Master Thesis, Department of Geosciences

Regional rainfall thresholds using global high resolution satellite precipitation estimates

A case study of landslides in Bangladesh

Knut-Johan Fagerland Kjelstad

UNIVERSITY OF OSLO

FACULTY OF MATHEMATICS AND NATURAL SCIENCES

Regional rainfall thresholds using global high-resolution satellite precipitation estimates

A case study of landslides in Bangladesh

Knut-Johan Fagerland Kjelstad

A Master Thesis in Geosciences

Discipline: Environmental geology and geohazards

Department of Geosciences

Faculty of Mathematics and Natural Sciences

University of Oslo

June 2011

Abstract

Landslides pose a risk to human life around the world. Since most landslides are related to intense rainfall, understanding the link between rainfall intensities and landslide triggering has been given great effort, especially through landslide mitigation by early warning. This requires estimating critical rainfall values for landslide initiation. Only 5 % of fatalities related to natural disasters occur in highly developed countries, reflecting the necessity of reducing risk in developing countries. Assessing critical conditions for landslide initiation in developing countries may be restricted due to insufficient rain gauge coverage; a potential solution to this is the use of satellite precipitation estimates (SPE). These data provide rainfall data at almost global coverage, at high temporal and spatial resolution in near-real time, but high uncertainty is related to their ability to capture the spatial and temporal rainfall variations.

This study is focused on the potential application of SPE data for assessment of critical rainfall values for initiation of landslides in areas with limited records of rain gauge and landslide data, using Bangladesh as a case study. An attempt was made for comparing TRMM based TMPA-RT rainfall estimates with a limited set of rain gauge data. The TMPA-RT product 3B42RT was applied for an initial study of the feasibility for applying SPE data in rainfall thresholds. Two multivariate techniques, classification tree analysis (CTA) and linear discriminant analysis (LDA), were tested. Scripts in Matlab (included as appendixes) were prepared both for capturing the satellite data from public servers and for performing the statistical analyses. These scripts can be reused in future studies that use these data sources and statistical methods for threshold assessment in other parts of the world.

It was found a generally poor correlation between rain gauge and TMPA-RT data; conversely intensity-duration plots of high intensity landslide triggering events displayed similar patterns. Rainfall data from Bangladesh proved to be highly homogeneous, resulting in low threshold performance for both methods. It is suggested that further studies focus on high-intensity events only, this applied for both thresholds analyses and for application of SPE data.

Acknowledgements

I would like to use this opportunity to thank the people who have helped making this thesis, and those who have motivated me these two years of study. Tanks to my supervisors Dr. Farrokh Nadim and Dr. José Cepeda, they have both given inspiring lectures in the fields of geohazards, and given interesting and challenging tasks. A special thanks to José for all the guidance during the work on this thesis, he seemed always to have time to help and to give advises, even in the latest hours. I really appreciate your help, I am sure I could not have had a better supervisor. Thanks to Jose and Mr. Rajinder K. Bhasin for getting the opportunity to do this thesis and for the helpful meetings in the early stages of my thesis. Thanks to Mr. Reshad Ekram, director of the Geological Survey of Bangladesh, and the Asian Disaster Preparedness Centre for providing the initial data for my thesis.

Thanks to my fellow student for all the interesting discussions, good talks during these two years, the same to all my friends at ICG, and thanks for being so welcoming and friendly.

I have saved the most important acknowledge to the one who deserves it the most: To my wife, Marianne, thank you for believing in me and encouraging me during these years, I would not have come this far without you. Your patience and your unconditional support has really been appreciated, I love you for it. I am truly a lucky man.

Front page photo: Courtesy of GSB / NGI

Table of contents

List of figures

List of tables

1 INTRODUCTION

1.1 Background

1.1.1 General

Landslides pose a threat to human life around the world. In the period 2002-2009, 201 landslide-related disasters were reported by the International Federation of Red Cross (IFRC 2010b), resulting in 7905 casualties. More than 95 % of the reported disasters and fatalities were caused by landslides of hydrological origin. Petley et al. (2007) found that more than 90 % of the landslide fatalities in Nepal are due to rainfall induced landslides. These numbers state the importance of landslide related studies and that assessing triggering conditions related to intense rainfalls is of importance.

In areas at high risk of landslides, mitigation measures are required to deal with the landslide threat. Mitigation of risk may be applied through appropriate physical measures like slope stabilization, lowering the probability (hazard) of landslide occurrence. In many landslide-prone areas such methods are however not feasible, e.g. if unstable slopes are too many or too expensive to stabilize. An alternative method for mitigation is early warning. By assessing and understanding the causes of landslide triggering, landslide events may potentially be predicted, and those exposed to the event can be warned and evacuated. Such systems are referred to as early warning systems (EWS).

As most landslides are triggered by intense rainfall, understanding the rainfall conditions causing landslides is important. This aspect has been studied extensively the last few decades, trying to establish the optimal condition (a threshold) for describing when (and possibly where) a landslide is expected to occur (Lumb 1975, Caine 1980, Crozier 1999, Guzzetti et al. 2007, Tiranti and Rabuffetti 2010). The analyses and methods applied for establishing a threshold have increased significantly the last years, including landslide susceptibility levels and applying advanced statistical methods (Jakob and Weatherly 2003, Santacana et al. 2003, Cepeda et al. 2009).

