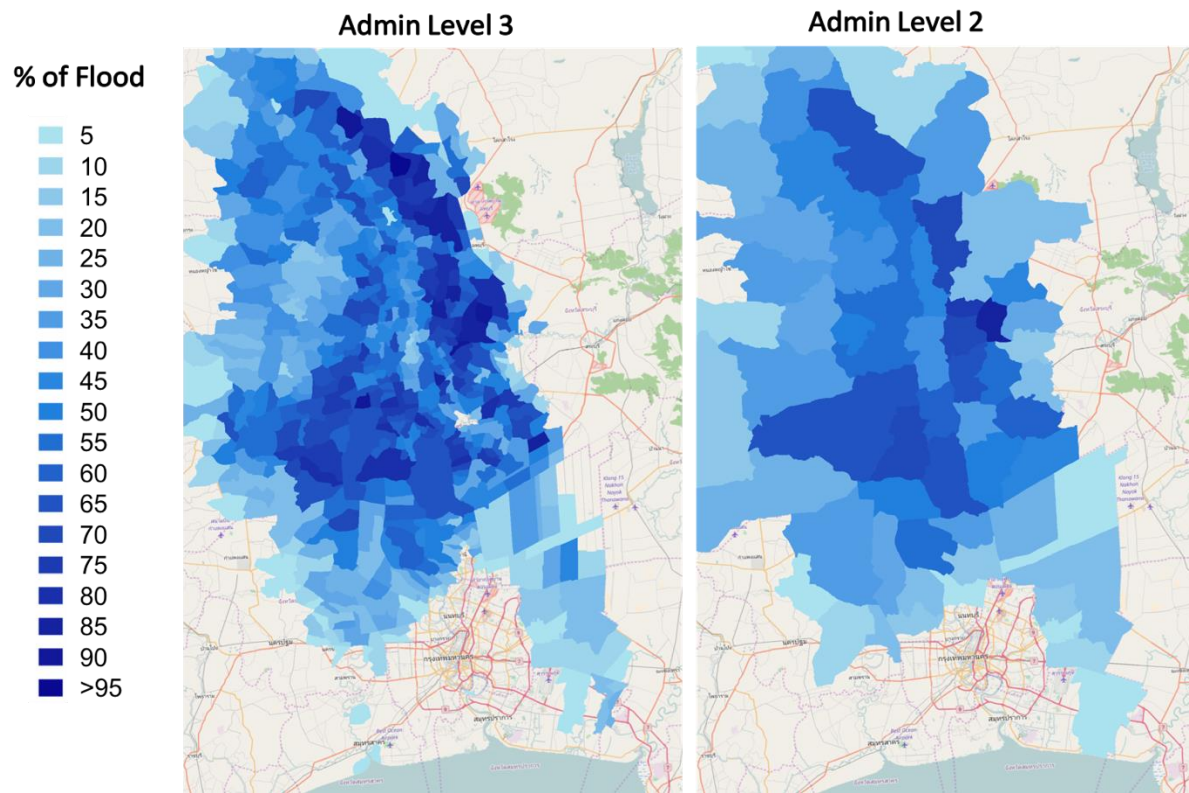


Figure 2C.4. **Percentage of flooded area** has been computed for each province and district, i.e., administrative level 3 and 2. The map shows in a shade of blue (from light to dark blue) the percentage of area flooded.

Annex 2D. Flood Occurrences

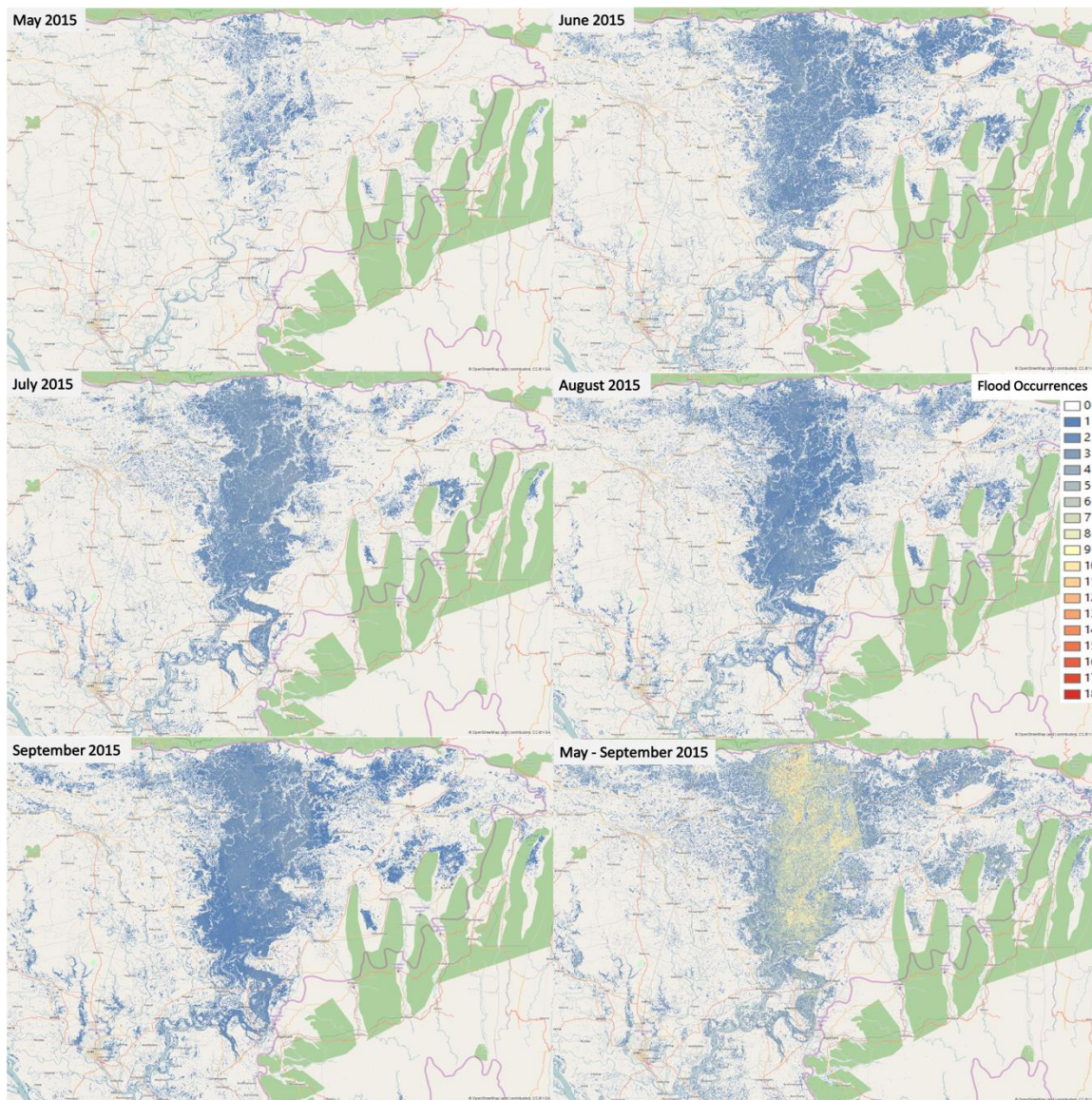


Figure 2D.1 Flood occurrences for Bangladesh 2015 flood. The scale indicates the number of times water was detected for each pixel, from blue (1 time) to red (18 times, which is the maximum value found for this case study).

Annex 2E. Flood Depth Estimation

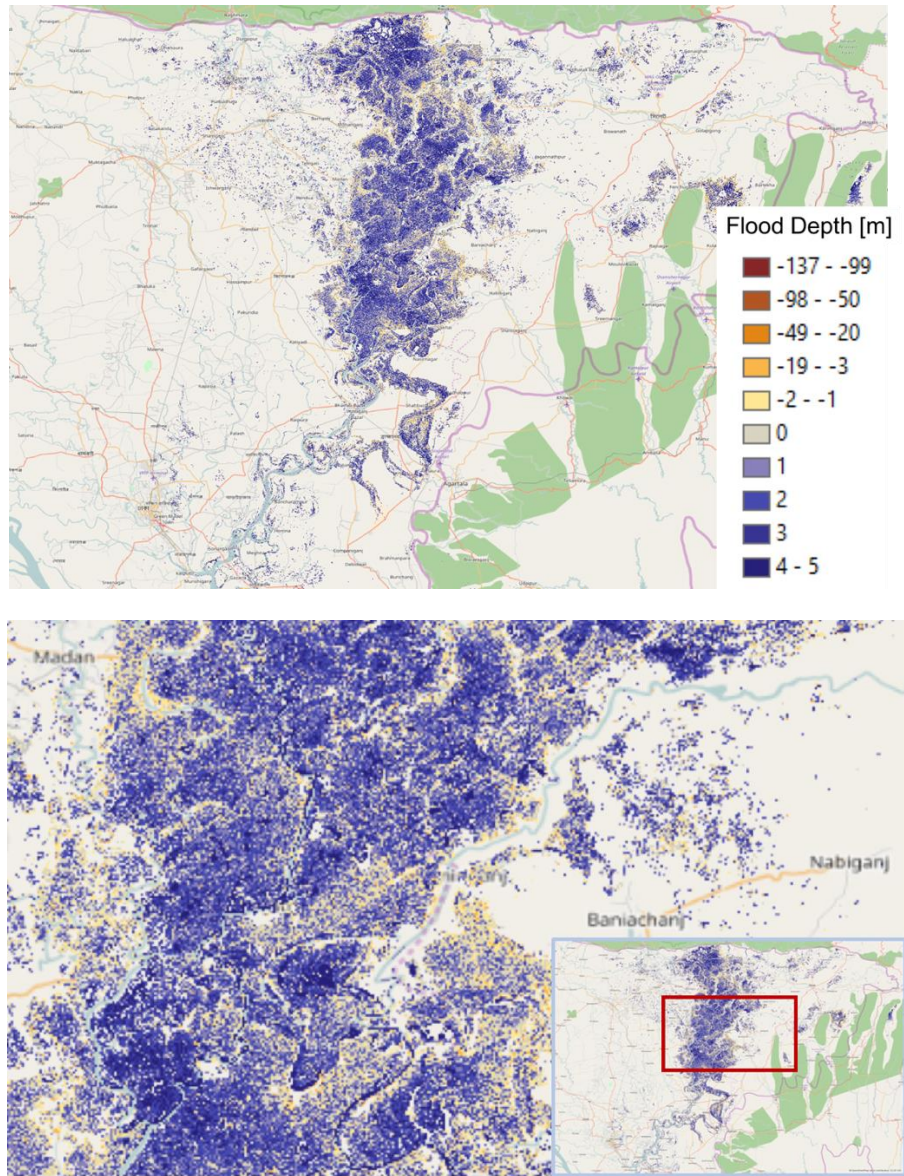


Figure 2E.1 Flood depth calculated for the flood extent of July 2015 in Bangladesh. As indicated in the scale bar, yellowish pixels indicated negative flood depth values. This means that the estimation of the water level was wrong. However, since negative values occur mainly at the edges of the flooded areas, this indicates a certain accuracy in the estimation, with errors mainly due to the coarse SRTM resolution, which does not allow a very precise water level estimation for every area. The errors are mainly in between 0 and -2 meters, which are reasonable considering the 1-meter vertical resolution of the DEM (see the zoom in the following image).

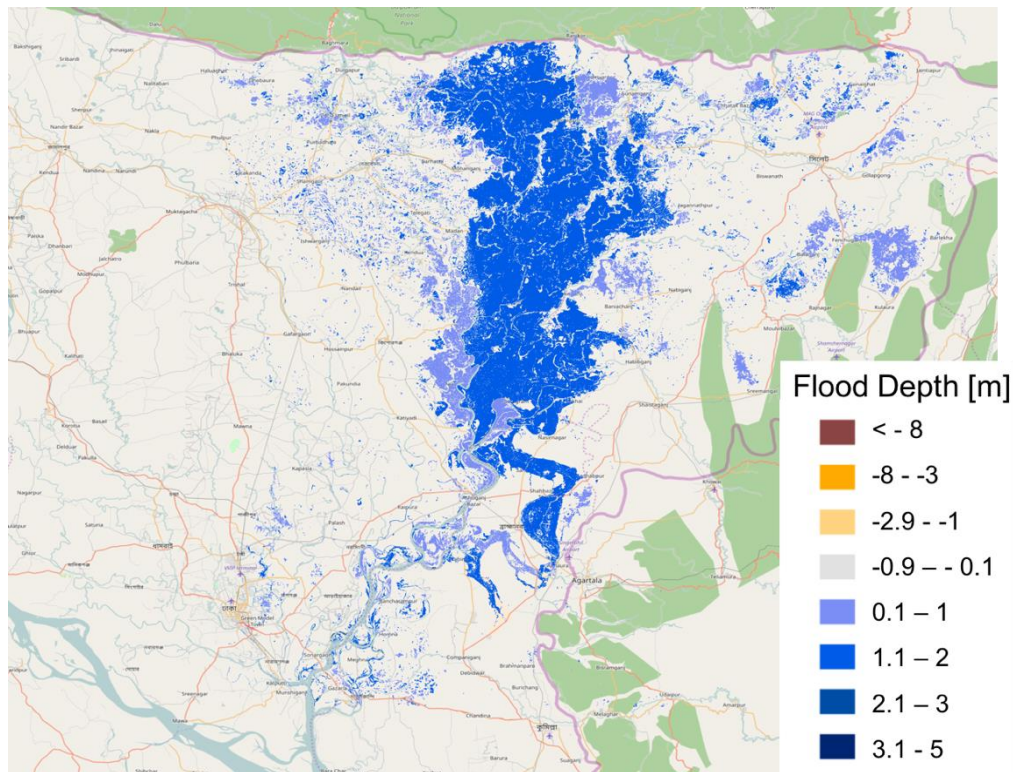


Figure 2E.2. The image shows the result when the median value of the flood depth for each flooded area was computed. As expected, negative values, being not frequent, disappear. This approximation means that we are supposing that the area is completely flat, which is a big approximation. Locally we may introduce an error in the order of 5 meters (the maximum value found was 5 meters; using the median the minimum value is close to zero), which would spoil completely the assessment of economic impacts. However, reasoning with a broader view, this value may be useful for the assessment of economic impacts. In Phase 2 of this project, we plan to explore more solid ways to improve flood depth estimation by means of low-resolution DEM.

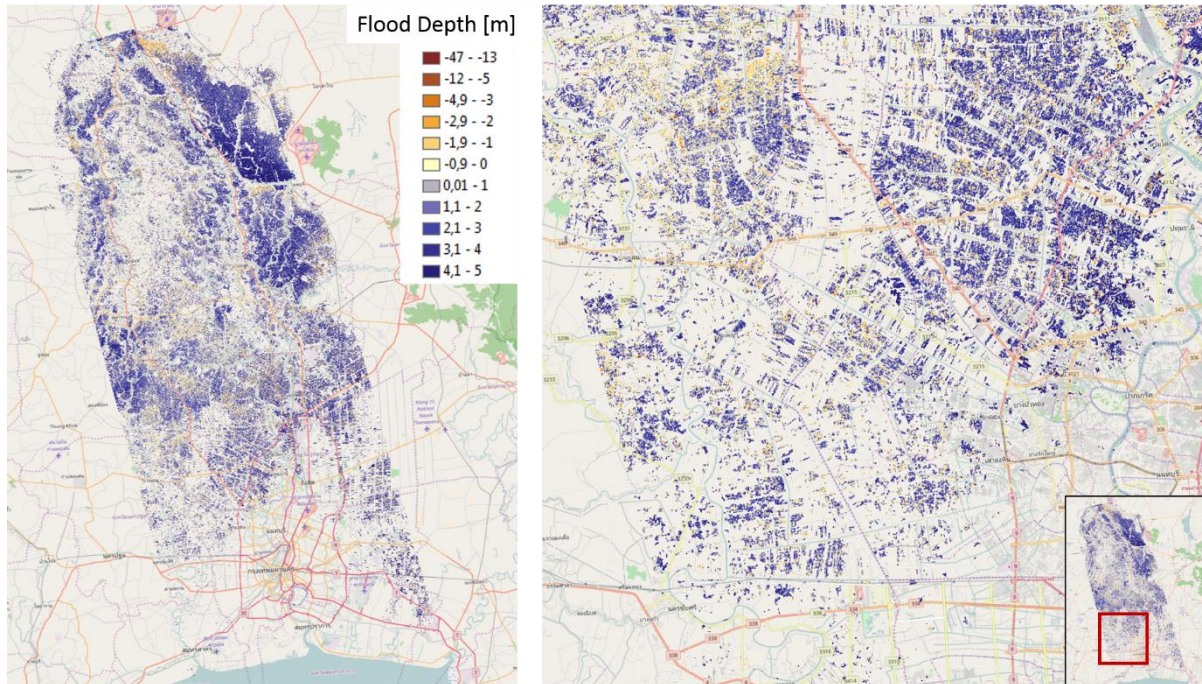


Figure 2E.3. Flood depth estimation for Thailand, October 22, 2011. Left: overview of the whole area of interest. Right: detail of an area near Bangkok where it is possible to appreciate the resolution of the result. The fluctuation of values due to the low resolution of the DEM is evident. However, negative values (i.e., wrong depth estimation) are mainly limited to -1 meter, therefore acceptable.

Annex 2F. Airbus Defence - WorldDEM

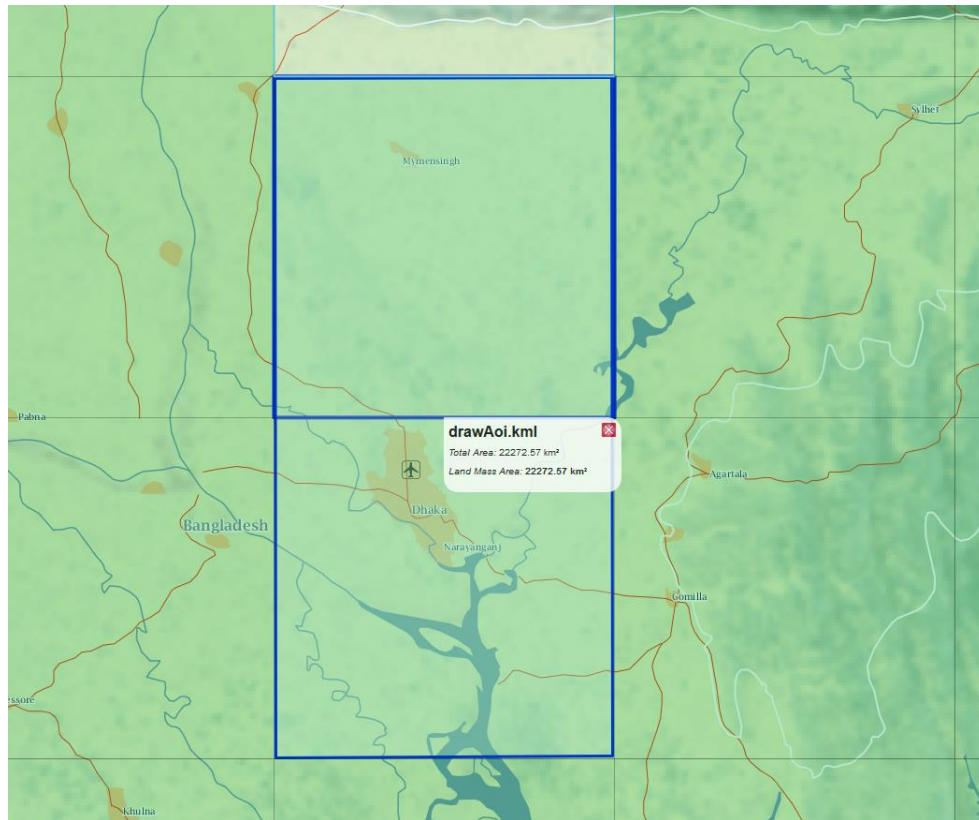


Figure 2F.1. To cover the area of interest for the flood in Bangladesh, we would need four DEM tiles, each 10,000 km². By shrinking the area of interest to only two tiles and purchase of the most basic product, and with the WorldDEM™ at the cost of 10€/km², the total cost would be €200,000 (10,000 km² x 2 x €10/km²).

Area of 1 tile ca. 10,000 km²

Delivery time starting from 10 days

Tiles of interest for Bangladesh: **N24 E89; N24 E90; N23 E89; N23 E90**

GEO Elevation

Elevation10			WorldDEM™			
Product	Description	Price	Product	Description	Price	
Elevation10 DSM _{raw}	DSM _{raw} : output of radar-grammetric processing using StripMap data pairs without any post-processing.	€ 16.00 / km ²	WorldDEM _{core}	WorldDEM _{core} : output of interferometric processing of StripMap data pairs without any post-processing.	€ 8.00 / km ²	
Elevation10 DSM _{basic}	DSM _{basic} is derived from the DSM _{raw} representing the surface of the Earth including buildings, infrastructure and vegetation.	Price DSM _{raw} + € 1.00 / km ²		WorldDEM™	WorldDEM™ is produced based on WorldDEM _{core} representing surface of Earth (incl. buildings, infrastructure and vegetation). Hydrological consistency is ensured.	Price WorldDEM _{core} + € 2.00 / km ²
Elevation10 DSM _{hydro}	DSM _{hydro} is derived from the DSM _{basic} , ensuring hydrological consistency.	Price DSM _{basic} + € 3.00 / km ²	WorldDEM DTM		In additional editing steps, WorldDEM™ will be transformed into a Digital Terrain Model (DTM) representing elevation of bare Earth terrain.	Price WorldDEM™ + € 6.00 / km ²
Elevation10 DTM	In additional editing steps, the DSM _{hydro} will be transformed into a Digital Terrain Model (DTM) representing elevation of bare Earth terrain.	Price DSM _{hydro} + € 10.00 / km ²			Optional	Additional Quality Layers (with WorldDEM™) including water body mask
Optional	Additional information layers • ORI: Orthorectified radar image layer (single look direction, no mosaicking) • Quality layers (incl. water body mask)	€ 3.00 / km ² € 1.00 / km ²	WorldDEM Bundle	Includes WorldDEM™, DTM, and quality layers		€ 19.00 / km ²
Elevation10 Package	Includes DSM _{hydro} , DTM, ORI and quality layers	€ 39.00 / km ²				

■ Conditions:

Elevation10:

- Min. order size (per AoI): 500km²
- Width of AoI ≥ 20km

WorldDEM™:

- Minimum order size (per AoI): 100km²
- Width of AoI ≥ 1km for WorldDEM™
- Width of AoI ≥ 2km for WorldDEM DTM
- Width of AoI ≥ 2km for WorldDEM Bundle

Table 2F.1. Description and Prices of Available AIRBUS DEM Products

■ GEO Elevation Pricing Summary:

Elevation10		WorldDEM™	
Product	Price	Product	Price
Elevation10 DSM _{raw}	€ 16.00 / km ²	WorldDEM _{core}	€ 8.00 / km ²
Elevation10 DSM _{basic}	€ 17.00 / km ²	WorldDEM™	€ 10.00 / km ²
Elevation10 DSM _{hydro}	€ 20.00 / km ²		
Elevation10 DTM	€ 30.00 / km ²	WorldDEM DTM	€ 16.00 / km ²
Elevation10 Package	€ 39.00 / km ²	WorldDEM Bundle	€ 19.00 / km ²
Option: ORI	€ 3.00 / km ²	Option: Quality Layers	€ 1.00 / km ²
Option: Quality Layers	€ 1.00 / km ²		

Annex 2G. Flood Maps Cross-Comparison L8 – S1

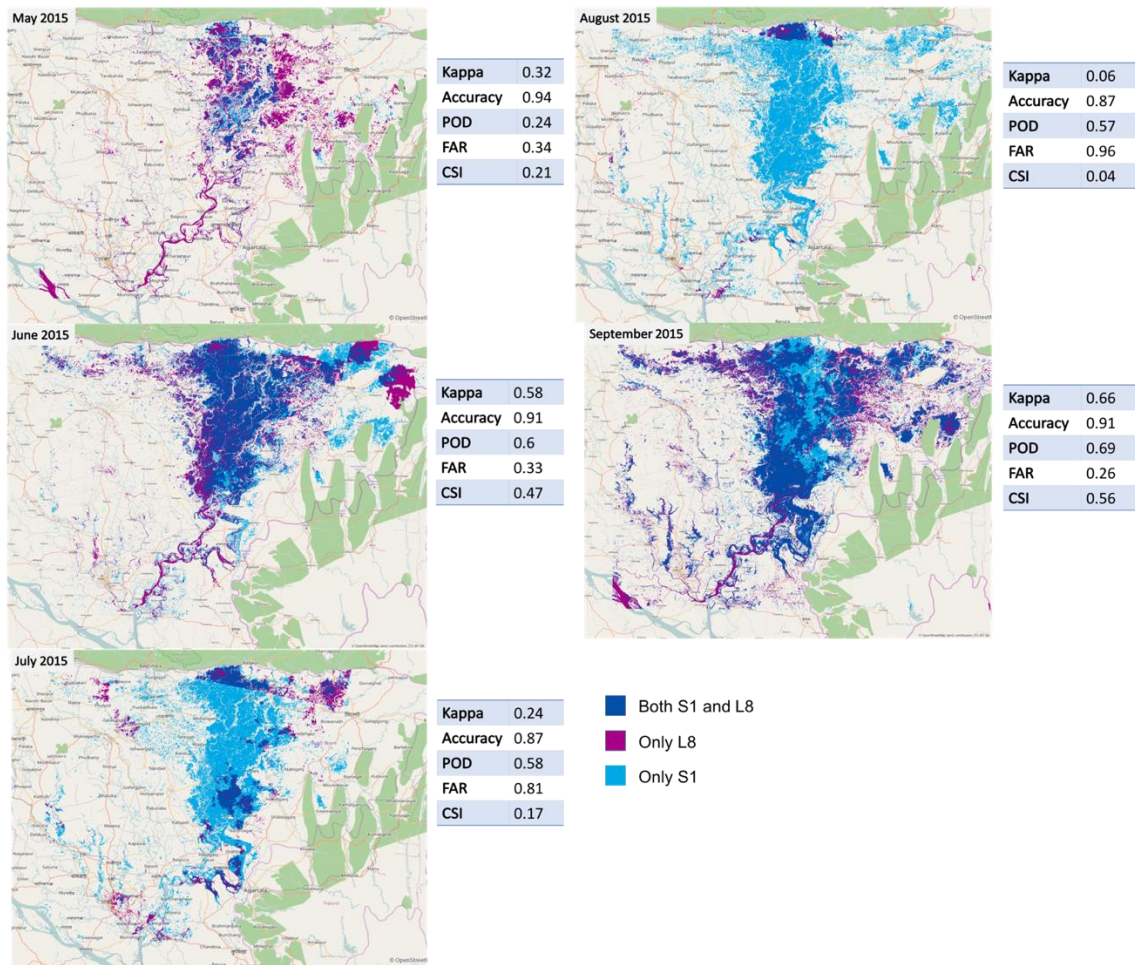


Figure 2G.1. Cross-comparison of flood maps obtained with Landsat 8 and S1 data for the months of May, June, July, August, and September 2015. For each cross-comparison, the table reports the values of the statistics employed.

Annex 2H. Improved Cross-Comparison of L8 and S1 Flood Maps

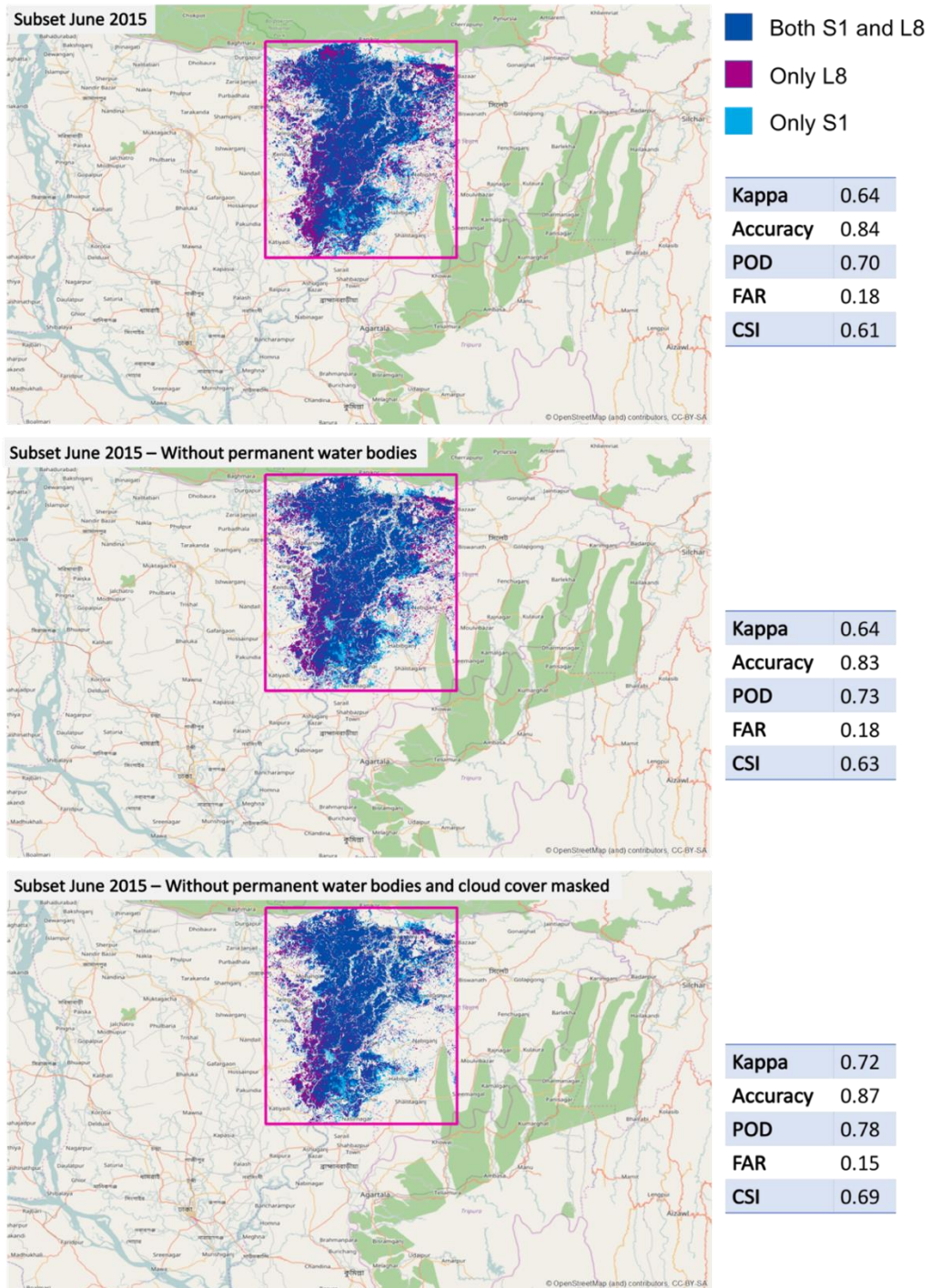


Figure 2H.1. Cross-comparison of flood maps obtained with Landsat 8 and S1 data for the month of June for a subset of the area of interest (top), removing permanent water bodies (middle), and masking cloud cover (bottom).

Annex 2I. Cross-Comparison of Flood Maps with NRT MODIS

Landsat 8-based maps

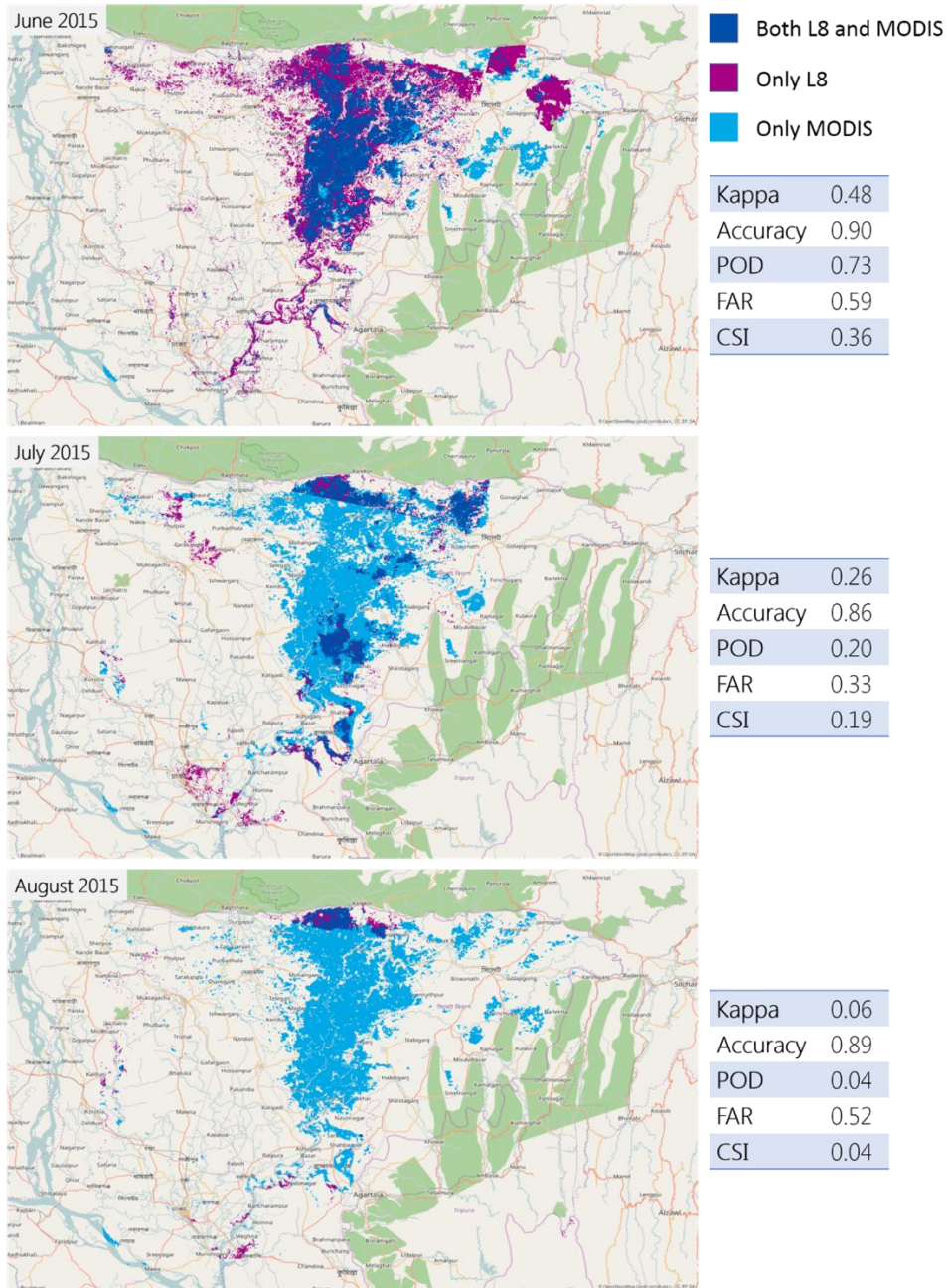


Figure 2I.1. Cross-comparison of flood maps obtained with Landsat 8 data with NRT-MODIS for the months of June, July, and August 2015. For each cross-comparison, the table reports the values of the statistics employed.

S1-based flood maps

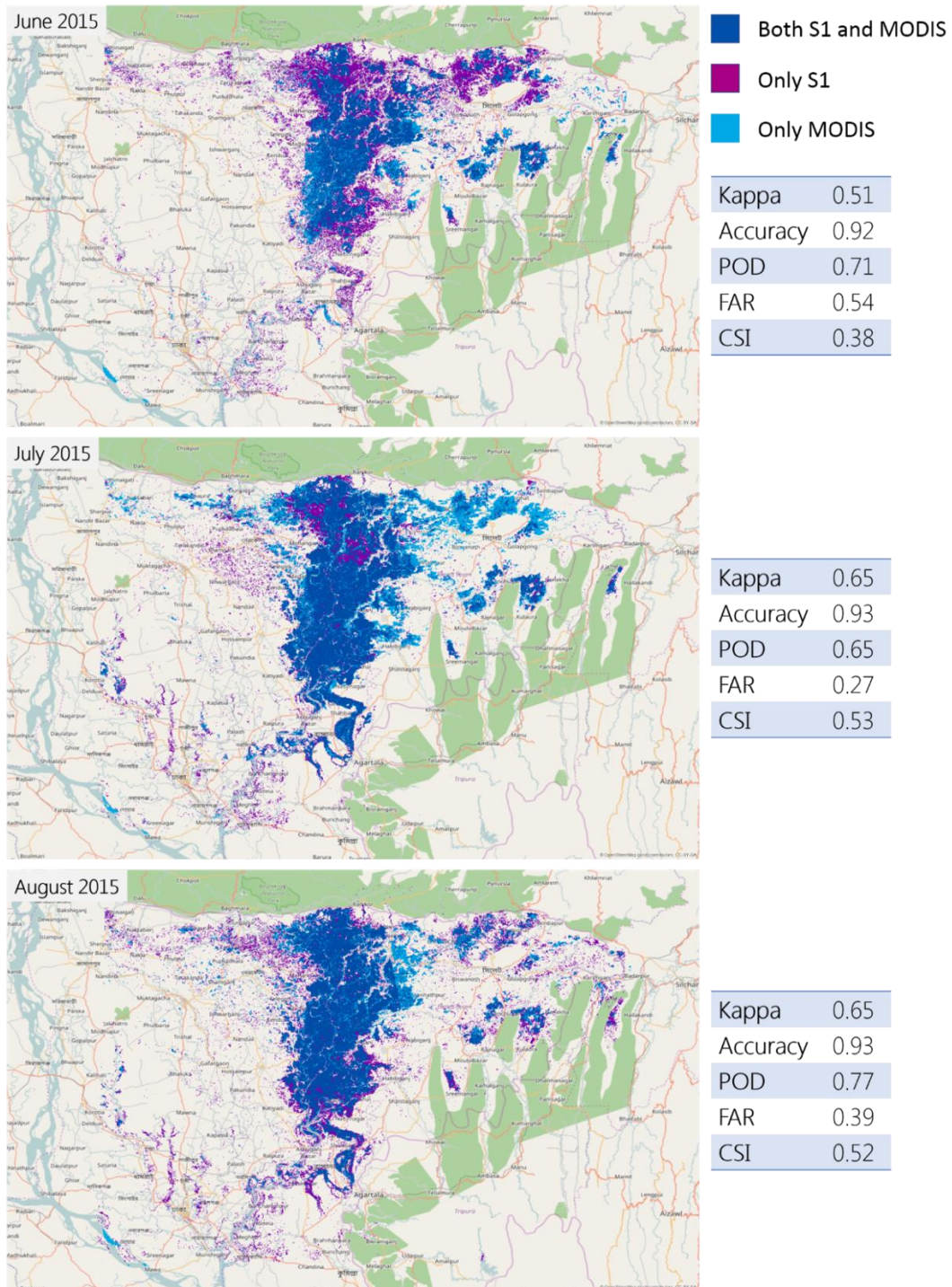


Figure 21.2. Cross-comparison of flood maps obtained with S1 data with NRT-MODIS for the months of June, July, and August 2015. For each cross-comparison, the table reports the values of the statistics employed.

L8 and S1 flood maps

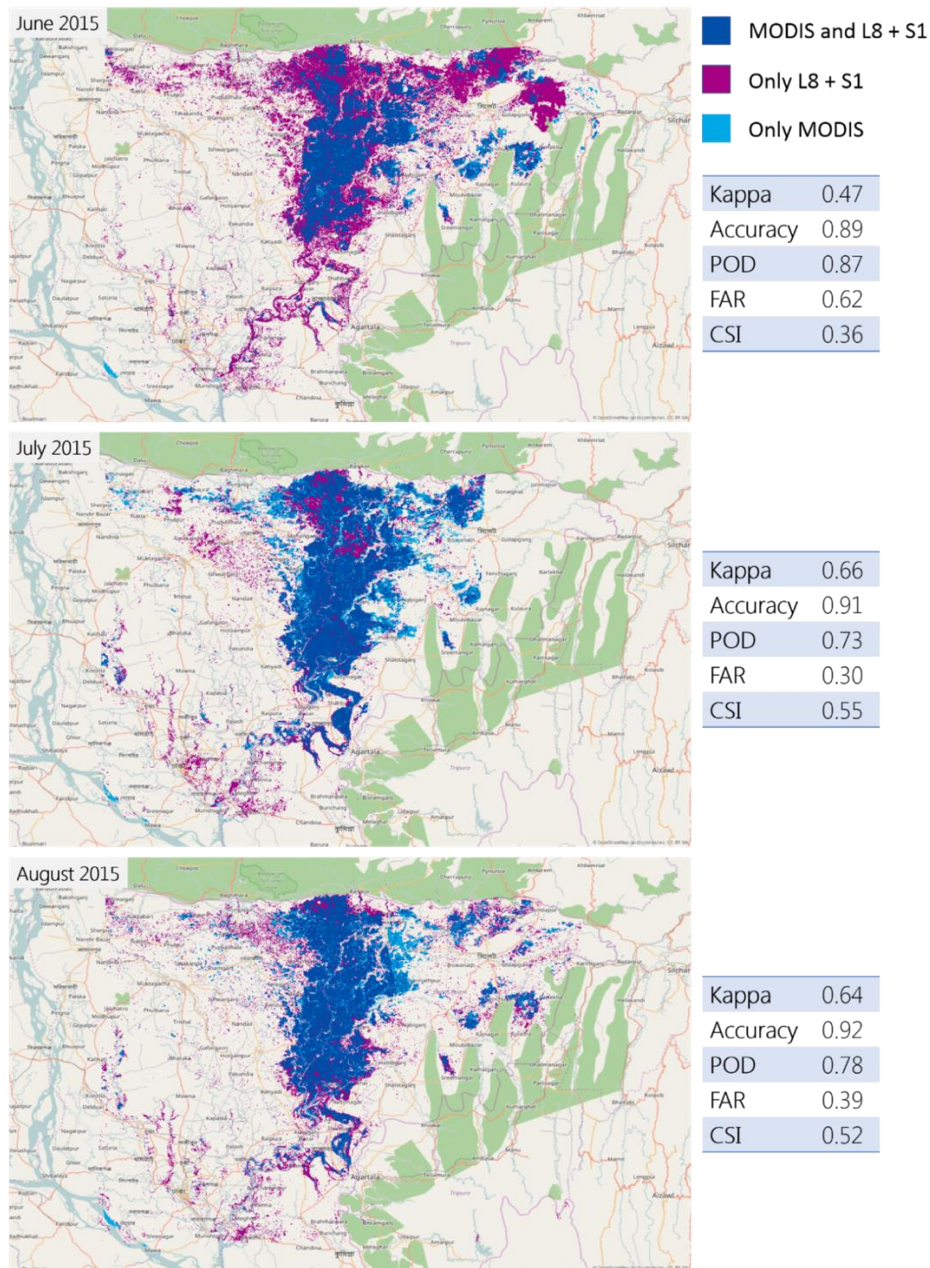


Figure 21.3. Cross-comparison of flood maps obtained with Landsat 8 and S1 data with NRT-MODIS for the months of June, July, and August 2015. For each cross-comparison, the table reports the values of the statistics employed.

Annex 2J. Cross-Comparison of GFMS Products with L8 and S1 Flood Maps

2015

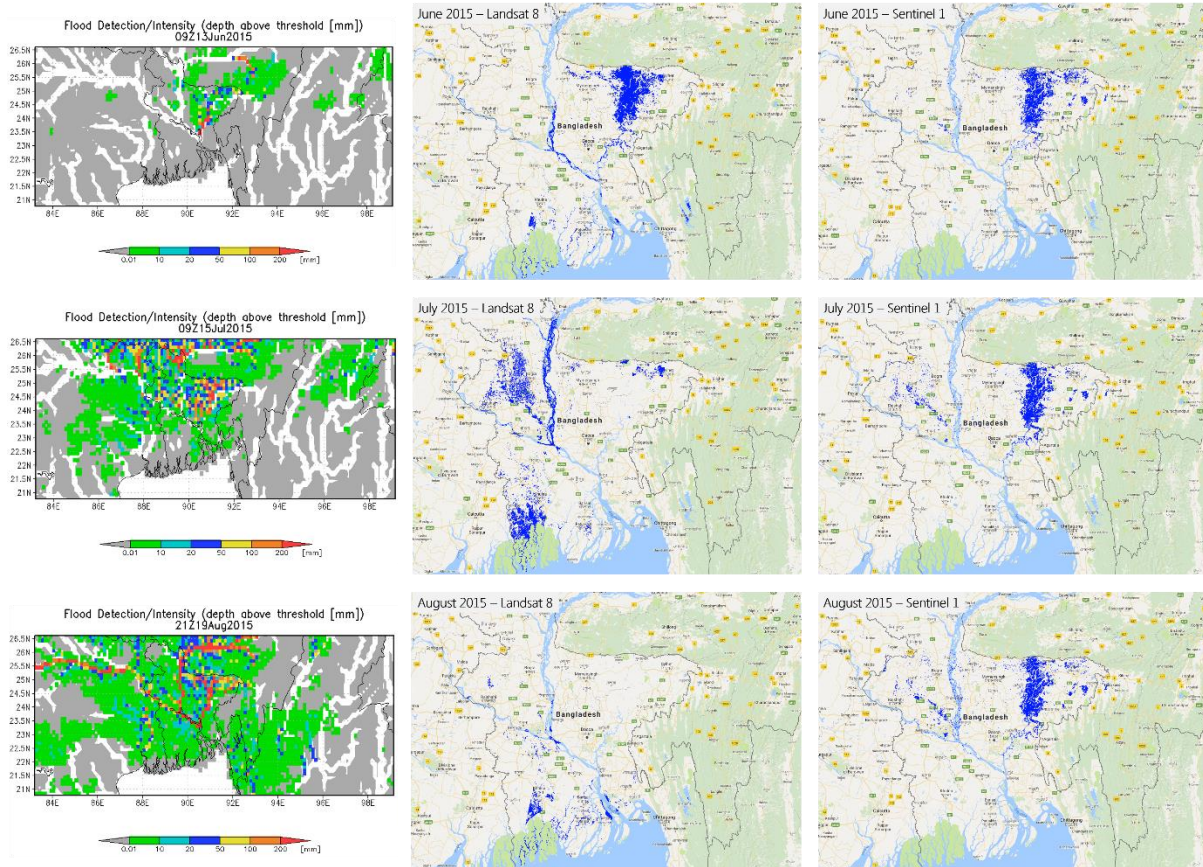


Figure 2J.1. Qualitative cross-comparison between GFMS products and L8 and S1 flood maps for the period June-August 2015 for Bangladesh

2016

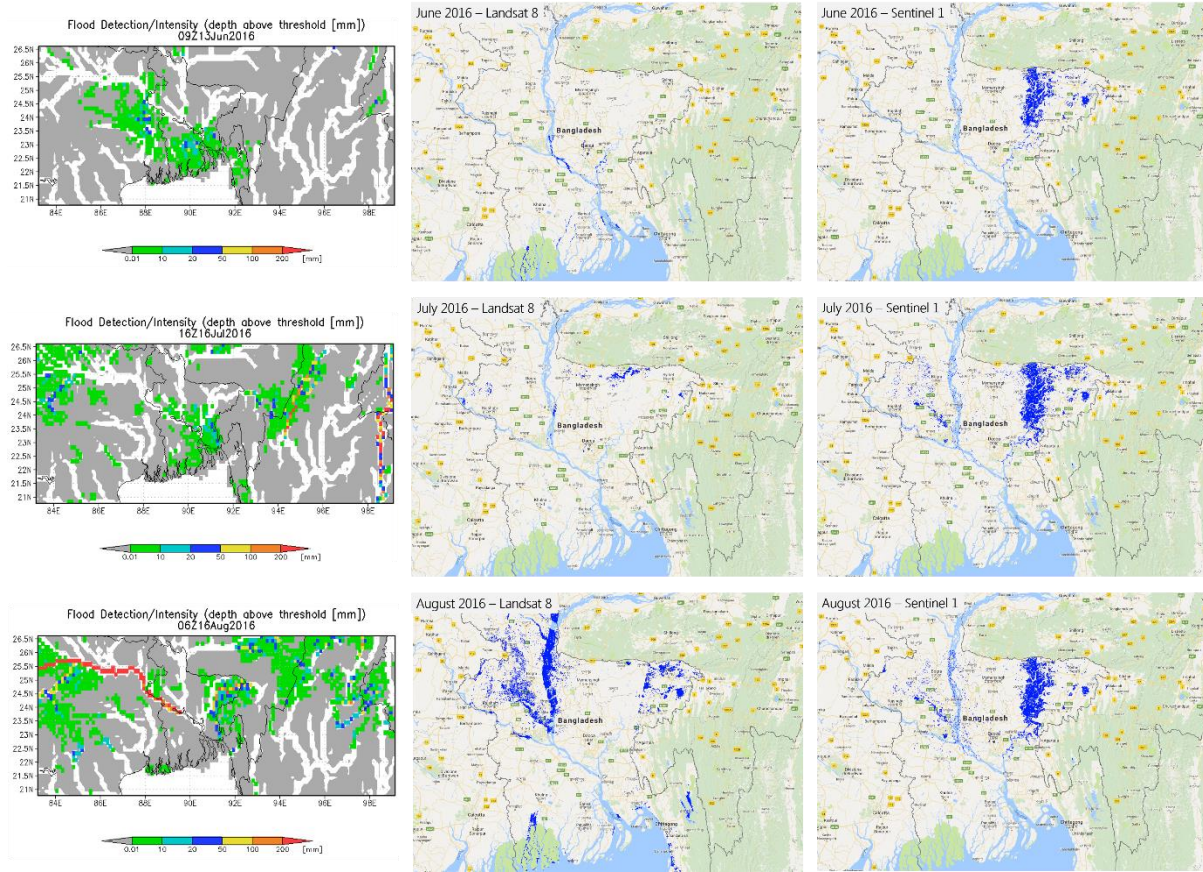


Figure 2J.2. Qualitative cross-comparison between GFMS products and L8 and S1 flood maps for the period June-August 2016 for Bangladesh

Annex 2K. Cross-Comparison of Population Data Sets

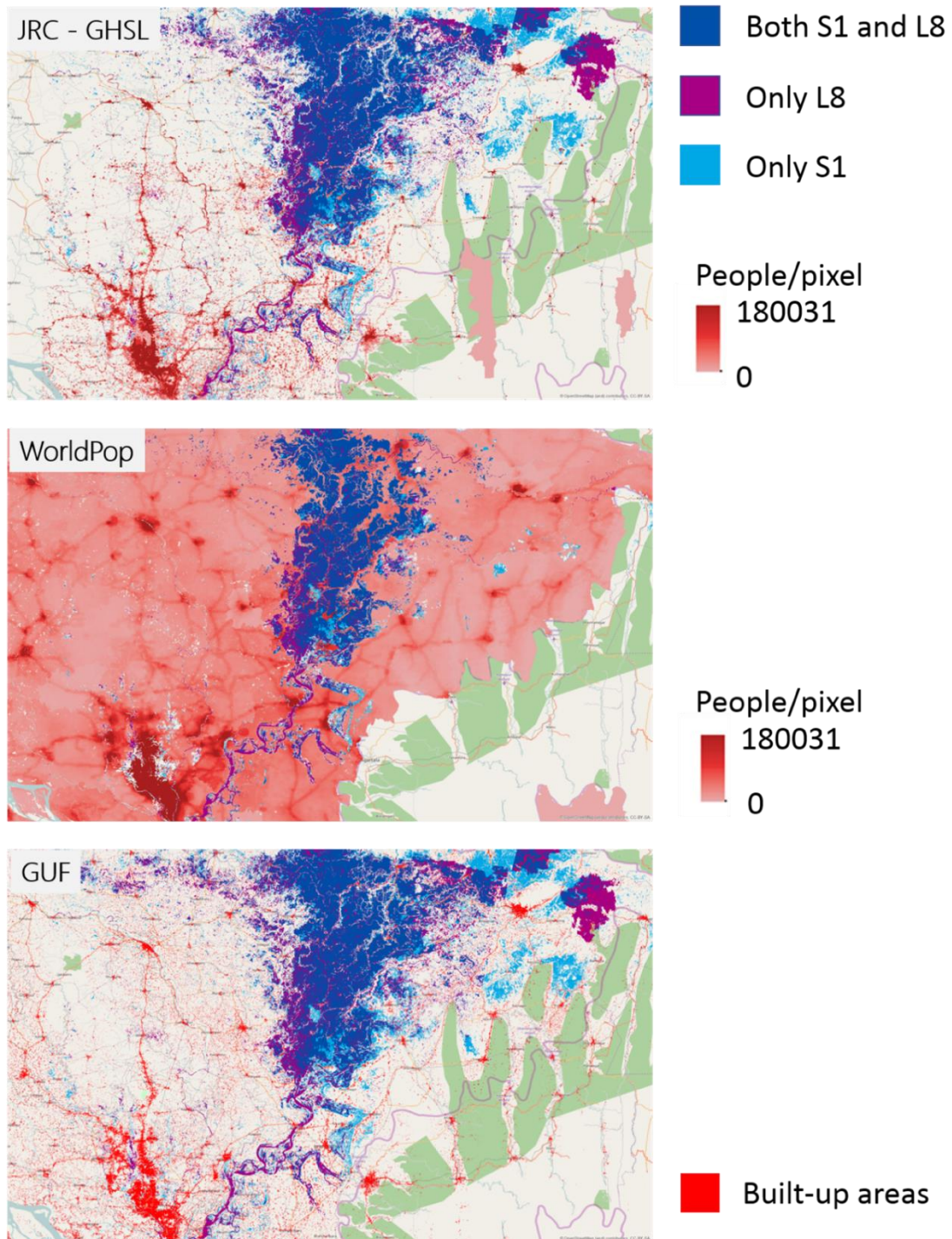


Figure 2K.1. **Top:** Global Human Settlements Population layer produced by the JRC overlaid on flood maps from LANDSAT, Sentinel-1, and combined products. **Middle:** WorldPop layer superimposed to the flood of June 2015. **Bottom:** Global Urban Footprint produced by DLR superimposed to the flood of June 2015.

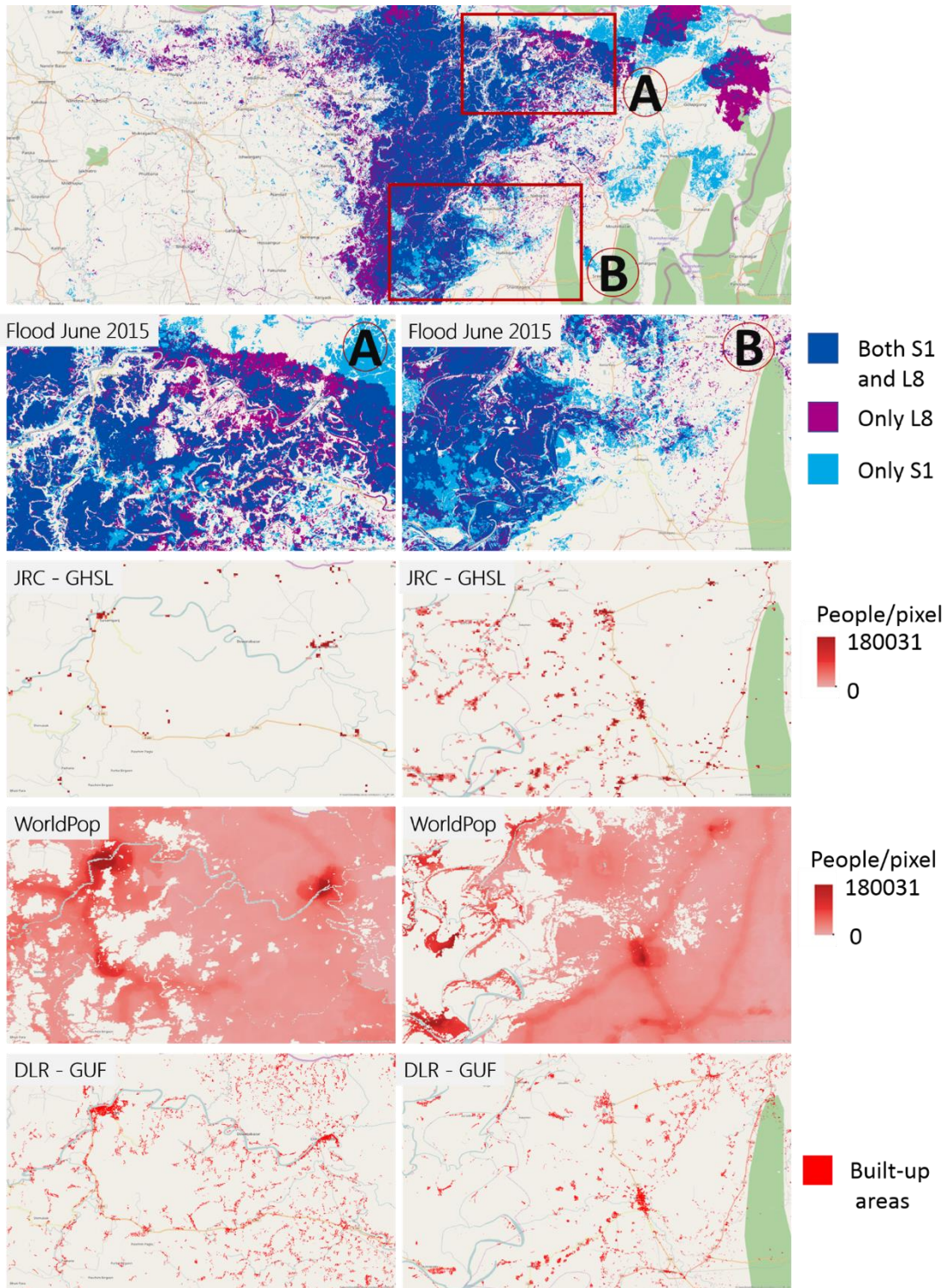


Figure 2K.2. Cross-comparison of different population data sets. A zoom in two areas affected by the flood of June 2015 shows the differences of the three data sets.

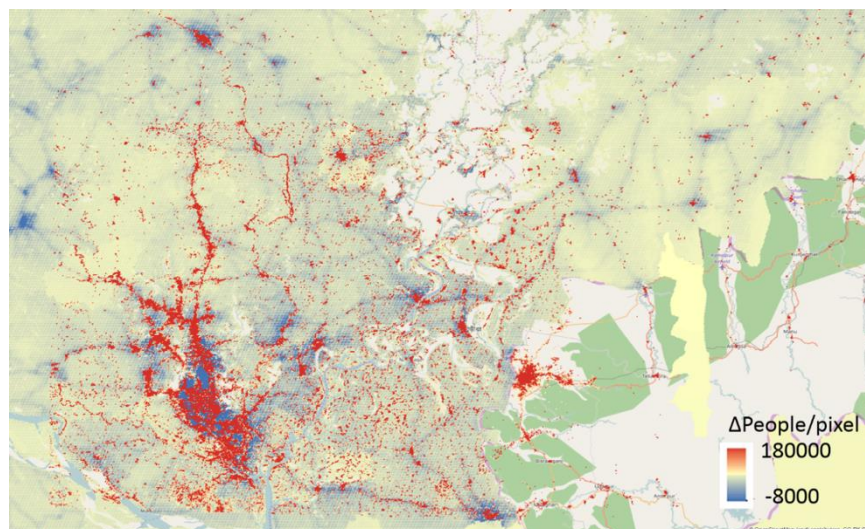


Figure 2K.3. Difference between the JRC GHSL and the WorldPop data set

Flood Map	GUf [Km ²]	WorldPop [million People]	JRC – GHSL-POP [million People]
S1 only	4.96	0.838	1.520
L8 Only	30.04	2.021	1.741
S1 and L8	1.52	1.042	1.291
S1 Tot	6.48	1.880	2.811
L8 Tot	31.56	3.062	3.032
S1+L8	36.52	3.901	4.552

Table 2K.2 Population Exposed to the Event Based on Different Population Data Sets for the Flood in Bangladesh in July 2015

	GUF [Km²]	WorldPop [million People]	JRC – GHSL-POP [million People]
S1 only	6.93	0.246	1.470
L8 Only	37.81	0.272	2.478
S1 and L8	0.96	0.132	0.311
S1 Tot	7.89	0.378	1.781
L8 Tot	38.77	0.403	2.789
S1+L8	45.7	0.649	4.259

Table 2K.2 Population Exposed to the Event Based on Different Population Data Sets for the Flood in Bangladesh in July 2015

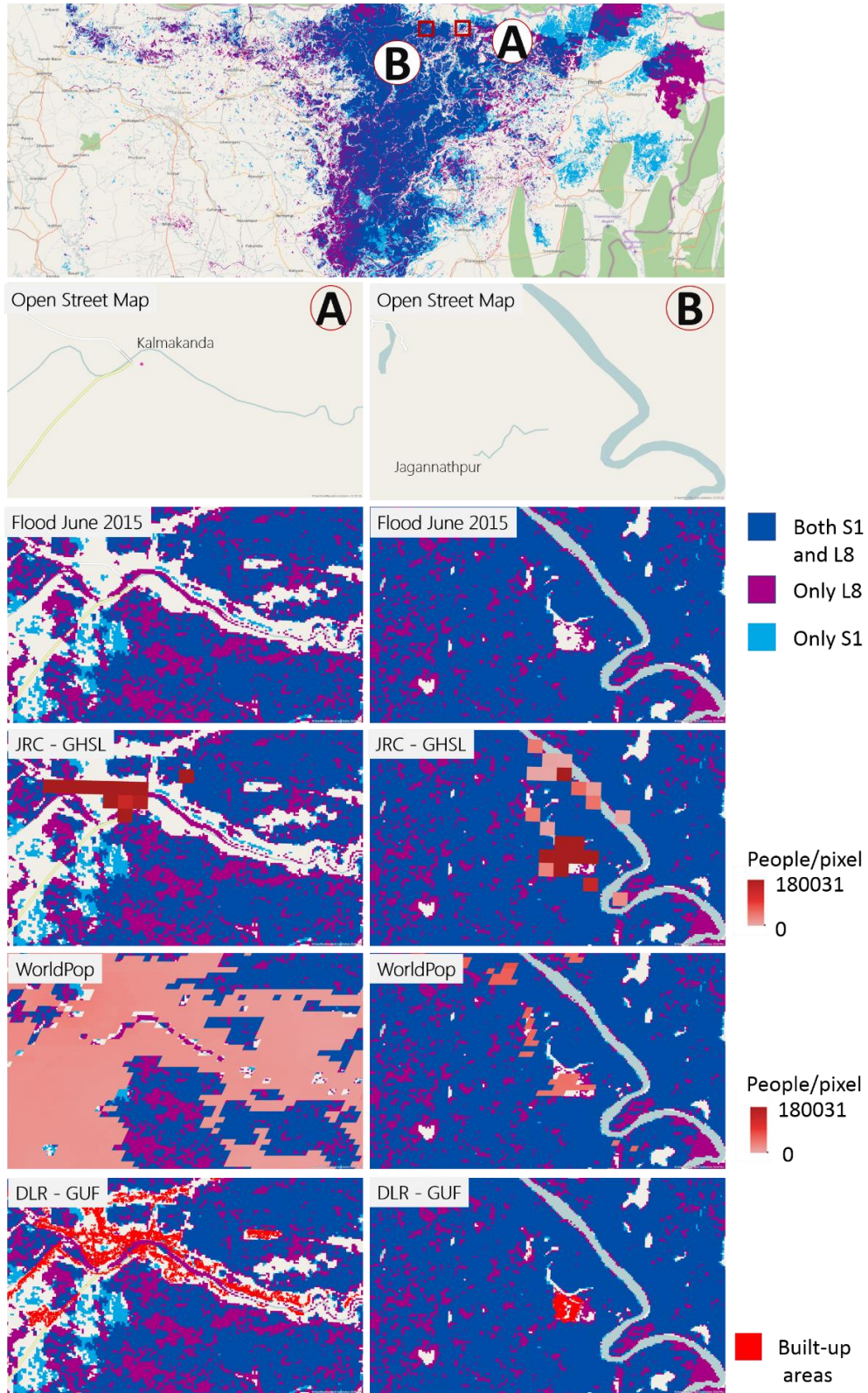


Figure 2K.4. Difference in exposure due to the different data sets

Annex 2L. Estimation of Yearly Surface Water Area

Year	Seasonal Water [Km ²]	Year	Seasonal Water [Km ²]
1984	0	2000	1.765.355.296
1985	0	2001	1.609.937.378
1986	0	2002	1.726.924.043
1987	1.783.817.543	2003	1.334.955.644
1988	2.142.843.334	2004	220.854.722
1989	1.601.588.355	2005	1.790.641.577
1990	1.459.737.465	2006	1.586.697.045
1991	1.699.440.436	2007	2.138.636.749
1992	7.898.126.465	2008	1.814.744.441
1993	1.428.054.532	2009	1.465.312.301
1994	9.170.815.631	2010	152.334.463
1995	3.218.126.524	2011	174.728.017
1996	3.742.900.387	2012	1.354.717.942
1997	1.185.772.505	2013	1.522.189.267
1998	0	2014	1.663.482.805
1999	1.397.740.193	2015	1.647.810.753

Table 2L.1. Estimation of Yearly Surface Water Area in Bangladesh Detected by Landsat from 1984 to 2015

Year	Seasonal Water [Km ²]	Year	Seasonal Water [Km ²]
1984	0	2000	8.170.858.639
1985	0	2001	7.530.938.763
1986	0	2002	91.883.444
1987	1.686.377.518	2003	6.579.454.513
1988	6.517.878.163	2004	8.161.864.173
1989	622.688.045	2005	832.207.395
1990	7.815.729.589	2006	1.178.788.699
1991	3.184.432.794	2007	8.055.437.958
1992	1.708.879.419	2008	9.259.733.444
1993	7.243.415.602	2009	8.408.409.409
1994	3.111.412.778	2010	1.004.011.635
1995	3.036.518.832	2011	133.766.084
1996	6.747.998.987	2012	6.723.351.506
1997	1.803.257.885	2013	9.489.592.354
1998	3.910.833.055	2014	7.479.152.267
1999	5.246.018.497	2015	4.904.897.594

Table 2L.2. Estimation of Yearly Surface Water Area in Thailand Detected by Landsat from 1984 to 2015

Section 3. Precipitation and Flood Return Period Analyses

Masahiko Haraguchi, Columbia Water Center

3.1 Summary

Several detailed investigations into the exceedance probability for flood event impacts were pursued with the goal of calculating return periods of floods in Bangladesh and Thailand. Historic data on flood hazards were used to estimate significant variables for flood areas, durations, and damage in order to calculate the return period of floods. This is preliminary work supporting a future view of pricing of any risk transfer mechanisms based on the impact assessment methodology. The key findings are the following:

Thailand

1. The association between flood hazards and rainfall amounts with different rainfall windows was first examined using both the full flood data and top 20 events in terms of flooded areas, durations, and damage. Flood data since 1985 were obtained from Dartmouth Flood Observatory (DFO). The results showed that 30-day windows are more dominant than other rainfall windows.
2. Using 5-day and 30-day windows of rainfall, return periods of 2-, 20-, and 100-year levels were calculated using generalized Pareto distributions of extreme value statistics. For the 5-day window, the estimated return levels were 84 mm/5 days, 91 mm/5 days, and 94 mm/5 days for the 2-year, 20-year, and 100-year levels respectively. For the 30-day window, the estimated return levels were 391 mm/30 days, 414 mm/30 days, and 425 mm/30 days, respectively.
3. The limitation of this approach is that we took averages of precipitation amounts that fall in the entire country.

Bangladesh

1. The same methodology was used to examine the association between flood hazards and different windows of rainfall. However, due to the complex nature of floods in Bangladesh, streamflow data (discharge and water level) for primary river systems (the Brahmaputra, Ganges, and Meghna Rivers) were added to the analysis. In addition, due to the complexity in DFO's flood data in Bangladesh, the annual data on flood-affected areas collected by the Bangladesh Water Development Board (BWBD) were added.
2. Using DFO's data of flood areas, it was hard to establish an association between flood areas and water levels in the three main rivers and flood areas for the lower and middle ranges of flood areas. In contrast, the upper 10th to 20th percentile events showed some relationships between water-level data and flood areas. The top 10 events consistently corresponded to positive standardized anomalies in the maximum water level in each location. This implies that there is a joint effect of streamflow from the three main river systems.

3. Local regression is used to fit a smooth curve among the predictors. The advantage of the local regression is that it can relax the linearity assumption of conventional regression methods. Based on generalized cross-validation scores, for the BWBD's flood area data, the selected predictor was the maximum water level in a rainy season. In contrast, for DFO's data, the maximum water levels in Rajshashi and Bhairab Bazar data were selected as the best predictor.
4. Because the analysis using the local regression showed that the linear model was adequate, the logistic regression was also conducted to select predictors of rainfall amounts and water levels, using the Least Absolute Shrinkage and Selection Operator (LASSO) method. The thresholds for the logistic regression were the upper 10th and 20th percentiles of flooded areas. For the BWBD's data that recorded the 10th percentile as floods, the best predictors were the maximum water levels in Rajshashi and Bhairab Bazar, which were consistent with the results of the previous local regression. For the BWBD's data that recorded the 20th percentile as floods, the maximum rainfall amounts for 3-day and 30-day windows and the maximum water level in Bhairab Bazar were selected as the best predictors. For the upper 10th percentile of the DFO's data, the maximum water level in Rajshashi was selected. For the upper 20th percentile data, the model using all variables except for the maximum water level in Bahadurabad was selected.

Conclusions and Recommendations

- 1) For Thailand, our recommendation is to use longer rainfall windows, such as a 30-day window, as a rainfall predictor. The estimated return levels are 13.0–16.8 mm/day, 13.8–18.2 mm/day, and 14.7–18.8 mm/day for the 2-year, 20-year, and 100-year levels, respectively. In addition, although this case analysis did not use streamflow data in Thailand, as presented in the case of Bangladesh, streamflow data on the ground might help predictions.
- 2) Flood mechanisms in Bangladesh seem complex because flows from India might be affecting flood hazards in Bangladesh. Therefore, rainfall recorded over Bangladesh is not effective for predicting floods in Bangladesh. Consequently, the river flows in the Brahmaputra and Ganges Rivers, which come from India, need to be considered. Our recommendation for predictors is to use data of the maximum water levels in Rajshashi (Ganges River) and Bhairab Bazar (Meghna River) in the combination of rainfall windows.
- 3) This case study shows that it is critical to collect data on flood hazards through both on-the-ground and remote sensing processing. In the case of Bangladesh, DFO, BWBD data, and remote sensing estimates demonstrated inconsistency. On the other hand, streamflow data will help estimate specific predictors in each river basin.
- 4) The qualitative assessment was also conducted to explore the applicability of the approaches. Considering their topography and river systems, all of Cambodia, Lao People's Democratic Republic, Myanmar, and Vietnam would need an approach similar to Bangladesh, although Myanmar could consider taking an approach similar to Thailand.

3.2 Thailand

Data

Flood areas data

Flood data (began and ended dates, flooded areas, flood durations, damages associated with floods) are obtained from Dartmouth Flood Observatory, and are available since 1985.

Precipitation data

The observational precipitation data are CPC 0.5 average (NOAA Climate Prediction Center's standardized metric) over the whole country from January 1, 1979, to January 9, 2017. Reanalysis data are from 1900 to the present (ERA/ ERA-interim, ECMWF). Three different windows are created: windows based on a day when a flood began, ones based on a day when a flood ended, and ones based on the middle of a flood period. But for Thailand, only the began-date data has higher correlations with different windows. Thus, only the began-date data are used.

Analysis

Correlation with Precipitation Amounts

The rainfall index is selected in the following ways:

- Identify the events for which we have data on flood area, duration, loss.
- Estimate the rainfall over the country or basins for windows of {1, 2, 3, 5, 10, 30 days} around the date of the event for each such event.
- Identify which window for rainfall gives the best correlation by ranks (Kendall's tau) with the flood statistics.
- Explore relation of flood loss to area, duration, and selected rain statistics to be sure that this is a good choice.
- Repeat with different sources of rainfall data as a check.

The results are shown in

Table 3.1. Though 30-day windows are more dominant than others such as in top 20 events in terms of areas, durations, and damages, there is no single variable to determine the relationship.

		Highest correlation
All	Area	w10 of began date data of reanalysis (0.23)
	Durations	w30 of began date data of CPC (0.28)
Top 20 events in terms of areas	Area	w30 of mid date data of CPC (0.33)
	Durations	w30 of mid date data of reanalysis (0.15)
	Damage	W1 –w2 of mid date data of reanalysis (0.36)
Top 20 events in terms of durations	Area	w30 of began date data of reanalysis (0.36)
	Durations	w30 of began date data of reanalysis (0.26)
	Damage	W1 of mid-date data of reanalysis (0.22)
Top 20 events in terms of damages	Area	W2 of began date of CPC , w4 of began date of reanalysis . (0.29)
	Durations	W5 of began-date data of CPC (0.31)
	Damage	W5 of began-date data of CPC (0.08)

Table 3.1. Highest Correlations between Different Windows of Rainfall and Areas/Durations/Damage (“All” means the analysis that uses all the data.)

Return Periods

Return periods are calculated in the following way:

1. Common 30-year period for flood attributes and rainfall
 - 1.1. For example, the rain statistic selected is the 30-day rainfall for the entire country centered around the flood event: W30c.
 - 1.2. Identify the smallest W30c value of flood-associated rainfall that led to a flood damage = W30cm.
2. Using 100-year rainfall data identify all W30 events that could cause a flood loss.
 - 2.1. Identify all rainfall amounts using a 30-day moving window.
 - 2.2. Retain all events for which W30c > W30cm.
 - 2.3. Choose the top 10th percentile of the W30c values as a threshold.
 - 2.4 Use the generalized Pareto distribution to estimate the return period of any potential trigger event for a catastrophic loss.

As examples, 5- and 30-day windows are selected. The threshold for a 5-day window is 77 mm/ 5 days (sixth-largest rainfall event), while it is 347 mm/ 30 days (sixth-largest rainfall event) for a 30-day window. The limitation of this approach is that our data are mean precipitation amounts that fall in the entire country. Therefore, the differences in precipitation amounts in each return level seem relatively small.

	2-year level	20-year level	100-year level
5-day window	84 (mm/ 5 days)	91 (mm/ 5 days)	94 (mm / 5 days)
30-day window	391 (mm/30 days)	414 (mm/30 days)	(m/30 days)

Table 3.2. Return Levels and Corresponding Rainfall Amounts.

3.3 Bangladesh

Data

Flood Area Data

Flood data are obtained first from Dartmouth Flood Observatory, whose data are available from 1985 until the present. However, since DFO data have some problems in data quality, two other annual data sources, Bangladesh Water Development Board and calculated seasonal surface water by our remote sensing team, are also examined. Three sets of data are plotted below (**Error! Reference source not found.3.1**). There are still big discrepancies among these three data sets. Looking at the correlation coefficients based on rankings (Kendall’s tau), DFO and BWBD have a higher coefficient (

Table 3.3). Since BWBD is government official data and referred to by other studies, such as Yang et al. (2015) and an ADB’s technical report (Ozaki, 2016), BWBD data for the purpose of demonstrating a methodology is used as a primary alternative to the DFO data.

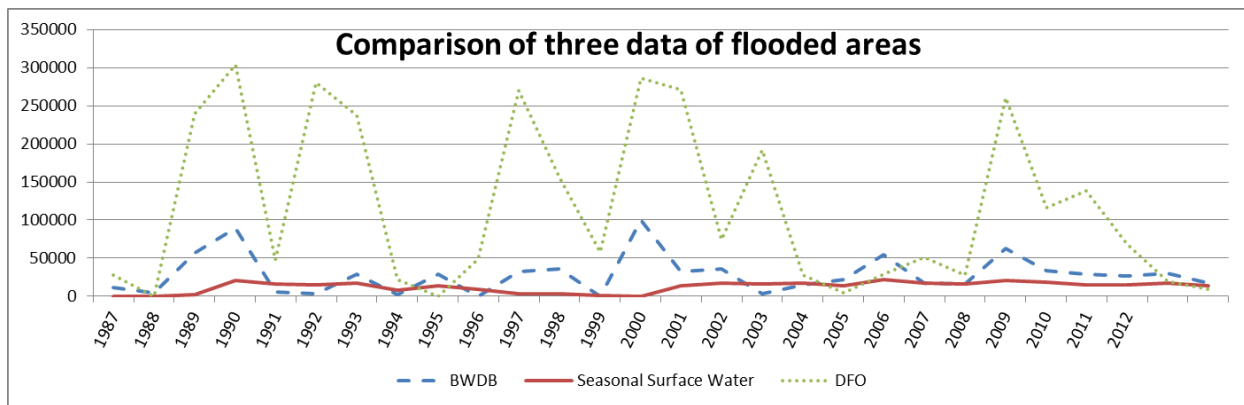


Figure 3.1. Comparison of the three data sets for flood-affected areas

	Seasonal Surface Water	DFO
BWBD	0.226	0.336
Seasonal Surface Water	-	0.079

Table 3.3. Correlation Coefficients among the Three Data Sets for Flood-Affected Areas (Kendall's tau)

Precipitation Data

The observational precipitation data are CPC 0.5 average over the whole country from January 1, 1979, to January 9, 2017. Reanalysis data are from 1900 to the present (ERA/ ERA-interim, ECMWF). In terms of windows, I have created three versions: windows based on a day when flood began, ones based on a day when flood ended, and ones based on the middle of a flood period.

Streamflow Data

The following streamflow data are obtained from Professor Soojun Kim, Inha University in South Korea and Professors Ethan Yang and Casey Brown at University of Massachusetts (Table 3.4). Each location is shown in

Figure 3.2 and

Figure 3.3. Data are available for the three main river systems that flow into Bangladesh: Bahadurabad for the Brahmaputra River, Rajshahi for the Ganges River, and Bazar Meghna for the Meghna River.

Site / Streamflow	Type
Bahadurabad in Bangladesh	<ul style="list-style-type: none"> ▪ Daily streamflow (1985–1992)¹ ▪ Monthly streamflow (1956–2000) no missing values¹ <ul style="list-style-type: none"> ▪ Daily (1998–Nov. 20, 2006)² ▪ Monthly (1956–1995, with NA)³ ▪ Monthly (1969–March 92, with NA)⁴ ▪ Monthly (1998–Nov. 2011 with NA)²
Beki River	<p>Daily streamflow (1994, 1995, 1997–2011)¹</p> <p>(But about 40% are missing values: Jan 96–July 97, Jan. 2001–Dec. 2002, May 2003–Sept. 2003, Jan. 2005–Dec. 2005, May 2008–Sept. 2008, Jan. 2010–Dec. 2010, July 2011–Dec. 2011)</p>
Brahmaputra site in Pandu, Guwahati	<ol style="list-style-type: none"> 1. Daily discharge (1992–April 1998) missing values are 4%.¹ 2. Once a week to get data for discharge (May 1998–Dec. 2009)¹ 3. Level discharge (2010–2013)¹
Water level	
Daily water level at Bahadurabad (m²)	Daily water level (April 1949–Oct. 2009) missing values are 4%.
Daily water level at Rajshashi (m²)	Daily water level (April 1922–Dec 2006), ² missing values are 25% (mostly April 1938–Dec. 1957).
Daily water level at Bhairab Bazar (m²)	Daily water level (April 1959–July 2006), missing values are 17%.
<p>1 is from Prof. Soojun Kim; 2 is from Prof. Ethan Yang; 3 is from Yu et al. through Yang; 4 is from GRDC 2013, through Yang).</p>	

Table 3.4. List of Streamflow and Water Level Data

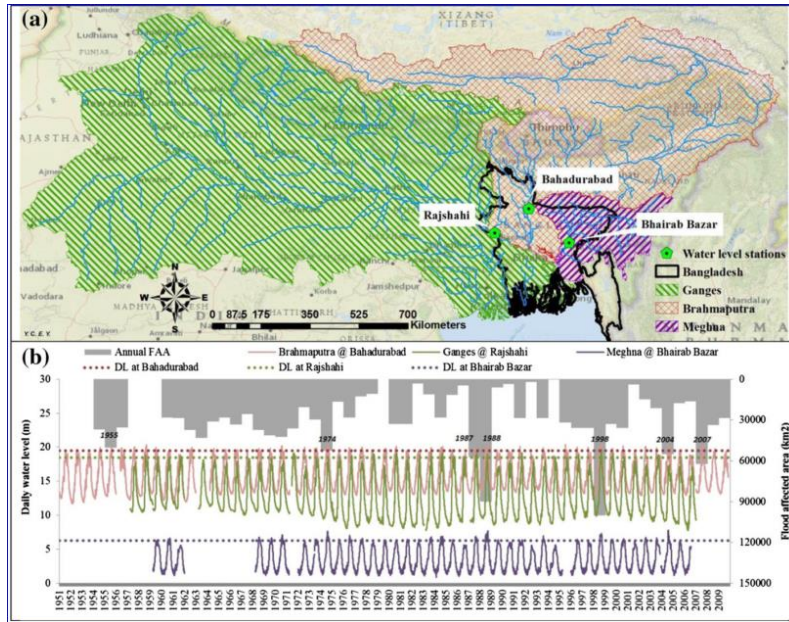


Figure 3.2. Locations of streamflow and water level data collected

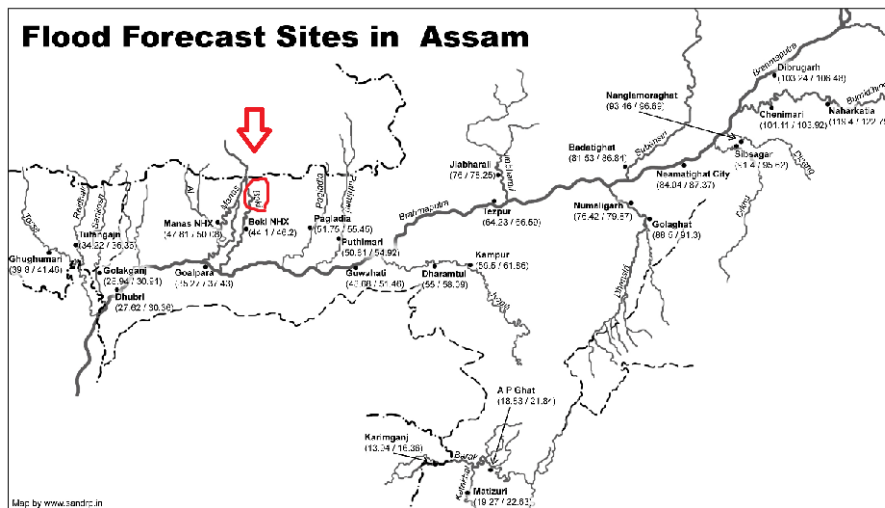


Figure 3.3. Location of Beki River

Analysis

Correlation with Precipitation Amounts, Streamflow, and Water Level

BDWB Data

The BDWB data are on an annual basis. Therefore, for the precipitation data, the annual maximum rainfall amounts for each window are calculated to estimate correlations with the annual flood-affected areas (

Table3.5 and Figure3.4). Rainfall data used here is reanalysis ERA and ERA interim data. Using these time windows and a threshold of the 10th percentile, the return periods are estimated based on generalized Pareto distributions (Table3.6). The shaper parameters (0.20 for one-day and 0.38 for two-day window) are improved to be positive compared to our previous analysis using DFO.

	1-day window	2-day window	3-day window	4-day window	5-day window	10-day window	30-day window
Annual flood-affected areas	0.107	0.115	0.057	0.050	0.042	0.053	0.039

Table 3.5 Correlation Coefficients between Flooded Areas and Different Rainfall Windows (Kendall's tau)

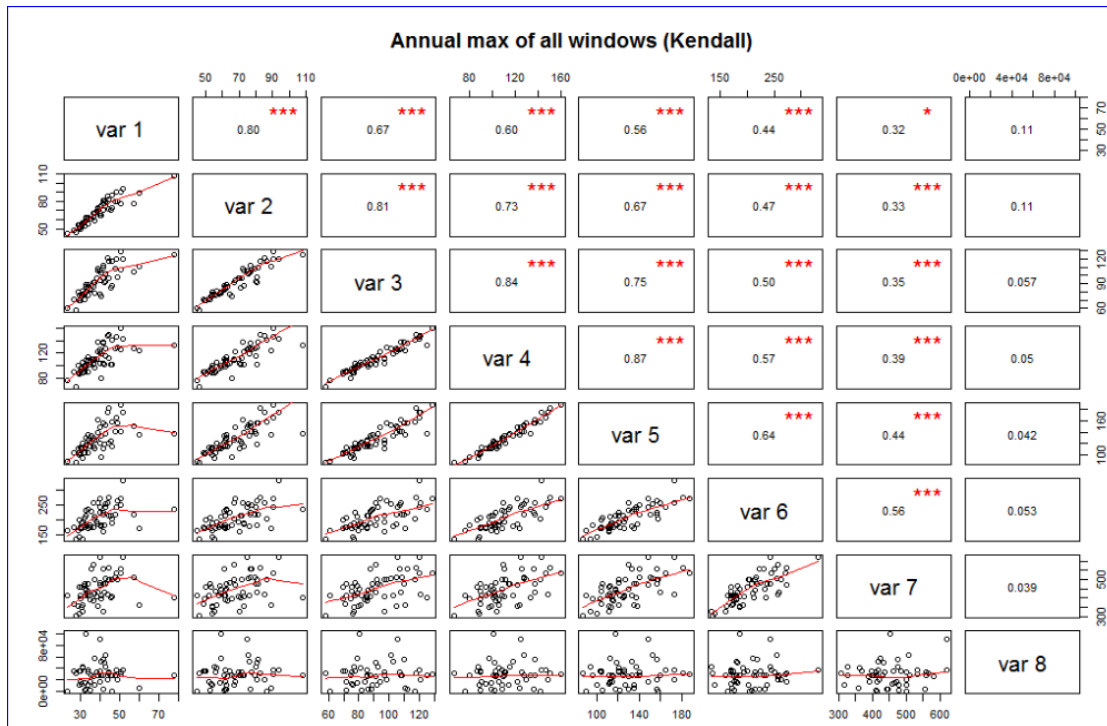


Figure 3.4. Correlation coefficients between flooded areas and different rainfall windows (Kendall's tau). Var1 is a 1-day window, Var2 is a 2-day window, Var3 is a 3-day window, Var4 is a 4-day window, Var5 is a 5-day window, Var6 is a 10-day window, Var7 is a 30-day window, and Var8 is flood-affected areas.

	2-year level	20-year level	100-year level
1-day window (mm/1 day)	97.5	145.7	194.9
2-day window (mm/2 days)	128.6	199.6	301.5

Table 3.6: Estimated Return Periods

However, the individual events are not associated well. For example, Figure3.5 shows the time series of highest time windows (one-day and two-day windows) and flood-affected areas. The largest flood event in 1998 is not associated with rainfall amounts. In addition, Figure3.4 shows that large rainfall events are not associated with flood occurrences in 1-day and 2-day windows while they are associated with flood occurrences in longer windows such as 30-day windows. One of possible reasons for this is that river inflow from India might be affecting the flood situations. Thus, I will examine the following two relationships:

- Streamflow data and rainfall amounts
- Streamflow data and flood-affected areas

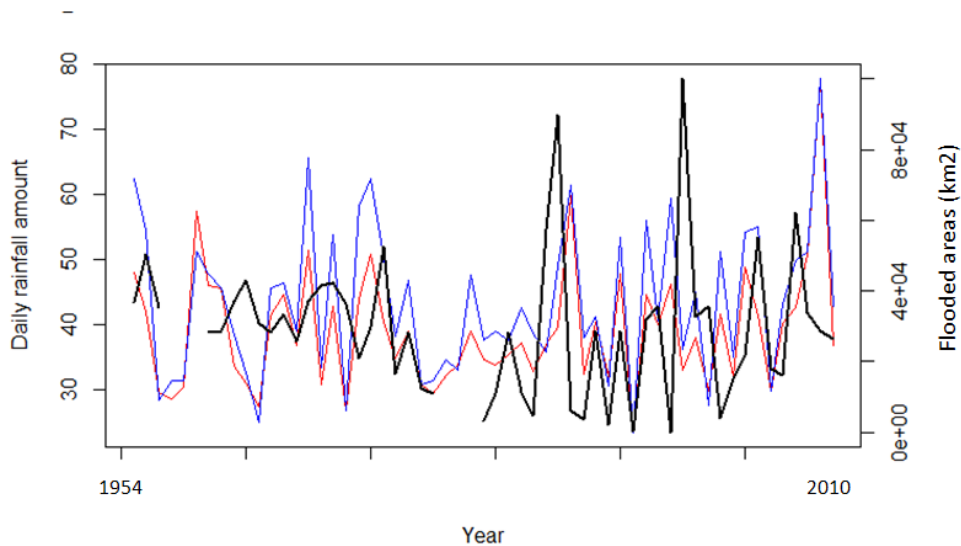


Figure 3.5. Time series of one-day and two-day windows and flood-affected areas. The blue line is daily rainfall amounts of one-day window, the red line is daily rainfall amounts of two-day window, and the black line is flood-affected areas (km²).

Examining Relationships among Seasonal Total Rainfalls, Streamflow/Water Levels, and Flooded Areas of BWDB

Using annual flood data from BWDB, the below analysis for three parts is conducted:

1. Relationships among annual maximum of different rainfall windows, streamflow, and flood areas.
2. Relationships among seasonal total rainfalls, maximum wet spell, streamflow, and flood areas. Seasons are defined thus: October–February as winter, March–May as summer, and June–September as rainy season. The sites are monthly streamflow (1956-2000) in Bahadurabad in Bangladesh, daily streamflow (1994, 1995, 1997–2011) in Beki River in India, and daily discharge (1992–2009) in Brahmaputra site in Pandu, Guwahati, India. Relationships among seasonal total rainfalls, maximum wet spell, water levels, and flood areas. Daily water level at Bahadurabad, daily water level at Rajshashi Ganges, daily water level at Bhairab, and Bazar Meghna are analyzed.

As the results attached in Annex 3A show, the correlation coefficients are low. Thus, the standardized anomaly is analyzed below.

Correlating Water Level Data with DFO's Flood Areas

Since the DFO data are on an event basis, the maximum values for water level at each location are used.

Figure 3.6 shows that there is an inability for variables of water level to predict for the lower range of flood-affected areas. But for top 10 events in terms of flood-affected areas, it seems that there are some relations. Therefore, for the next step, 10-20% events are examined below with a logistic regression.

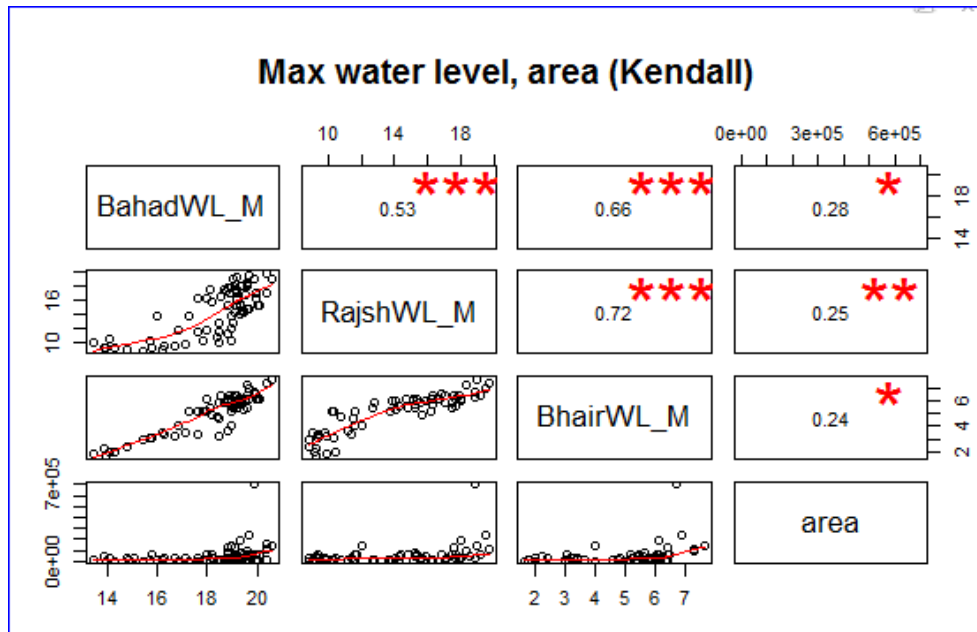


Figure 3.6. Correlation among the maximum water level in Bahadurabad, Rajshashi, and Bhairab Bazar, and flooded areas (Kendall's tau)

Standardized Anomaly

Using BWBD Data

Due to the low correlations coefficients shown above, the patterns of the standardized anomaly of the Bangladesh rainfall are identified. The standardized anomalies of the seasonal rainfall and standardized anomaly of the streamflow/ water level for each one are examined in the following way:

1. Rank by flooded areas from the largest to the smallest floods
2. Look at the patterns of each standardized anomaly of predictors

Some of the floods would be due to three factors: (1) heavy rainfall in Bangladesh itself; (2) the river of Brahmaputra without the influence of rainfall; and (3) the combination of them. If patterns are a small positive standardized anomaly for all of them for some of those events, it can be considered as the joint effects of these factors. In terms of a threshold of standardized anomaly, $p=-0.9$ ($z=-1.28$), the results of selecting predictors are as below (

Table 3.7.,

Table 3.8,

Table 3.9.,

Table 3.10).

Year	Total rainfall during a rainy season	Total summer rainfall	Total winter rainfall	Max precipitation during a rainy season	Total streamflow during a rainy season	Total streamflow during summer	Total streamflow during winter	Flooded areas (km ²)
1998	0.81	-0.28	0.68	-0.64	1.02	0.23	1.87	100250
1988	2.22	0.13	0.50	0.88	1.39	1.58	-0.39	89970
1987	1.00	0.15	-1.05	0.58	1.07	0.68	-0.96	57300
1974	0.37	2.02	-0.72	0.97	1.42	1.12	-1.43	52600
1963	-0.94	0.27	0.75	-0.59	-0.72	-0.69	0.06	43100
1970	1.88	-0.93	0.92	0.03	0.85	2.05	-1.05	42400
1969	-0.04	-1.31	-1.48	-0.1	0.81	0.40	-1.88	41400
1962	-0.31	-0.01	-0.86	0.21	-0.44	-0.54	0.28	37200
1968	0.79	-0.81	-0.87	2.20	-0.49	-0.57	0.18	37200
1971	0.73	-0.97	-0.05	-0.77	-4.22	-2.30	-3.34	36300

Table 3.7. Streamflow at Bahadurabad in Bangladesh (1956-2000) for the Top 10 Events in Terms of Flood-Affected Areas

Year	Total rainfall during a rainy season	Total summer rainfall	Total winter rainfall	Max precipitation during a rainy season	Total water level during a rainy season	Total water level during summer	Total water level during winter	Flooded areas (km ²)
1998	0.81	-0.28	0.68	-0.64	0.56	0.33	0.33	100250
1988	2.22	0.13	0.5	0.88	0.22	0.52	0.58	89970
2007	1.56	-1.38	0.57	1.09	0.19	0.22	0.28	62300
1987	1	0.15	-1.05	0.58	0.18	0.3	0.42	57300
2004	0.99	-0.38	0.15	1.08	0.11	0.5	0.39	55000
1974	0.37	2.02	-0.72	0.97	0.46	0.55	0.43	52600
1955	0.69	-0.72	0.95	-0.08	0.16	-0.04	-2.73	50500
1963	-0.94	0.27	0.75	-0.59	NA	NA	NA	43100
1970	1.88	-0.93	0.92	0.03	0.28	0.51	0.29	42400
1969	-0.04	-1.31	-1.48	-0.1	0.2	0.27	0.1	41400

Table 3.8. Water Levels at Bahadurabad (1954-2009) for the Top 10 Events in Terms of Flood-Affected Areas

Year	Total rainfall during a rainy season	Total summer rainfall	Total winter rainfall	Max precipitation during a rainy season	Total water level during a rainy season	Total water level during summer	Total water level during winter	Flooded areas (km ²)
1998	0.81	-0.28	0.68	-0.64	1.77	0.83	0.75	100250
1988	2.22	0.13	0.5	0.88	0.29	0.29	-0.08	89970
1987	1	0.15	-1.05	0.58	-0.17	-0.07	0.13	57300
2004	0.99	-0.38	0.15	1.08	-0.57	0.04	-0.04	55000
1974	0.37	2.02	-0.72	0.97	-0.34	0.27	0.19	52600
1963	-0.94	0.27	0.75	-0.59	0.9	-1.36	0.84	43100
1970	1.88	-0.93	0.92	0.03	0.4	0.96	0.49	42400
1969	-0.04	-1.31	-1.48	-0.1	0.18	0.75	0.58	41400
1962	-0.31	-0.01	-0.86	0.21	NA	NA	NA	37200
1968	0.79	-0.81	-0.87	2.2	-0.07	0.97	0.61	37200

Table 3.9. Water Levels at Rajshashi (1960 - 2006) for the Top 10 Events in Terms of Flood-Affected Areas

Year	Total rainfall during a rainy season	Total summer rainfall	Total winter rainfall	Max precipitation during a rainy season	Total water level during a rainy season	Total water level during summer	Total water level during winter	Flooded areas (km ²)
1998	0.81	-0.28	0.68	-0.64	1.1	0.05	0.32	100250
1988	2.22	0.13	0.5	0.88	1.63	0.45	0.56	89970
1987	1	0.15	-1.05	0.58	0.22	-0.21	0.57	57300
2004	0.99	-0.38	0.15	1.08	0.78	1.36	0.47	55000
1974	0.37	2.02	-0.72	0.97	1.16	0.61	0.75	52600
1963	-0.94	0.27	0.75	-0.59	NA	NA	NA	43100
1970	1.88	-0.93	0.92	0.03	0.58	0.37	1.08	42400
1969	-0.04	-1.31	-1.48	-0.1	0.2	-0.43	-0.09	41400
1962	-0.31	-0.01	-0.86	0.21	NA	NA	NA	37200
1968	0.79	-0.81	-0.87	2.2	0.4	-1.34	0.15	37200

Table 3.10. Water Levels at Bhairab Bazar for the Top 10 Events in Terms of Flood-Affected Areas

Using DFO Data

Error! Reference source not found. shows the standardized anomaly for top 10 events (the total number of recorded events in DFO is 78 events). Almost all show a positive standardized anomaly, except for one year for Rajshashi and Bhairab Bazar, but some are lower than 1.0. The year of 53 is negative standardized anomaly, which indicates that there is a possibility that local rain might be affecting it. Therefore, rainfall variables are included to conduct local regression analysis in the next section.

Flood ID	Maximum water level in Bahadurabad	Maximum water level in Rajshashi	Maximum water level in Bhairab Bazar	Flooded areas (km ²)
65	0.92	1.12	1.06	702361.7
5	0.81	1.23	1.15	240065.2
60	0.6	0.96	0.82	190726.7
73	1.21	NA	NA	150900
8	1.33	1.15	1.46	138855.3
53	0.42	-0.2	-0.06	136001.8
cc56	0.49	0.87	0.95	132357.5
78	0.33	NA	NA	128962.3
54	1.2	1.28	1.32	116176.7
7	0.76	1.02	1.3	98283.8

Table 3.11. Standardized Anomaly in Water Levels in Bahadurabad, Rajshashi, and Bhairab Bazar for the Top 10 Flood Events in Terms of Flood Areas

Local Regression

I conducted the local regressions for both BWBD and DFO data.

Conducting Local Regressions of BWBD Data

The results for BWBD's data are shown in Table 3.12. Results of Local Regressions with GCVs for BWBD Data 3.12. The predictor of the maximum water level in the rainy season in Bhairab Bazar shows the minimum GCV value (Generalized Cross-Validation), which means that it is an adequate model.

Predictors	GCV
Rain_Precip	4.804642e+08
RainWL	2.922312e+08
SummerWL	7.013183e+08
WinterWL	6.248909e+08
Rain_Precip, RainWL	4.586061e+08
Rain_Precip, SummerWL	6.692367e+08
Rain_Precip, WinterWL	9.628752e+08
RainWL, SummerWL	3.676626e+08
SummerWL, WinterWL	1.048159e+09
Rain_Precip, RainWL, SummerWL	8.903630e+08
Rain_Precip, RainWL, winterWL	2.940885e+09
Rain_Precip, SummerWL, WinterWL	1.870981e+09
RainWL, SummerWL, WinterWL	1.091053e+09
All	1.678401e+10

Table 3.12. Results of Local Regressions with GCVs for BWBD Data

DFO Data

Table 3.13. shows the GCV values for local regressions using DFO data. The maximum water level in Rajshashi Bhairab Bazar data shows the lowest GCV.

Predictors	GCV scores
Bahadurabad	7.347994e+09
Rajshashi	6.823339e+09
Bhairab Bazar	6.613980e+09
Bahadurabad + Rajshashi	7.522490e+09
Bahadurabad+Bhairab Bazar	9.506928e+09
Rajshashi+ Bhairab Bazar	8.536507e+09
all	1.227129e+10

Table 3.13. Results of Local Regressions with GCVs for the DFO Data

Logistic Regression with LASSO

After recording an event, if flood events exceed a threshold of top 10 and 20 percentiles, logistic regression is conducted with Lasso (Least Absolute Shrinkage and Selection Operator). Lasso is a shrinkage method that adds a penalty term.

Using BWBD Data (Annual, 1960–2006)

Figure 3.7 and Figure 3.7. The path of the coefficient of the logistic regression against the L1-norm of the whole coefficient vector as λ varies when the upper 10th percentile of BWBD's data was recorded as floods. The axis above is the number of nonzero coefficients at the current λ . Labels—1: The maximum precipitation with 1-day window; 2: the maximum precipitation with 2-day window; 3: the maximum precipitation with 3-day window; 4: the maximum precipitation with 4-day window; 5: the maximum precipitation with 5-day window; 6: the maximum precipitation with 10-day window; 7: the maximum precipitation with 30-day window; 8: the maximum water level in Bahadurabad (Brahmaputra River); 9: the maximum water level in Rajshashi (Ganges River); 10: the maximum water level in Bhairab Bazar

(Meghna River). 3.8 show the relationship between L1 norm and coefficients of the logistic regression for the case where the top 10 percentile events are recorded as floods and the case of the top 20 percentile, respectively.

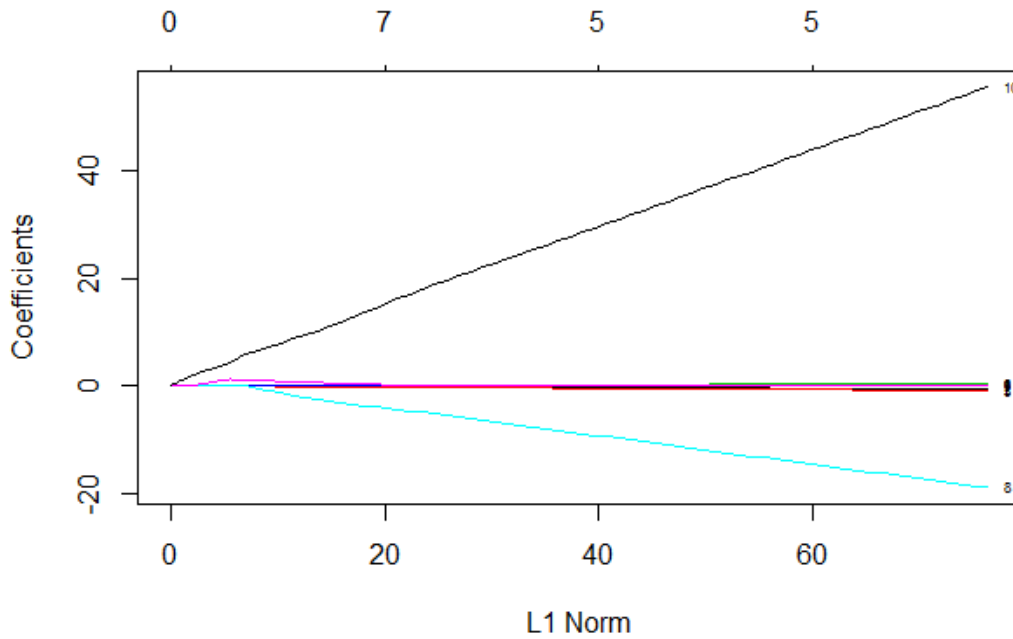


Figure 3.7. The path of the coefficient of the logistic regression against the L1-norm of the whole coefficient vector as λ varies when the upper 10th percentile of BWBD's data was recorded as floods. The axis above is the number of nonzero coefficients at the current λ . Labels—1: The maximum precipitation with 1-day window; 2: the maximum precipitation with 2-day window; 3: the maximum precipitation with 3-day window; 4: the maximum precipitation with 4-day window; 5: the maximum precipitation with 5-day window; 6: the maximum precipitation with 10-day window; 7: the maximum precipitation with 30-day window; 8: the maximum water level in Bahadurabad (Brahmaputra River); 9: the maximum water level in Rajshahi (Ganges River); 10: the maximum water level in Bhairab Bazar (Meghna River).

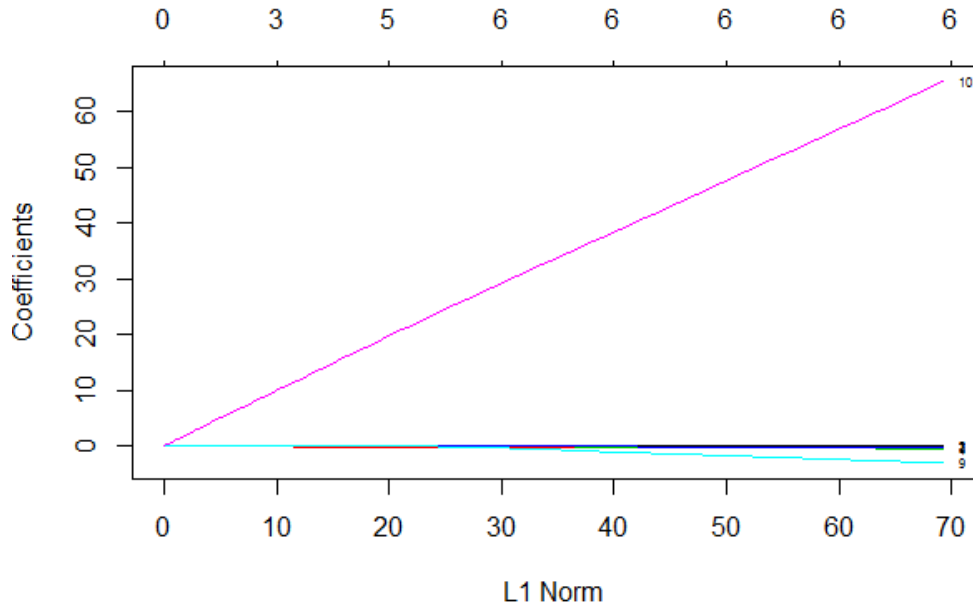


Figure 3.8. The path of the coefficient of the logistic regression against the L1-norm of the whole coefficient vector as λ varies when the upper 20th percentile of BWBD’s data was recorded as floods. The axis above is the number of nonzero coefficients at the current λ . Labels—1: The maximum precipitation with 1-day window; 2: the maximum precipitation with 2-day window; 3: the maximum precipitation with 3-day window; 4: the maximum precipitation with 4-day window; 5: the maximum precipitation with 5-day window; 6: the maximum precipitation with 10-day window; 7: the maximum precipitation with 30-day window; 8: the maximum water level in Bahadurabad (Brahmaputra River); 9: the maximum water level in Rajshashi (Ganges River); 10: the maximum water level in Bhairab Bazar (Meghna River).

Because BWBD’s data that uses top 10 percentile as a threshold has missing values, statistical (percentiles) data was used as an input instead. To select variables, the coefficients of cross-validation (with 10-fold) based on the minimum λ are examined as above.

Table3.14, as other tables below, shows validated variables based on cross-validation. For the data that records top 10 percentile as floods, the selected model uses the maximum water level in Rajshashi (Ganges River) and Bhairab Bazar (Meghna River).

Variables	Coefficients
Intercept	-35.74
Maximum rainfall amounts in 1-day window	-
Maximum rainfall amounts in 2-day window	-
Maximum rainfall amounts in 3-day window	-
Maximum rainfall amounts in 4-day window	-
Maximum rainfall amounts in 5-day window	-
Maximum rainfall amounts in 10-day window	-
Maximum rainfall amounts in 30-day window	-
Max Bahadurabad Water Level	-
Max Water Level in Rajshashi	0.65
Max water level in Bhairab Bazar	3.14

Table 3.14. Variable Selection Based on 10-fold Cross-Validation for the Upper 10th Percentile in BWBD's Data Recorded as Floods

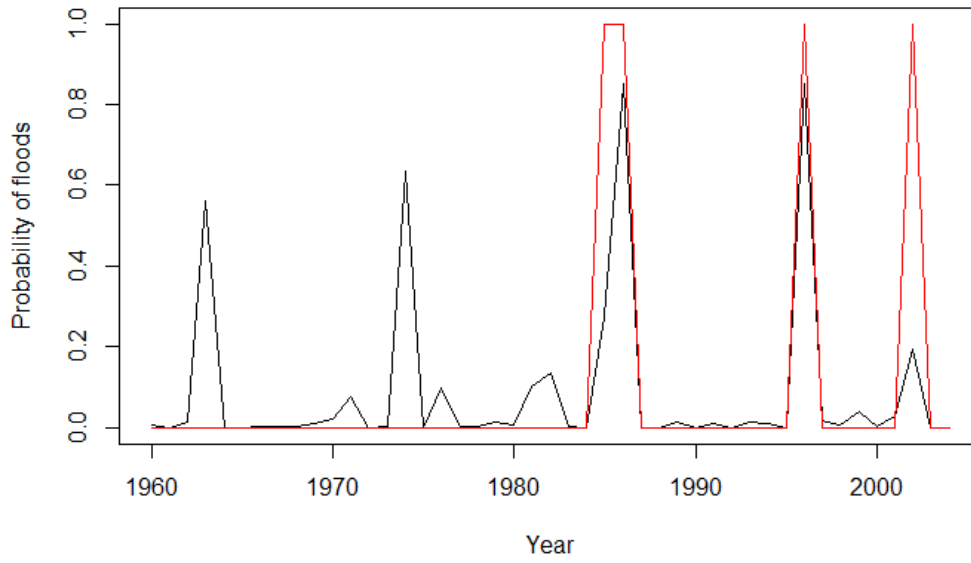


Figure shows the time series of the observations (red) and predicted values (black) by the best model.

Figure 3.9. Time series of observations and predicted values by the selected model for the top 10 percentile data

The selected model for the data of the top 20 percentile as a threshold uses the maximum rainfall amounts for 3-day and 30-day windows, and the maximum water level in Bhairab Bazar (Meghna).

Variables	Coefficients
Intercept	-42.42
Maximum rainfall amounts in 1-day window	-
Maximum rainfall amounts in 2-day window	-
Maximum rainfall amounts in 3-day window	-0.03
Maximum rainfall amounts in 4-day window	-
Maximum rainfall amounts in 5-day window	-
Maximum rainfall amounts in 10-day window	-
Maximum rainfall amounts in 30-day window	-0.01
Max Bahadurabad Water Level	-
Max Water Level in Rajshashi	-
Max water level in Bhairab Bazar	7.03

Table 3.15. Variable Selection Based on 10-fold Cross-Validation for the Upper 20th Percentile in BWBD's Data Recorded as Floods

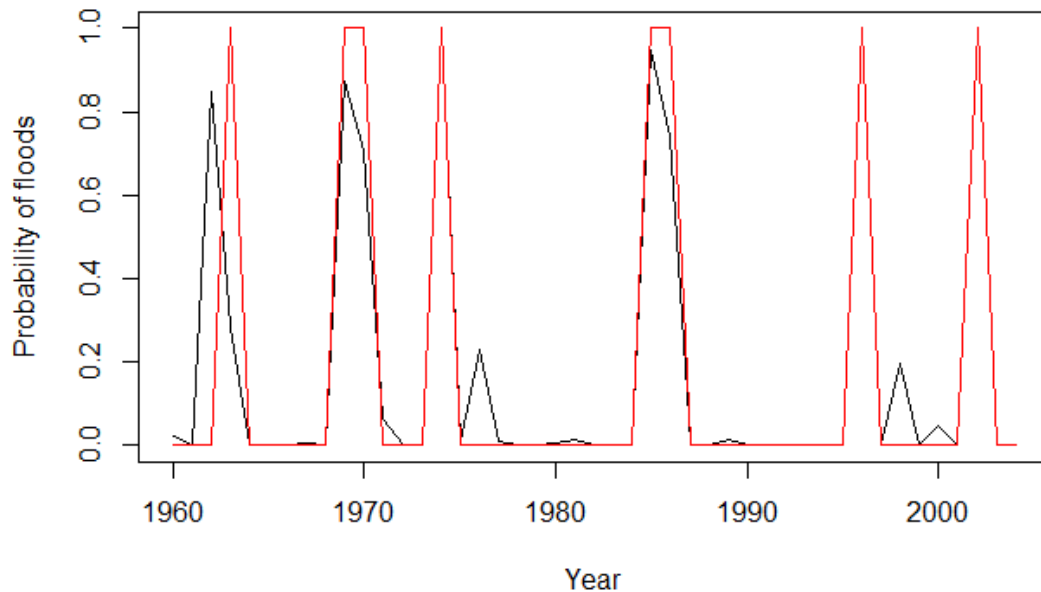


Figure 3.10 shows the time series of the observations (red) and predicted values (black) by the best model.

Figure 3.10. Time series of observations and predicted values by the selected model for the top 20 percentile data

Using the DFO Data (1985–2010)

Figure 3.11 and

Figure 3.11. The path of the coefficient of the logistic regression against the L1-norm of the whole coefficient vector as λ varies when the upper 10th percentile of DFO data was recorded as floods. Black: The maximum water level in Bahadurabad (the Brahmaputra River); red: the maximum water level in Rajshahi (Ganges River); green: the maximum water level in Bhairab Bazar (Meghna River); blue: the maximum precipitation with two-day window; and light blue: the maximum precipitation with two-day window.

3.12 show the relationship between L1 norm and coefficients of the logistic regression for the case where the top 10 percentile events are recorded as floods and the cause of the top 20 percentile, respectively. For the smaller L1 norm, the logistic regression is simpler because it uses fewer variables, while for the maximum L1 norm the result is equivalent to a normal multivariate logistic regression.

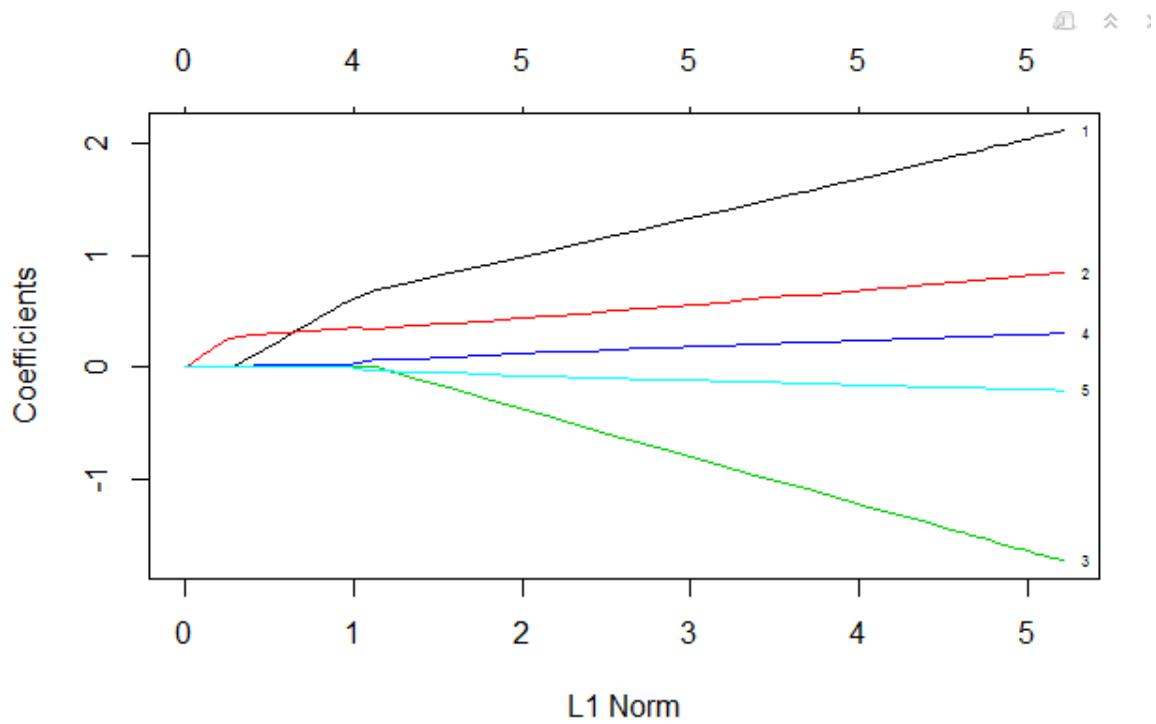


Figure 3.11. The path of the coefficient of the logistic regression against the L1-norm of the whole coefficient vector as λ varies when the upper 10th percentile of DFO data was recorded as floods. Black: The maximum water level in Bahadurabad (the Brahmaputra River); red: the maximum water level in Rajshashi (Ganges River); green: the maximum water level in Bhairab Bazar (Meghna River); blue: the maximum precipitation with two-day window; and light blue: the maximum precipitation with two-day window.

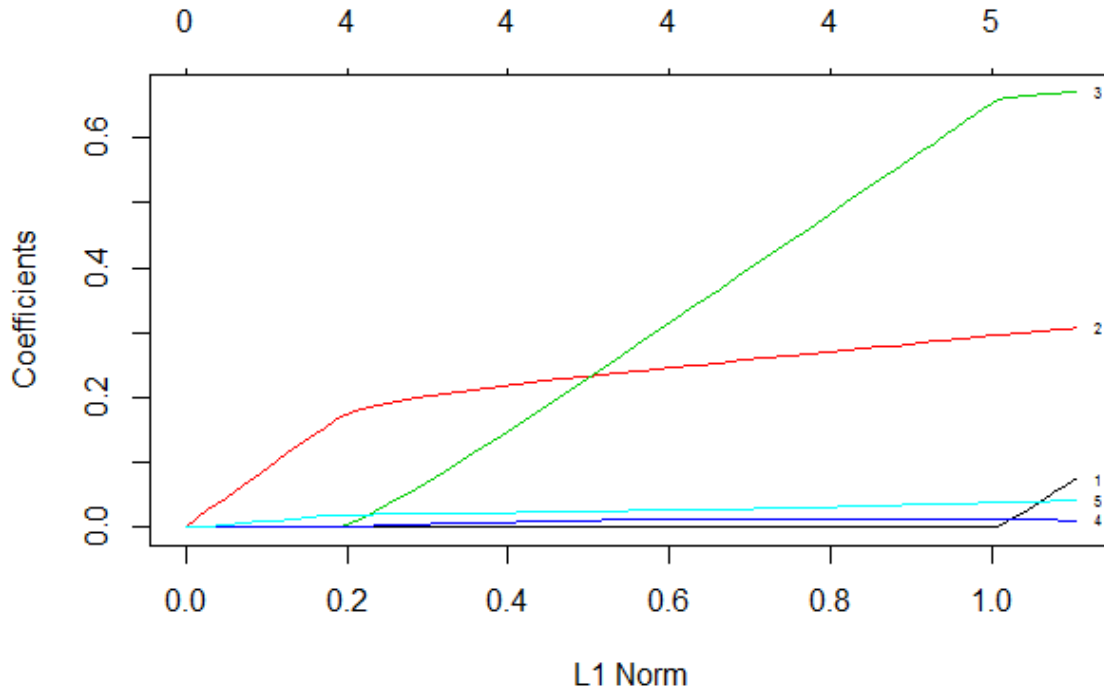


Figure 3.1.2 The path of the coefficient of the logistic regression against the L1-norm of the whole coefficient vector as λ varies when the upper 20th percentile of DFO data was recorded as floods. Black: the maximum water level in Bahadurabad (Brahmaputra River); red: the maximum water level in Rajshashi (Ganges River); green: the maximum water level in Bhairab Bazar (Meghna River); blue: the maximum precipitation with two-day window; and light blue: the maximum precipitation with two-day window.

To select variables, coefficients are examined by cross-validation (with 10-fold) based on the minimum λ . The selected model uses the maximum water level in Rajshashi (RajshWL_M) and two-day window (Table).

Variables	Coefficients
Intercept	-5.76
Max Bahadurabad water Level	-
Max water level in Rajshashi	0.22
Max water level in Bhairab Bazar	-
Maximum rainfall amounts in 2-day window	0.01
Maximum rainfall amounts in 3-day window	-

Table 3.16. Variable Selection Based on 10-fold Cross-Validation for the Upper 10th Percentile in DFO’s Data Recorded as Floods

Error! Reference source not found. 3.13 shows the time series of the observations (red) and predicted values (black) by the best model that uses RajshWL_M.

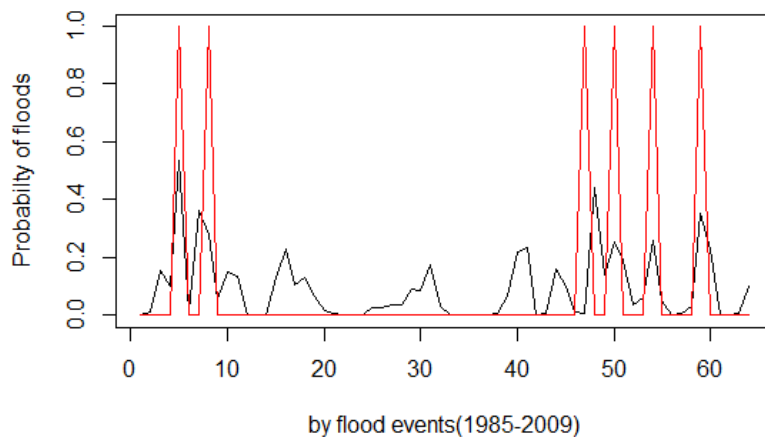


Figure 3.13. Time series of observations and predicted values by the selected model

For DFO data where the top 20% is recorded as flood above the threshold, the model that uses all variables except for BahadWLM is selected. The selected variables are the maximum rainfall amount in two-day and three-day windows, and the maximum water level in Rajshashi (Ganges) and Bhairab Bazar (Meghna).

Variables	Coefficients
Intercept	-7.85
Max Bahadurabad water level	-
Max Water level in Rajshashi	0.23
Max water level in Bhairab Bazar	0.20
Maximum rainfall amounts in 2-day window	0.01
Maximum rainfall amounts in 3-day window	0.02

Table 3.17. Variable Selection Based on 10-fold Cross-Validation for the Upper 20th Percentile in DFO’s Data Recorded as Floods

Error! Reference source not found.3.14 shows the time series of the observations (red) and predicted values (black) by the best model.

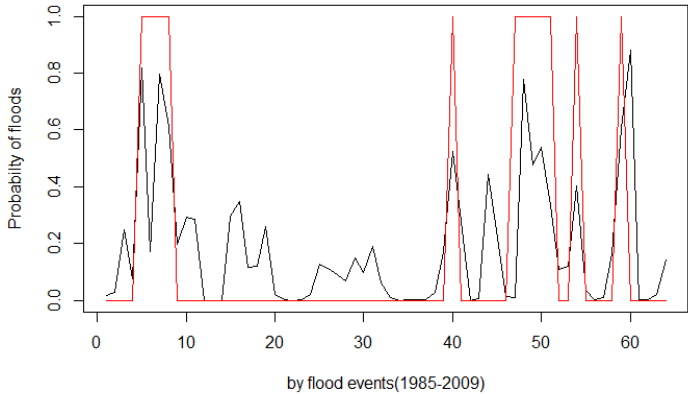


Figure 3.14. Time series of observations and predicted values by the selected model for the top 20 percentile data

3.4 Rainfall Analysis Summary

Recommendations

For Thailand, our recommendation is to use longer rainfall windows, such as a 30-day window, as a rainfall predictor. The estimated return levels are 13.0–16.8 mm/day, 13.8–18.2 mm/day, and 14.7–18.8 mm/day for the 2-year, 20-year, and 100-year levels, respectively. In addition, although this case analysis did not use streamflow data in Thailand, as presented in the case of Bangladesh, streamflow data on the ground might help predictions.

Flood mechanisms in Bangladesh seem complex because flows from India might be affecting flood hazards in Bangladesh. Therefore, rainfall recorded over Bangladesh is not effective for predicting floods in Bangladesh. Consequently, the river flows in the Brahmaputra and Ganges Rivers, which come from India, need to be considered. Our recommendation for predictors is to use data of the maximum water levels in Rajshahi (Ganges River) and Bhairab Bazar (Meghna River) in the combination of rainfall windows.

This case study shows that it is critical to collect data on flood hazards either on the ground or through remote sensing processing. In the case of Bangladesh, the DFO, BWBD data, and remote sensing estimates demonstrated inconsistency. On the other hand, streamflow data will help estimate specific predictors in each river basin.

Applicability of the Approach to Other Countries

This section characterizes the two countries and demonstrates which approaches will be applied to other countries, specifically Cambodia, Lao PDR, Myanmar, and Vietnam.

Thailand's topography is flat and rainfall is brought by monsoon and tropical storms. The main rivers (Chao Phraya River, and Mun and Chi Rivers, which are tributaries of Mekong River) originate in Thailand. In contrast, Bangladesh's topography is flat as well as low elevation. The three main rivers (Brahmaputra, Ganges, and Meghna Rivers) originate outside the country. Only 7.5% of the catchment area of the three main rivers lies in Bangladesh, and 92.5% lies outside the country. Precipitation in Bangladesh is brought by monsoon and tropical storms. The analysis of Bangladesh required using streamflow data in addition to precipitation analysis, and the results were not successful.

Cambodia is in the delta of the Mekong River, which is a major international river and originates in neighboring countries, such as China, Thailand, and Lao PDR. The Mekong River as well as Tonle Sap Lake cause floods in Cambodia. The country is very low elevation, and the elevation of one-third of the country is less than 25 meters above sea level. Rainfall is caused by tropical storms and monsoon. These characteristics meant that the analysis of floods in Cambodia is similar to the approach used for Bangladesh.

Lao PDR's main river system is the Mekong River. Streamflows in the river are influenced by rainfall and streamflow outside the country because the river originates in China. The majority of the country is mountainous, which makes the flood analysis challenging and which necessitates an approach similar to that for Bangladesh.

One of Myanmar's two major rivers, Ayeyarwardy River, originates in the country. In contrast, the other major river, Salween River, originates in China. The topography is a flat plain along the Ayeyarwardy River. The central flat plain is surrounded by mountains in the west and highlands in the east. Considering these characteristics, the approach to floods in Myanmar is not clear—that is, it is not clear whether either Thailand's approach or Bangladesh's approach is applicable.

Many rivers in Vietnam originate in neighboring countries. Topographies are different among the North (plain), Middle (mountainous), and South (delta) regions. Three-fourths of the country is mountains and highlands. These features make Vietnam take an approach similar to that of Bangladesh

Country	Bangladesh	Thailand	Cambodia	Lao PDR	Myanmar	Vietnam
River systems	<ul style="list-style-type: none"> - Three main river systems (Brahmaputra, Ganges, and Meghna Rivers) originate outside the country - Streamflows in the main rivers are influenced by rainfall and streamflow outside the country 	<ul style="list-style-type: none"> - The main rivers originate in the country 	<ul style="list-style-type: none"> - Mekong River flows through China, Thailand, and Lao PDR - Streamflows in the main rivers are influenced by rainfall and streamflow outside the country 	<ul style="list-style-type: none"> - Mekong River flows through China, Thailand - Streamflows in the main rivers are influenced by rainfall and streamflow outside the country 	<ul style="list-style-type: none"> - One of the two major rivers originates in the country and the other originates outside the country 	<ul style="list-style-type: none"> - Many rivers originate in other neighboring countries - Streamflows in the main rivers are influenced by rainfall and streamflow outside the country
Topography	<ul style="list-style-type: none"> - Flat topography in most of the land 	<ul style="list-style-type: none"> - Low and flat topography 	<ul style="list-style-type: none"> - Low topography (the elevation of one-third of the land is less than 25 meters above sea level) - Large lake (Tonle Sap Lake) 	<ul style="list-style-type: none"> - Mountainous 	<ul style="list-style-type: none"> - Flat plain along the Ayeyarwardy River - The flat plain is surrounded by mountains in the west and highlands in the east 	<ul style="list-style-type: none"> Topographies are different among the North (plain), Middle (mountainous), and South (delta) regions. - Three-fourth of the country is

						mountainous and highlands
Precipitation patterns	- Rainfall is brought by monsoon and tropical cyclones	- Rainfall is brought by monsoon and tropical cyclones	- Rainfall is brought by monsoon and tropical cyclones		Rainfall in the mountains and highlands causes floods in the plain.	
Approach	n/a	n/a	Bangladesh	Bangladesh	Bangladesh/ Thailand	Bangladesh

Table 3.18. Applicability of the Approaches to Other Countries in Southeast Asian Countries

3.5 Flood Index Development

A combination of different data sources is proposed to develop a scalable rapid prototyping strategy for a rain-induced regional flood risk index that can be readily customized to a particular setting. The intention is to provide an index that is suitable for a macro level regional project for index risk from an extreme event. As a second step, local organizations may disaggregate the risk index to potential local exposure mapping and secondary insurance or financial products at the local level that are effectively backed up or reinsured by the regional index. As these disaggregated products are deployed by a local agent, the need for the magnitude of the index purchase or reinsurance would become directly evident. Thus a scalable strategy that links traditional insurance and structural flood risk instruments to a regional index, or directly uses a regional index for coverage, becomes available. The primary data sources we consider are:

1. Historical inundation data from a variety of remote sensing sources—Landsat, Sentinel, etc.—available through NASA and other satellite sources, the Google Earth Engine, and the Colorado Flood Observatory. Area inundated and duration of inundation, as well as estimates of damages or people affected, may be available.
2. Daily and sub-daily rainfall records at different resolutions, as available from a variety of global climate data products based on observations, climate re-analysis, sub-seasonal to seasonal climate forecasts, and climate scenarios for future conditions. The re-analysis data sets may be up to a century long, thus providing the opportunity to reduce the uncertainty associated with short inundation records.
3. Daily streamflow and water level data from major contributing river systems, if available.
4. Digital elevation model and/or a hydrologic drainage network (e.g., hydrosheds) associated with the region of interest.
5. Population density, demographics, and infrastructure asset maps

Satellite-based inundation data are available for approximately 30 years over much of the earth. For the regional flood index for the region of interest, let us consider that a data set of maximum area inundated per event A_i , $i=1\dots l$ is available, over the n years of record, such that we retain only those events for which the inundated area is greater than some nominal threshold A^* . The corresponding date t_i and duration d_i of each event can also be extracted from the data using procedures similar to those used by the Colorado Flood Observatory, or taken directly from their pre-processed data.

A possible algorithm for index development can then be developed as follows:

1. For each event i , from the hydrologic drainage network identify the contributing drainage area for the inundated section of the region. Let us label this as C_i .
2. From the rainfall data set, identify $P_{i\tau}$, $\tau=1\dots T$, where T is an index for selected durations of rainfall, e.g., 1 day, 3 days, 5 days, ...30 days, and $P_{i\tau}$ is the precipitation averaged over C_i over a duration τ preceding the date of inundation t_i . We are interested in identifying a duration of rainfall that is best correlated to the potential inundation. For large area inundation, the extreme rainfall over the contributing area over one day immediately preceding the flooding is likely to be important. However, prior rainfall amounts in the immediately preceding days or even the prior week may also be important, since they would have contributed to runoff as well as the saturation of soil moisture leading to enhanced runoff for the most recent day. Consequently, we may be interested in an index that covers one or more time scales for precipitation, and would like to identify this from the data.
3. From the streamflow and water level data, identify total or maximum streamflow, SF_{ir} and water level WL_{ir} for each river system r . In the case of Bangladesh, there are three main river systems to consider: the Brahmaputra, Ganges, and Meghna Rivers.
4. Typically, the $P_{i\tau}$ will be highly correlated for different values of τ , and hence using a classical stepwise regression approach to choose the best values of τ to predict A_i or d_i from an appropriate subset of the $P_{i\tau}$ may not be robust. From a pragmatic perspective we would like to design an index such that no more than two values of τ are considered for our final index. These would correspond to the immediate rainfall and rainfall over an appropriate pre-conditioning period. Then for the region of interest, we consider a selection problem for τ' , to maximize the conditional log likelihood:

$$LL = \sum_{i=1}^n \log(f(A_i | P_{i1}, P_{i\tau'})) \quad (1)$$

The conditional probability model would be fit using an appropriate multivariate model, e.g., a generalized Pareto distribution for A_i with an appropriate linear or nonlinear model relating A_i to P_{i1} and $P_{i\tau'}$. Let's term this Model M_1 .

5. When SF_{ir} and WL_{ir} are available, consider a selection problem for predictors for SF_{ir} and WL_{ir} , maximize the conditional log likelihood:

$$LL = \sum_{i=1}^n \log(f(A_i | P_{i1}, P_{i\tau'}, SF_{i,r}, WL_{i,r})). \quad (2)$$

6. An alternative approach is to consider only the extreme inundation events, which have a return period greater than R years. For this consider a binary time series, b_i , $i=1\dots I$ such that $b_i = 1$ if $A_i > A_R$, and 0 else. A_R can be estimated by any method appropriate for partial duration series (since there may be multiple events per year). Next we consider a generalized linear model with a logistic link function between b_i and P_{i1} , $P_{i\tau'}$, $SF_{i,r}$, and $WL_{i,r}$. Then one can use BIC or AIC or Lasso to select the appropriate τ' and the predictors. Note that only P_{i1} may be selected by either procedure suggested.

7. Let us say that the selected logistic model is

$$b = \beta_0 + \beta_1 P_{i1} + \beta_2 P_{i\tau'} + \beta_3 SF_{i,r} + \beta_4 WL_{i,r} \quad (3)$$

Then if a selected combination of $(P_{i1}, P_{i\tau'}, SF_{i,r}, WL_{i,r})$ leads to $b=0.5$, then this combination corresponds to a median prediction of A_R . This implies that our index can be defined as b , and a value of $b \geq 0.5$ constitutes an exceedance of A_R and reflects a payout condition.

8. If the above approach is only using $P_{i\tau}$, the approach is similar to Khalil et al. (2007). If flood hazard series such as flood-affected areas are directly used to specify a trigger threshold for index insurance, then the trigger level is the value of the series corresponding to a desired probability of exceedance, $Prob_{exc}$. For example, $Prob_{exc}=0.25$ corresponds to an event with a return period of four years. The corresponding threshold can be identified from an observed cumulative distribution function of historic series of flood-affected areas under the assumption that the flood-affected areas are independently and identically distributed.

a. Let's say that $Prob_{exc}$ is 0.25. A corresponding threshold A^* such that the flood area is exceeded on average with probability $Prob_{exc}$ can be identified. This is the 75th percentile of the flood area as given

$$E [\text{prob}(A > A^*)] = 0.25 \quad (4)$$

b. We can identify a trigger value $P_{i\tau}$ for two different precipitation windows, $P_{i\tau}$, such that on average the probability of exceedance of the corresponding threshold for the flood area is the desired probability $Prob_{exc}$. Given that we use the two highest correlated windows of precipitation, we expect that if a certain threshold A^* is exceeded, then on average a corresponding threshold $P_{i\tau}^*$ is exceeded with some probability. We aim to identify such an A^* .

c. For flood-affected areas A , consider the binomial process as b_i as before ($b_i=1$ if $A > A^*$, $b_i=0$ if $A < A^*$). Now, consider a logistic regression of b_i on $P_{i\tau}$, we can estimate the conditional probability $E[\text{prob}(b_i | P_{i\tau}^*)]$.

9. To measure performance of predictors (or classifiers) using a trigger level, a contingency table is created (**Error! Reference source not found.**). Each case is recorded with either positive (p) or negative class (n) labels. Namely, the positive class means a case when a flood event is recorded above a certain degree; the negative one means a case when a flood event is not recorded below the same threshold. A classifier is a prediction of a flood or not. To make a distinction between the actual class and the predicted class, we use the labels $\{Y, N\}$ for the class predictions by a model. Following a predicting classifier and actual instance, there are four possible outcomes (**Error! Reference source not found.3.16**). We will use true positive rate (sensitivity) and true false ratio (specificity).

		True class			
		p	n		
Hypothesized class	Y	True Positives	False Positives	fp rate = $\frac{FP}{N}$	tp rate = $\frac{TP}{P}$
	N	False Negatives	True Negatives	precision = $\frac{TP}{TP+FP}$	recall = $\frac{TP}{P}$
Column totals:		P	N	accuracy = $\frac{TP+TN}{P+N}$	F-measure = $\frac{2}{1/precision+1/recall}$

Figure 3.16. Contingency table and its performance metrics. Source: Captured from Fawcett (2006)

The above approach is based on event i . However, we can do the same approach for annual data. For the annual data, for a year t , we define $P_{t,season,r}$, $SF_{t,season,r}$, and $WL_{t,season,r}$.

Based on the above approaches, we are going to present our case studies using Thailand and Bangladesh data.

Results of Our Case Studies in Thailand and Bangladesh

Remote Sensing

Processed satellite images from Sentinel-1 and LANDSAT estimate flood extent, depth, and duration. Flood depth is approximated with a digital elevation model (DEM). One of the current challenges is cloud penetration. For example, cloud cover in July-August 2015 in Bangladesh does not allow the detection of flood. The detection of flood during July and August can only be performed using SAR data. However, Sentinel-1 and LANDSAT show a high agreement after removing clouds and permanent water. After these processes, the statistics to assess uncertainties are the following: 0.72 for the Kappa Coefficient, 0.78 for Probability of Detection (POD), and 0.69 for Critical Success Index (CSI).

Precipitation and Streamflow Analysis (Thailand)

Since there is no major river coming from the outside of the country, the association between rainfalls and flood extent was reasonably well established, so that a rainfall-based index can be created. Thus, we follow approach 8 above. The correlation analysis selected 5- and 30-day windows for τ . For the demonstration purpose, a selected $Prob_{exc}$ is 0.2, which is the 80th percentile of the flood areas index.

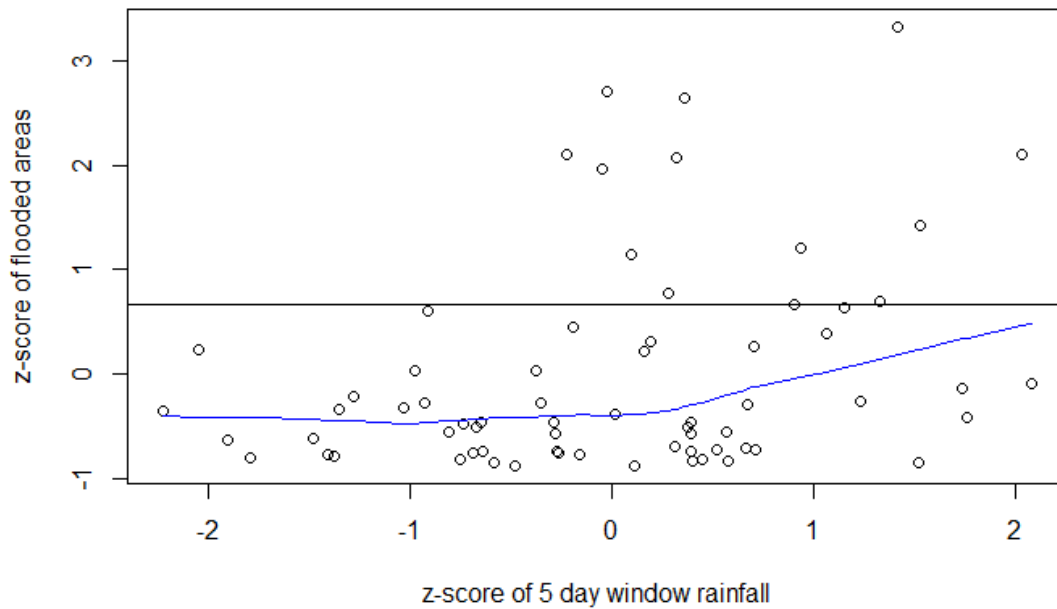


Figure 3.17. Flood areas versus five-day window rainfall using 68 events since 1985. The horizontal line represents the 80th percentiles of the distribution of the flood areas based on 68 events of data, and the blue curve represents a lowess smooth of the data.

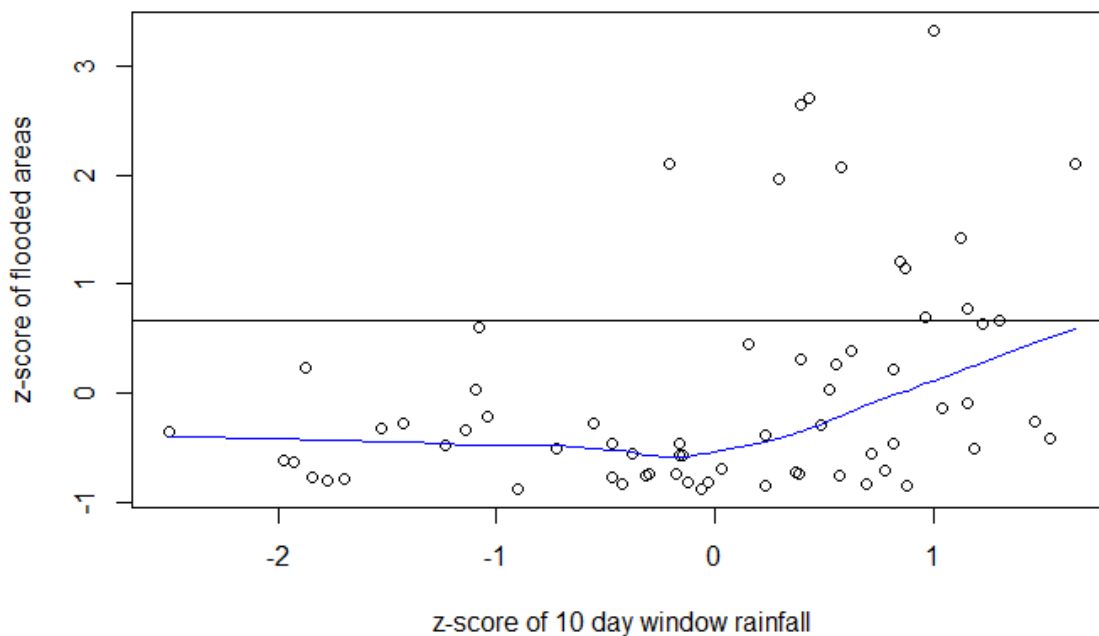


Figure 3.18. Flood areas versus 10-day window rainfall using 68 events since 1985. The horizontal line represents the 80th percentiles of the distribution of the flood areas based on 68 events of data, and the blue curve represents a lowess smooth of the data.

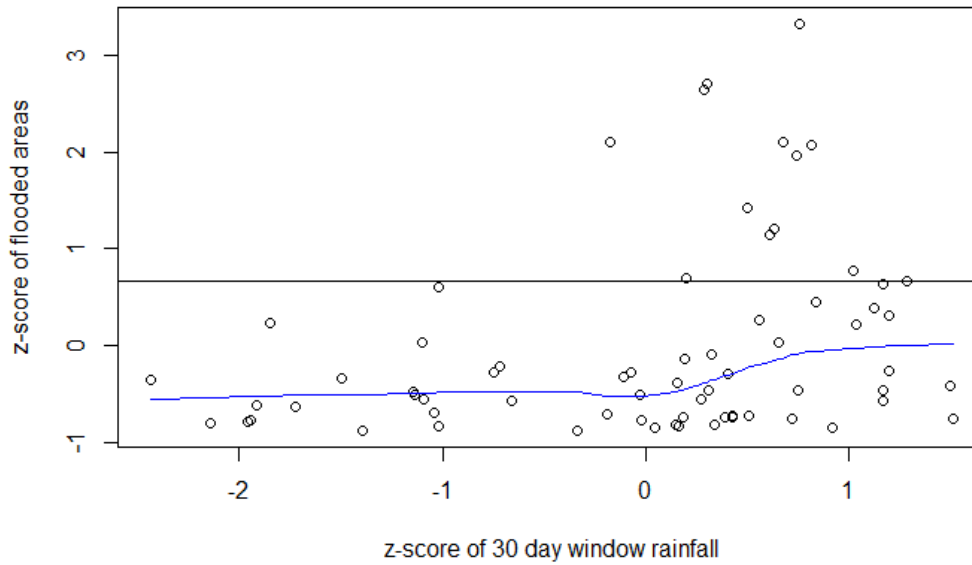


Figure 3.19. Flood areas versus 30-day window rainfall using 68 events since 1985. The horizontal line represents the 80th percentiles of the distribution of the flood areas based on 68 events of data, and the blue curve represents a lowess smooth of the data.

We consider the binomial process, b_i as $b_i=1$ if $A>A^*$, $b_i=0$ if $A<A^*$, where A^* is the 80th percentile of the flood area. Then, we estimated the following logistic regressions:

$$b = \beta_0 + \beta_1 P_{i5} \tag{5}$$

$$b = \beta_0 + \beta_1 P_{i10} \tag{6}$$

$$b = \beta_0 + \beta_1 P_{i30} \tag{7}$$

The results are in Table 3.19.

Coefficients	Estimate	Std. Error	z value	Pr(> z)
Model 5 (AIC=62.16)				
β_0	-1.7122	0.3802	-4.503	6.69e-06
β_1	0.9925	0.3829	2.592	0.00955
Model 6 (AIC=56.6)				
β_0	-2.1122	0.5183	-4.075	4.59e-05
β_1	1.6849	0.5902	2.855	0.00431
Model 7 (AIC=63.54)				
β_0	-1.7308	0.4001	-4.326	1.52e-05
β_1	1.0426	0.4750	2.195	0.0282

Table 3.19. Result of Single Logistic Regressions of b on 5-Day, 10-Day, or 30-Day Windows for the 80th Percentile Floods

For multivariate logistic regressions, we selected the models based on AIC. The lowest AIC is the Model (8) as below (**Error! Reference source not found.3.20**).

$$b = \beta_0 + \beta_1 P_{i3} + \beta_2 P_{i4} + \beta_3 P_{i10} \quad (8)$$

Coefficients	Estimate	Std. Error	z value	Pr(> z)
Model 8 (AIC=54.3)				
β_0	-2.2817	0.5628	-4.054	5.03e-05
β_1	-4.2091	1.9875	-2.118	0.0342
β_2	4.1505	2.2280	1.863	0.0625
β_3	1.6209	0.8228	1.970	0.0488

Table 3.20. Result of the Multivariate Logistic Regression of b on 3-Day, 4-Day, and 10-Day Windows of Model (8)

Using these estimated coefficients, the probability, *Prob*, that the flood area series will exceed the specified threshold, conditional on precipitation window values, P_{i10} and $P_{i3,i4,i10}$, is estimated as

$$\text{prob} = \frac{1}{1 + \exp(-z_i)} \quad (9)$$

For P_{i10} , we have the following estimated equations:

$$z_i = -2.1122 + 1.6849P_{i10} \quad (10)$$

$$z_i = -2.2817 + -4.2091P_{i3} + 4.1505 P_{i4} + 1.6209P_{i10} \quad (11)$$

Consequently, for the 80th percentile, the esimated trigger level P_{i10} of Model (6) is 1.25361 as a z-score for 10-day window rainfall (

Figure3.20). For Model (8), $P_{i10} + 2.59677P_{i3} - 2.56061P_{i4} \geq 1.40767$ is a trigger level (

Figure3.21). In

Figure3.21, the trigger level is above the colored plain.

As conclusions, for the case of Thailand, if either of these two models exceed these thresholds, they reflect payout conditions.

Following Step 9 above, to examine true positive and true negative ratio, sensitivity and specificity are calculated when using these trigger levels (

Table 3.21). Specificity values (0.96 and 0.92) work relatively well, while the sensitivity values (0.15 and 0.23) do not. Thus, it is necessary to improve the true positive ratio.

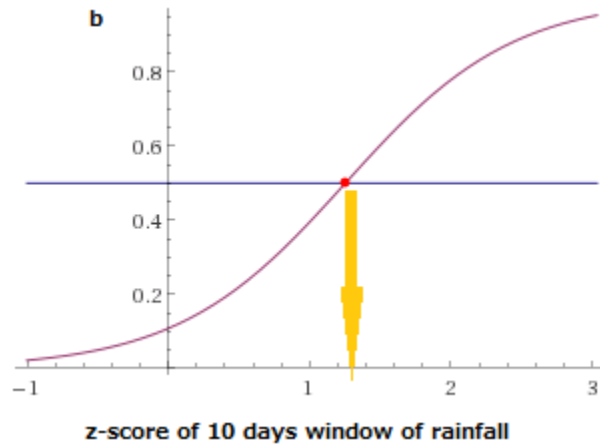


Figure 3.20. Predicted $E [Prob|Pi10]$ from the logistic regression. The trigger level is greater than 1.25361 of the z-score of 10-day window rainfall for the 80th percentile flood.

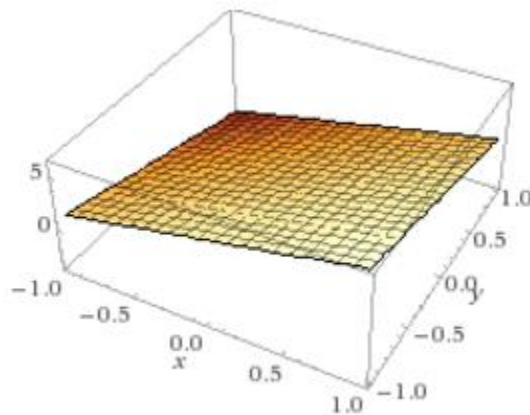


Figure 3.21. Predicted $E [Prob|Pi3, Pi4, Pi10]$ from the logistic regression. The trigger level is $Pi10 + 2.59677Pi3 - 2.56061Pi4 \cong 1.40767$ for the 80th percentile floods. The trigger level is above the colored plain.

	80th percentile flood occurred ($b_i=1$)	80th percentile flood did not occur ($b_i = 0$)
Model (6)		
$b \geq 0.5$ (payout)	2	2
$b < 0.5$ (no payout)	11	53
	Sensitivity = $2/13=0.15$	Specificity = $53/55=0.96$
Model (8)		
$b \geq 0.5$ (payout)	3	4
$b < 0.5$ (no payout)	10	51
	Sensitivity = $3/13=0.23$	Specificity = $51/55=0.92$

Table 3.21. Contingency Table for Model (6) and Model (8)

Precipitation and Streamflow Analysis (Bangladesh)

For the case of Bangladesh, we use annual data. Seasons are categorized into three seasons: summer (March–May), rainy season (June–September), and winter (October–February). Streamflow and water level data are collected from three stations from the three main river systems: Bahadurabad in Brahmaputra River (abbreviated *bahad*), Rajshashi in Ganges River (abbreviated *rajsh*), and Bhairab and Bazar in Meghna River (abbreviated *bhair*) are analyzed. Water levels in Bahadurabad during the rainy season for a year of t are shown as $WL_{t,rainy,bahad}$.

We follow approach 8 above. For the demonstration purpose, a selected $Prob_{exc}$ is 0.1, which is the 90th percentile of the flood area index.

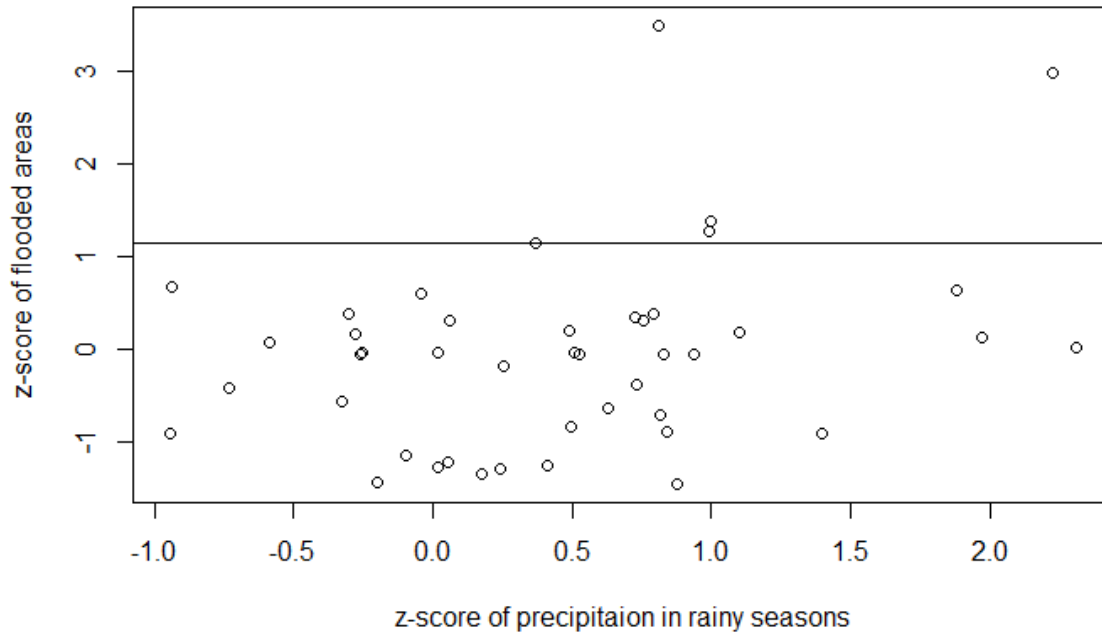


Figure 3.22. Flood areas versus precipitation during the rainy season using 46 years of events between 1960 and 2005. The horizontal line represents the 90th percentiles of the distribution of the flood areas based on 46 years of data.

We consider the binomial process, b_i as $b_i=1$ if $A>A^*$, $b_i=0$ if $A<A^*$, where A^* is the 90th percentile of the flood area. Then, we estimated the following logistic regression:

$$b = \beta_0 + \beta_1 P_{t,rainy} \quad (12)$$

Coefficients	Estimate	Std. Error	z value	Pr(> z)
β_0	-4.350	1.590	-2.735	0.00623
β_1	4.571	1.993	2.293	0.02184

Table 3.22. Result of the Logistic Regression of b on Water Level in Rainy Season in Bhairab and Bazar

For multivariate logistic regressions, we selected the models based on AIC. The lowest AIC is Model 13 and Model 14 as below.

$$b = \alpha_0 + \alpha_1 P_{t,rainy} + \alpha_2 WL_{t,rainy,bhair} \quad (13)$$

$$b = \alpha_0 + \alpha_1 WL_{t,rainy,bahad} + \alpha_2 WL_{t,rainy,bhair} \quad (14)$$

Coefficients	Estimate	Std. Error	z value	Pr(> z)
β_0	-5.339	2.281	-2.341	0.0192
β_1	1.054	1.169	0.902	0.3671
β_2	4.874	2.295	2.124	0.0337

Table 3.23. Result of the Logistic Regression of b on Water Level in Rainy Season in Bhairab and Bazar (Model 13)

Coefficients	Estimate	Std. Error	z value	Pr(> z)
β_0	-4.110	1.696	-2.423	0.0154
β_1	-2.952	5.396	-0.547	0.5844
β_2	5.456	2.771	1.969	0.0490

Table 3.24. Result of the Logistic Regression of b on Water Level in Rainy Season in Bahadurabad and Rainy Season in Bhairab and Bazar (Model 14)

Consequently, for 90th percentile, the estimated trigger level P_i rainy is 0.95 as a z-score for rainy season's precipitation (Table 3.25). For Models 14-16, the trigger levels are expressed in a linear function. For Model 13,

$$P_{rainy} + 0.216249 * WL_{t,rainy,bhair} \geq 1.0954 ,$$

while for Model 14,

$$WL_{t,rainy,bahad} - 0.541056 * WL_{t,rainy,bhair} \geq 0.753299 .$$

For these, the trigger levels are above the linear lines.

Model	Linear predictor of z_i	AIC	Trigger levels
Model 12	$z_i = -4.350 + 4.571 P_{t,rainy}$	19.96	0.95
Model 13	$z_i = -5.339 + 1.054 P_{t,rainy} + 4.874 WL_{t,rainy,bhair}$	21.10	$P_{t,rainy} + 0.216249 * WL_{t,rainy,bhair} \geq 1.0954$
Model 14	$z_i = -4.110 + -2.952 WL_{t,rainy,bahad} + 5.456 WL_{t,rainy,bhair}$	21.69	$WL_{t,rainy,bahad} - 0.541056 * WL_{t,rainy,bhair} \geq 0.753299$

Table 3.25. Trigger Levels for Models 12-14

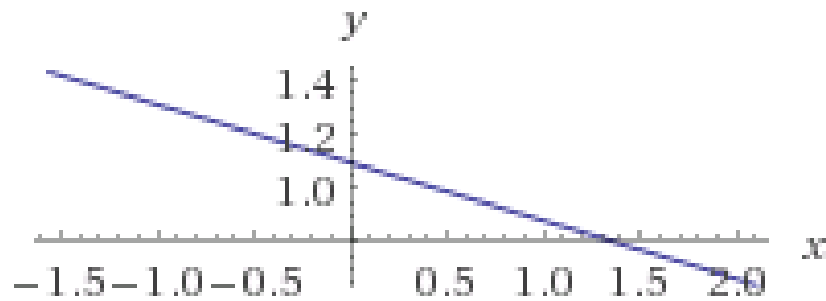


Figure 3.23. A trigger line of Model 13. The x axis is a z-score of water level in Bhairab and Bazar in rainy seasons, while the y axis is a z-score of rainfall in rainy seasons. The upper right part above the line is an area with the trigger level.

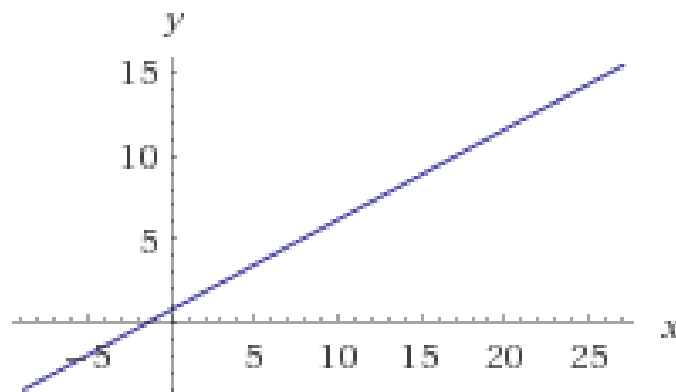


Figure 3.24. A trigger line of Model 14. The x axis is a z-score of water level in Bhairab and Bazar in rainy seasons, while the y axis is a z-score of water level in Bahadurabad in rainy seasons. The upper left part above the line is an area with the trigger level.

As conclusions, for the case of Bangladesh, if either of these models exceed these thresholds, they reflect payout conditions.

Following Step 9 above, in order to examine true positive and true negative ratio, sensitivity and specificity are calculated when using these trigger levels (

Table 3.26). Specificity values (1.0, 0.87, and 1.0) are working relatively well. In addition, the sensitivity values (0.6, 0.4, and 0.6) are also doing better than the case of Thailand. Thus, we can say that the addition of predictors of water levels improves models.

	10th percentile flood occurred ($b_i=1$)	10th percentile flood did not occur ($b_i = 0$)
Model 12		
$b \geq 0.95$ (pay out)	3	0
$b < 0.95$ (no payout)	2	33
	Sensitivity = $3/5 = 0.6$	Specificity = $33/33=1.0$
Model 13		
$b \geq 0.5$ (pay out)	2	4
$b < 0.5$ (no payout)	3	27
	Specificity = $2/5=0.4$	Sensitivity = $27/31=0.87$
Model 14		
$b \geq 0.5$ (pay out)	3	0
$b < 0.5$ (no payout)	2	31
	Specificity = $3/5=0.6$	Sensitivity = $31/31=1.0$

Table 3.26. Contingency Table for Models 12–14.

Conclusions and Recommendations

- 1) For the case of Thailand, the association between precipitation that falls in Thailand and flooded areas is relatively higher. However, the association between rainfall that falls in Bangladesh and flood extent was weak because there are large influences of the streamflow from India. Also, the association between streamflow and water levels with flood extent was not significant.
- 2) However, the performance of flood index in Bangladesh was better than Thailand in terms of sensitivity and specificity. This is partly because water level data contribute more to the performance of index.
- 3) As a recommendation for a future step, streamflow data are considerably helpful for Thailand, given that the situation has improved in Bangladesh if we include water level data.
- 4) As a recommendation for the study of Bangladesh, there are still large discrepancies in estimates of flooded areas among different sources. The direct use of flood estimates might be an option. However, the estimates of flood extent are not well estimated currently. Historic data of DFO and BWDB and results of processed remote sensing products are not consistent with each other (
- 5) Figure3.25).

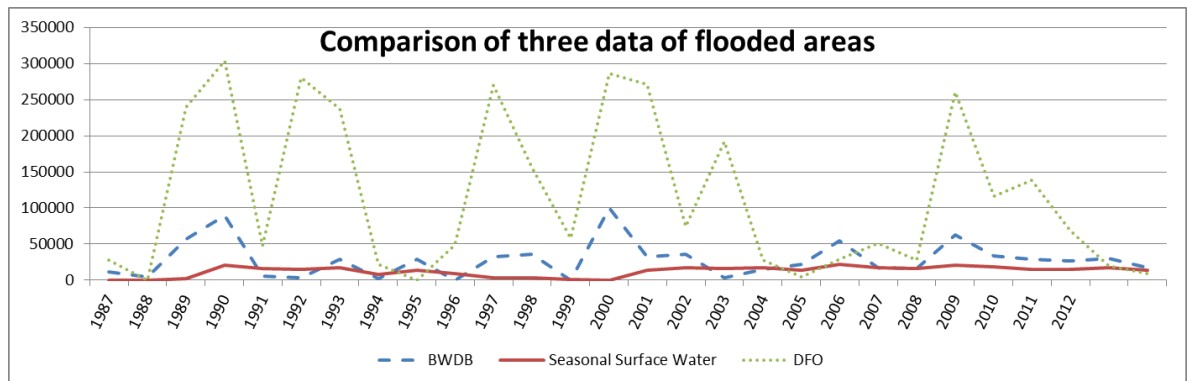


Figure 3.25. Comparison of three data sets of flooded areas in Bangladesh.

6) In future work, a perfect simultaneous acquisition of optical and SAR data over a flooded area would be desirable in order to improve the estimates of flooded areas. In addition, we will need to obtain higher-resolution data below the district level.

7) Also, for future work on the index approach, we can refine this procedure to account for longer-term variations as represented in the longer rainfall data sets of n_1 years that may be available.

This can be done as follows:

- Randomly draw with replacement n_1/n events from the set of I inundation events contained in the n years of data, and assign these to the n_1 years of data.
- Identify the combination of $(P_{i1}, P_{i\tau'})$ such that if there are k events in year t that have the same C_i , then $(P_{i1}, P_{i\tau'})$ are selected such that the k largest values of b are chosen.
- Apply model M_1 to get the expected value of A_i corresponding to each $(P_{i1}, P_{i\tau'})$ combination.
- We can now update step 4 to select an improved value of A_R , and then update the corresponding model in step 5.
- This procedure can be applied recursively to convergence.
- The final model would have the form of Equation 2, and given the coefficients of the model, one could compute values of b for future events and use them for a payoff.
- A disaggregation of the index to local areas could be done by associating flooding duration with subareas, and also a depth based on the digital elevation models. However, primarily our index is most useful to cover the potential exposure that would occur for a regional government so that it can rapidly mobilize recovery and response, since the details of each candidate inundated area from the historical record may actually be quite different.

3.6 Future Steps: Machine Learning Applications to Storm Characteristic Identification, Event Return Period Estimation, and Automatic Infrastructure Identification

Two main directions are considered. The first focuses on a hydrometeorological analysis of catastrophic floods in the region, with the goal of reducing the uncertainty associated with the assessment of catastrophic flood risk in each of the countries. The second is an exploratory effort to identify key infrastructure elements and assets at risk in the event of a flood, using RGB and depth data from satellites, with the goal of reducing the uncertainty estimated with potential losses.

Hydrometeorological Assessment

The current investigation used the flood events and flood area identified in the Dartmouth Flood Observatory data to assess potential relationships between the flooded area and rainfall over multiple day windows and over different areas. The rainfall information used was from century-long re-analysis products. The assumption was that as the flooded area and flood duration of concern increased, large-scale, persistent rainfall events could be responsible and could be identified. This effort was compromised by the fact that the DFO data often covered multi-country flooding and reported flooded area associated with an event that was much greater than the total country area (e.g., Bangladesh), and were hence poorly correlated with the rainfall space and time windows we used. The historical DFO data on flood area, duration, loss, and population are the main long-term source of flood impacts data, but it is not clear how best to assess the uncertainty in their products. Many of the events recorded have long durations, e.g., around 20% are longer than 15 days and 7% are longer than 30 days. Time series of streamflow data from the countries of interest are often not available with adequate space or time coverage. We propose to acquire as much of these data as is possible for providing context and verification. If possible, the Delft group working with the World Bank will be requested to provide its calibrated model runs to reproduce historical floods as well. Thus, the first task is to consider developing a regional database or catalog of all extreme river floods, the associated rainfall, and the associated mechanism (e.g., tropical cyclone, typhoon, or tropical-extratropical interaction). The underlying hypothesis is that by taking a regional approach, rather than a country or river basin approach, one can identify the key mechanisms of rainfall associated with extreme flooding, explore whether the pooling of such storms over the region for return period estimation makes sense, and develop a return period perspective from these extreme storms that could be used with the Delft or other hydrologic models to assess flooded area risk in specific locations, or used directly for financial instrument design. The catalog of potential risk events would allow the assessment of their likely probability from a regional or national or river basin level. The ability to visualize such storms and their transposition from one region to another, and the resulting potential flood impacts, could then be done in a possible Phase 3 of the project. In Phase 2, we will develop machine learning algorithms (k-nearest neighbor and support vector classifiers, and multivariate copula-based methods) to guide the storm transposition and probability assessment across the region.

Identification of Infrastructure Elements at Risk

Once rainfall occurs, topography determines flow and inundation associated with flooding. Areas of greatest concern for flooding are typically located in the relatively flat floodplains adjoining the river network. Often, these are areas with intensive agriculture; infrastructure such as bridges, canals, levees, thermal power stations; and population centers with varying types of urban building assets and roadway networks. For catastrophic floods, the economic losses are dominated by impacts that relate to these elements. The hydraulic modeling of flood flows as well as the remote sensing–based identification of flooded areas is also complex as one approaches the urban and other infrastructure elements. Consequently, it is important to get a representation of these elements for reducing the uncertainty associated with hydraulic modeling and for economic impact assessment.

Building and engineering permits are required and collected in nearly all countries, and major real assets are also mapped using GPS services. However, it is not easy to get such information in a digital form that allows for better hydraulic representation or economic valuation. Our exploratory work shows that many of the typical infrastructure elements, such as bridges, power plants, water and wastewater treatment plants, hospitals, and civic facilities can be found in Google and Garmin maps, and moreover are searchable. This brings up the possibility of integrating this information into the Google Earth Engine–based product for economic loss assessment. Often ancillary information on such assets, e.g., the power output, or treatment capacity, or the number of households served, or the design number of vehicles for a bridge or roadway, may also be available electronically. Linking such information in a spatial model and integrating it with the potential flooded area could provide a pathway for loss uncertainty reduction.

A further degree of uncertainty reduction can be realized by considering the use of emerging machine learning methods developed by computer scientists at Columbia University (Shih Fu Chang, <http://www.ee.columbia.edu/dvmm>) and Rutgers University (Kristen Dana, <http://www.ece.rutgers.edu/~kdana/>). They have a research project working with the Defense Advanced Research Projects Agency (DARPA) that is aimed at recovering and classifying 3D objects from RGB and depth point clouds with training using markov random field classifiers and other machine learning tools (see references below). The goal of their project is 3D geometric model construction based on multi-view satellite images. In our context, we would explore the identification of building height and footprint, roof material type, lanes of traffic, and traffic density, and integrate these with available digital elevation models in the Google Earth Engine. The resulting application would focus on an assessment of population vulnerability, both in terms of the potential income levels of the people directly affected, and the people indirectly affected by the flooding of critical infrastructure.

We propose to explore both the use of map databases and the 3D geometry recovery in Phase 2, with the idea that a mock-up of the capability and utility can be demonstrated, to be followed by a detailed implementation in Phase 3. This task also provides an integration of the work done across the other tasks.

3.7 References

Feng, Jie, Wang, Yan, & Chang, Shih-Fu. (2016). 3D Shape Retrieval Using Single Depth Image from Low-cost Sensors. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*.

Khalil, Abedalrazq F., et al. (2007). El Niño–Southern Oscillation–based index insurance for floods: Statistical risk analyses and application to Peru. *Water Resources Research*, 43(10).

Ozaki, M. (2016). *Disaster risk financing in Bangladesh*. Retrieved from <https://www.adb.org/sites/default/files/publication/198561/sawp-046.pdf>.

Wang, Yan, Feng, Jie, Wu, Zhixiang, Wang, Jun, & Chang, Shih-Fu. (2014). From Low-Cost Depth Sensors to CAD: Cross-Domain 3D Shape Retrieval via Regression Tree Fields. In *European Conference on Computer Vision (ECCV)*.

Yang, Y. C. E., Ray, P. A., Brown, C. M., Khalil, A. F., & Yu, W. H. (2015). Estimation of flood damage functions for river basin planning: a case study in Bangladesh. *Natural Hazards*, 75(3), 2773–2791. <http://doi.org/10.1007/s11069-014-1459-y>.

Annex 3A.

Three analyses were conducted:

- Relationships among annual maximum of different rainfall windows, streamflow, and flood areas (Figures 3A.1–3A.3).
- Relationships among seasonal total rainfalls, maximum wet spell, streamflow, and flood areas (Figures 3A.4-3A.6). Seasons are October-February: winter; March-May: summer; June-September: rainy season. The sites are monthly streamflow (1956-2000) in Bahadurabad in Bangladesh, daily streamflow (1994, 1995, 1997–2011) in Beki River in India, and daily discharge (1992–2009) in Brahmaputra site in Pandu, Guwahati, India.
- Relationships among seasonal total rainfalls, maximum wet spell, water levels, and flood areas (Figures 3A.7-3A.9). Sites are daily water level at Bahadurabad, daily water level at Rajshashi Ganges, and daily water level at Bhairab and Bazar Meghna.

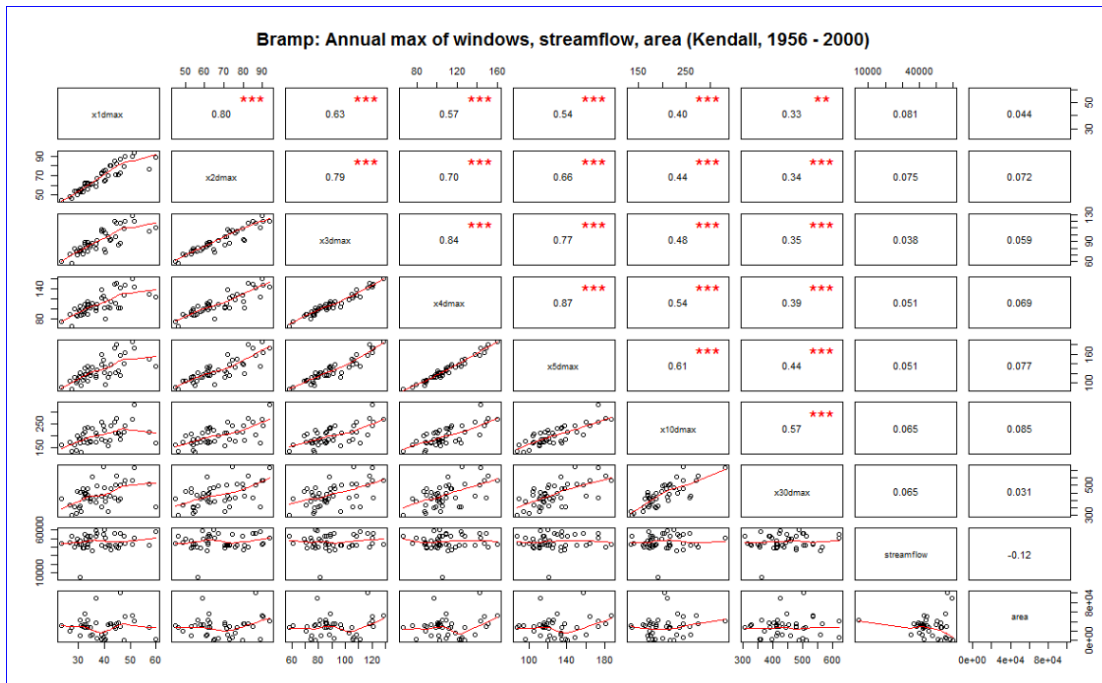


Figure 3A.1

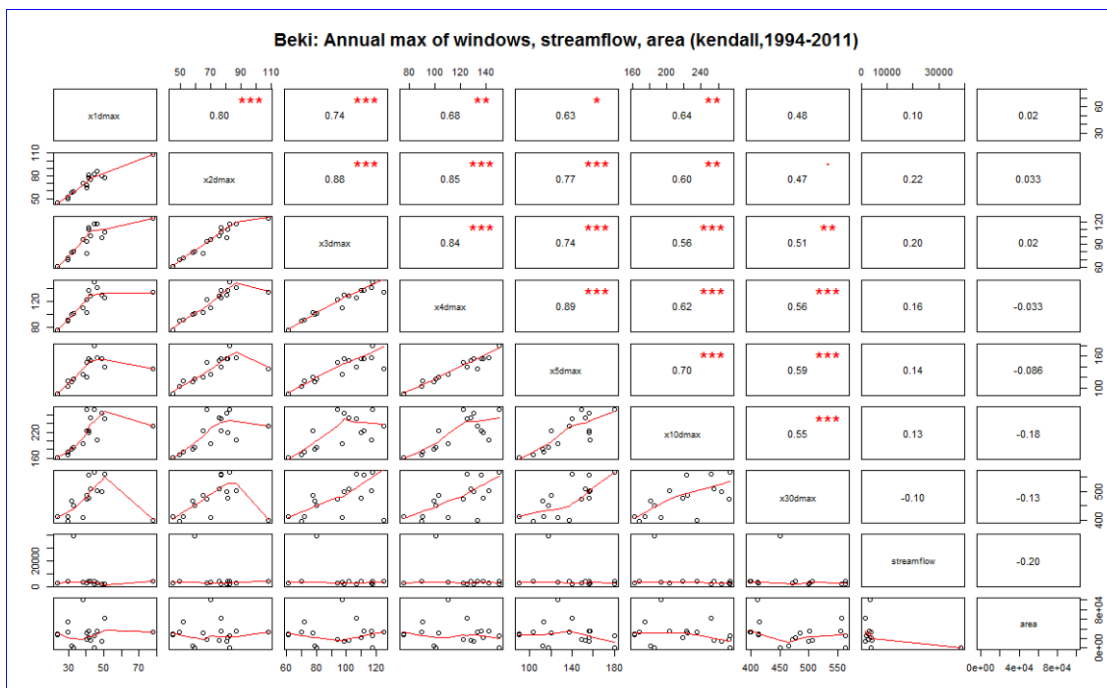


Figure 3A

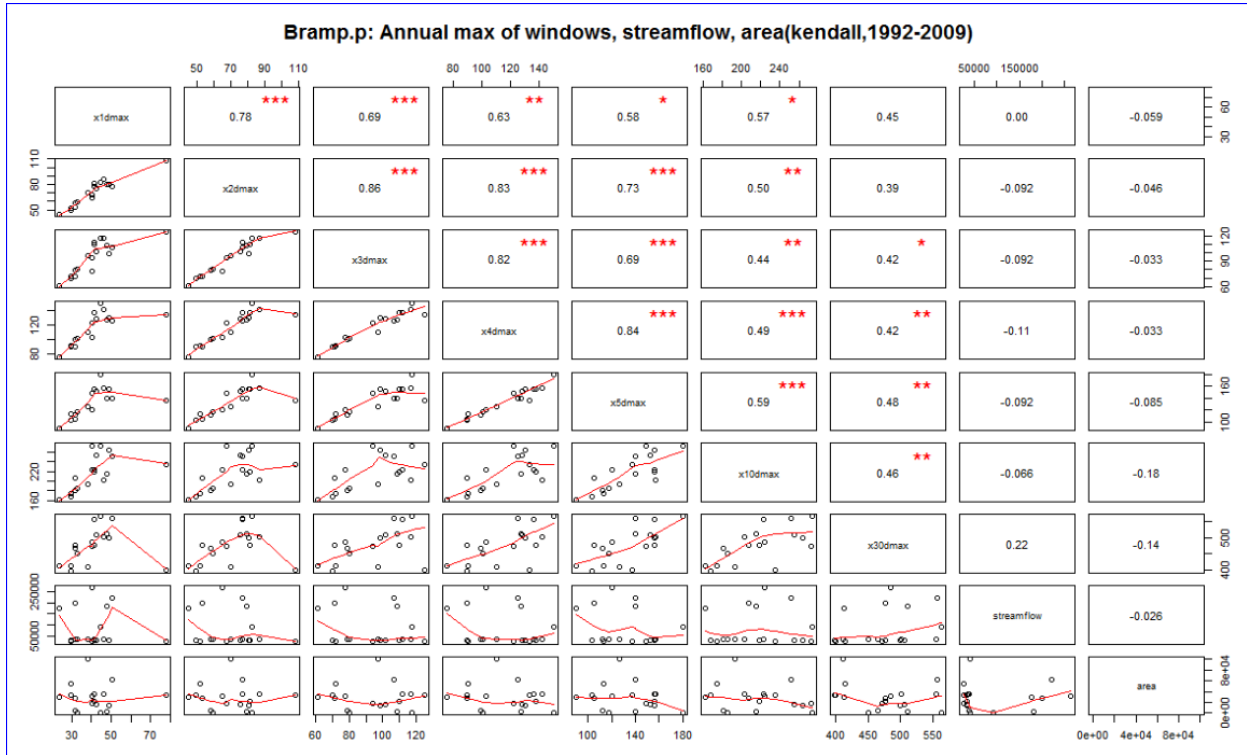


Figure 3A.3

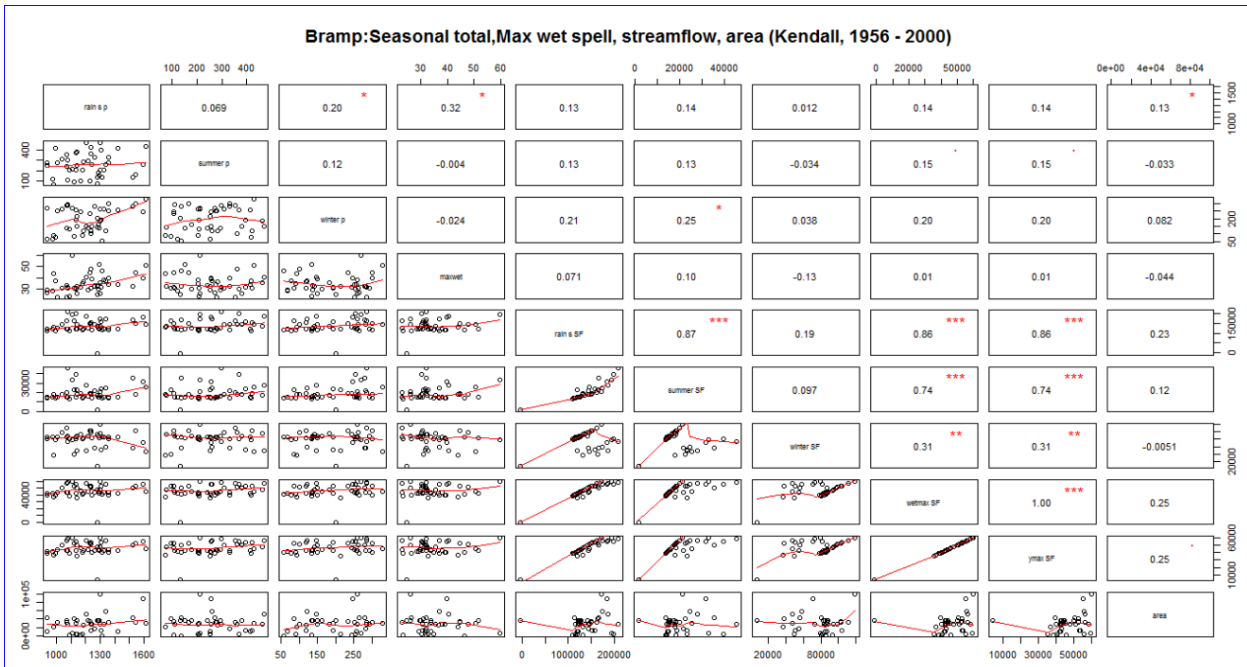


Figure 3A.4

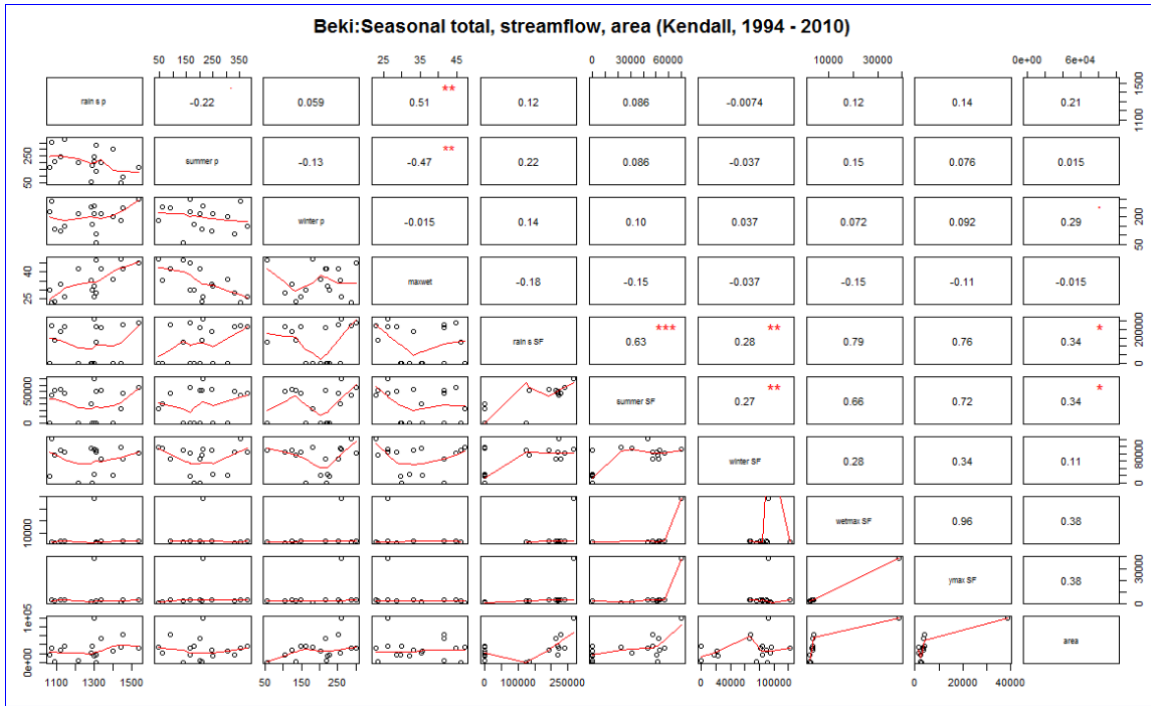


Figure 3A.5

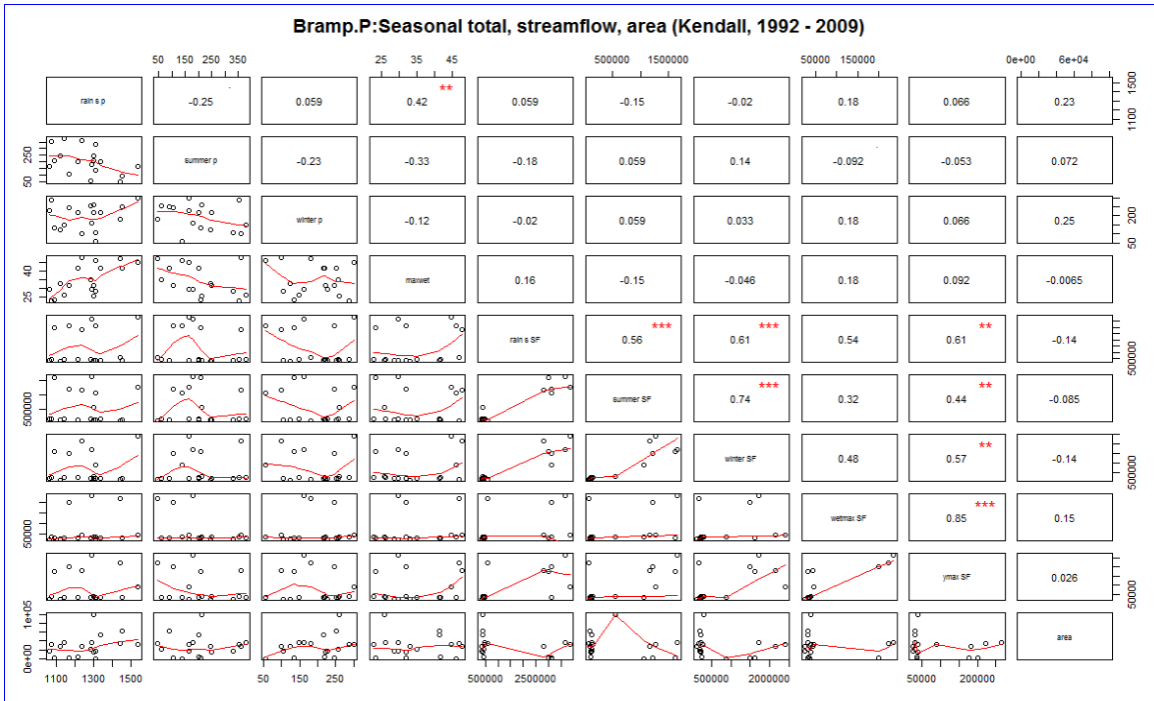


Figure 3A.6

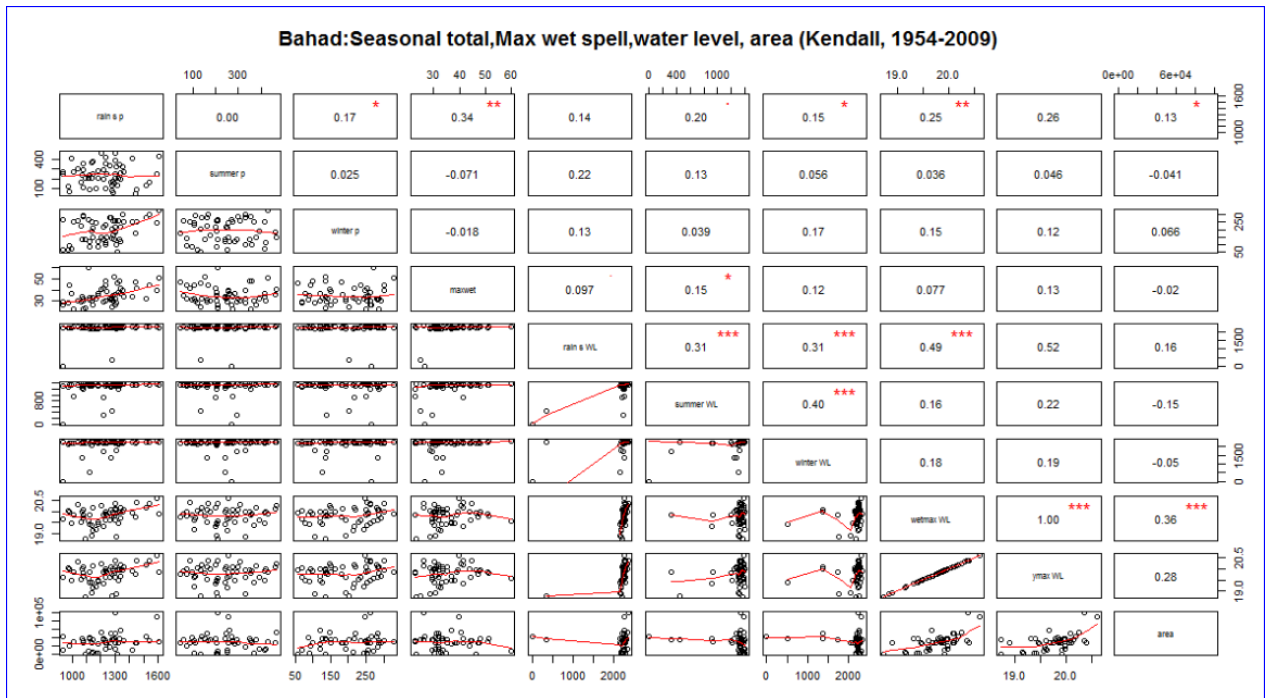


Figure 3A.7

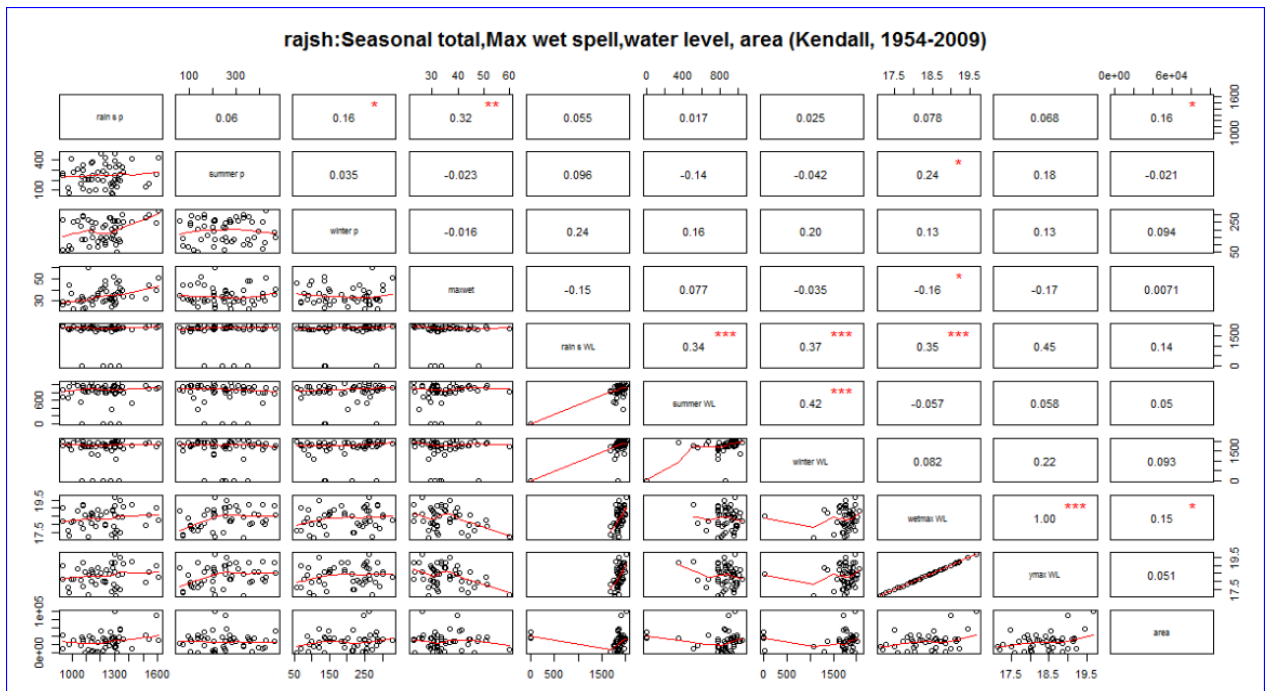


Figure 3A.8

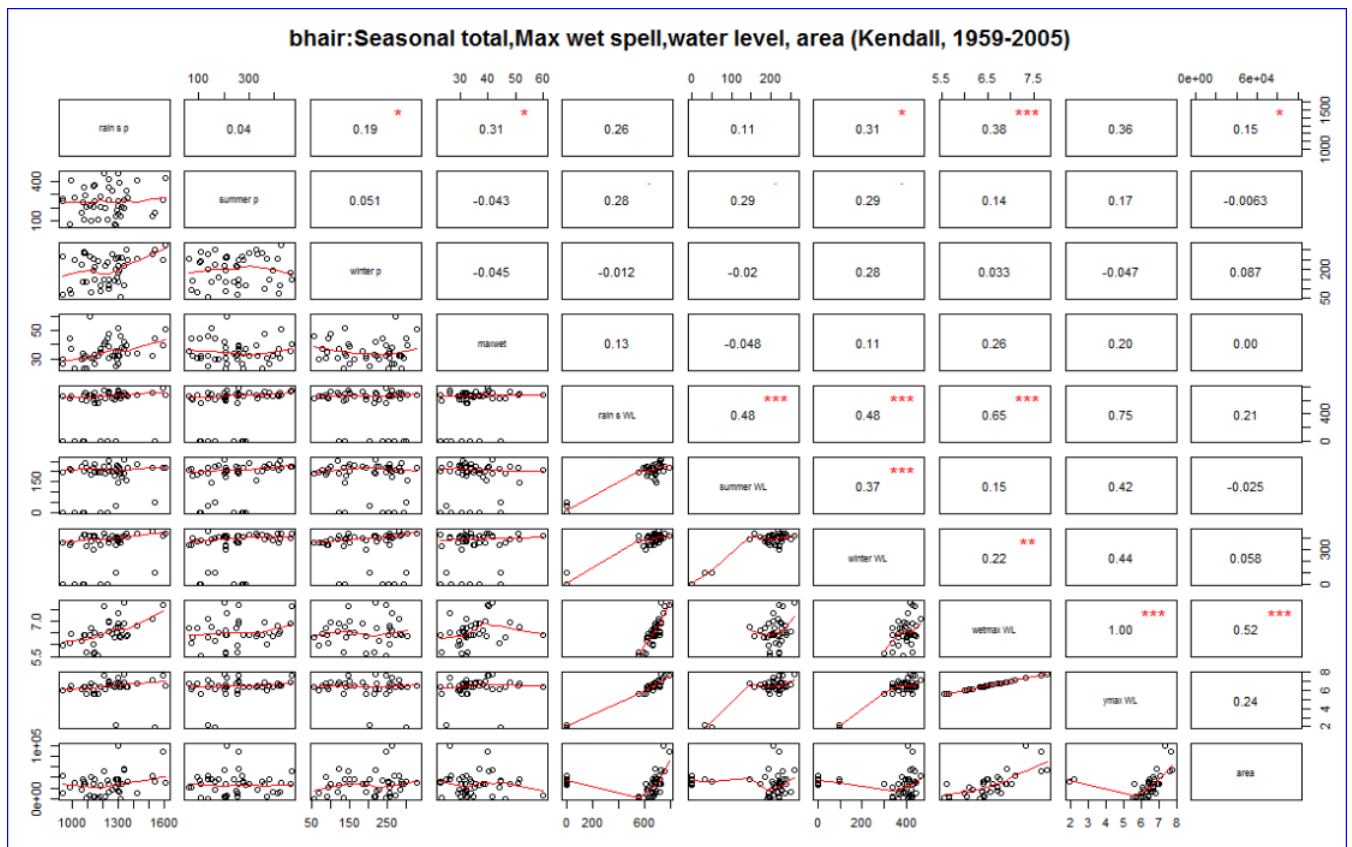


Figure 3A.9