1.1.2 Situation for developing countries

The highly complex methods used to assess landslide susceptibility and establish hydro-meteorological thresholds today cause a high demand of detailed data for analysis purposes. Ideally, this should be yielded by high-density rain gauge networks recording at high temporal resolution (hourly or more frequently), detailed maps of soil conditions, digital elevation maps, other slope data, etc. (Guzzetti et al. 2006, Baeza et al. 2010). Geographical information systems (GIS) methods are often applied. Such data are for many regions of the world not readily available and operational rain gauges may be sparse, usually providing only daily measurements. Records of natural hazard events may also be deficient, especially in developing countries, making good analyses even harder.

Although accessibility of data is not the best in developing countries, study of natural hazards in these areas is of utmost importance. Studies show that developing countries are greatly affected by landslides, and Lacasse and Nadim (2009) found that natural disasters in highly developed countries stand for only 5 % of the causalities. Developing countries often have high population densities and problems related to poverty and housing situations; overcrowded poorly built houses in illegal

settlements on dangerous sites are common (Ekram and Khan 2008, Gov. Bangladesh 2010, IFRC 2010b).

Bangladesh is one of the countries facing the problems of illegal settlements and also illegal slope cutting (Ekram and Khan 2008, IRIN 2008). This is a problem in the eastern hilly regions of the country, and landslides are frequent. Bangladesh is situated by the Bay of Bengal experiencing tropical monsoon climate and frequent cyclones; mean annual precipitation above 3600 mm and daily rainfall exceeding 400 mm have been registered in some areas (Peel et al. 2007). In recent years landslides have caused more than 300 casualties, 127 of these during one single rainstorm event in 2007. The region is sparsely covered by rain gauges, and due to annual high intensity storms triggering landslides, a better coverage would be an advantage. The recent fatal landslide events suggest the need for mitigation measures for landslides in Bangladesh. No rainfall thresholds for Bangladesh were found in the literature.

1.1.3 Satellite precipitation estimates

The use of satellites has increased rapidly over the last few decades. Satellite data has to become easily available, resulting in a wide range of applications. One area of application has been for hydrological studies where infrared and passive microwave techniques are used for estimating precipitation (Vicente et al. 1998, Kidd et al. 2003, Scofield and Kuligowski 2003, Huffman et al. 2007). At present, these satellite precipitation estimates (SPE) provide almost full global coverage at high temporal and spatial resolution in almost real-time data. SPE data can hence provide precipitation estimates where ground based measurements are limited or absent. Many SPE products are freely available in digital files, providing an opportunity for low cost application of these data.

1.2 Current thesis

This study was based on the application of satellite precipitation estimates in non-instrumented areas, with the goal of estimating a regional rainfall threshold for landslide initiation in a study area. Bangladesh was chosen as a study area, as this is a region facing the problems addressed above.

The thesis may be divided into two parts; one part with the objective to evaluate if satellite precipitation estimates (SPE) can be easily and successfully applied to a region without (or with only limited amounts of) rainfall data for validation. The second part would aim to assess if critical rainfall values for landslide initiation in the region of Bangladesh can be established using a SPE product.

1.2.1 Validation of SPE products

For the first part a few selected SPE products were to be compared with only a limited amount of rain gauge data. This included developing a method for accessing the data using easily available computer processing power and software, and Matlab was chosen for the task. The first goal was to download satellite estimated rainfall products and create a series of daily rainfall corresponding to

the position of a rain gauge and subsequently compare the SPE rainfall series to the rainfall measured by the rain gauge. Several aspects of the rainfall data were explored: Seasonal variation, correlation of SPE and gauge data, ability to estimate landslide triggering storms, etc.

This part of the study would address uncertainty related to satellite data, and hopefully also reflect the challenges related to application of SPE without the possibility of high quality data validation for the assessed area. Due to the limited amount of data, it was expected large uncertainty in the data, and the goal was to indicate the applicability of such data given these limitations.

1.2.2 Establishing a threshold

The second part of the thesis was to assess the feasibility for establishing a rainfall threshold for the region of Bangladesh using SPE data. Such an application of SPE data was only found for a global scale threshold (Hong et al. 2006), and uncertainty was naturally expected for this part as well. Applicability and uncertainty were to be examined. Common statistical methods were applied in analyses aimed to establish the threshold(s), using the same software and SPE data as for the validation part. The statistical methods ability to establish reliable threshold conditions was evaluated.

1.2.3 Matlab

The computational programming language and software Matlab and its Statistical Toolbox where used to create scripts for capturing of data from public FTP-servers and for handling and performing statistical analysis of the data. The Matlab-scripts produced are included in appendix C, and can be reused in future studies that use these data sources and statistical methods for threshold assessment in other parts of the world.

1.3 Limitations

Limitations were especially related to the uncertainty of SPE data. Uncertainty exists in these data events when adjusted for local conditions using highly dense rain gauge networks (Shen et al. 2010, Sohn et al. 2010). The resolution of SPE data also limits their ability to pick up small scale temporal and spatial rainfall variations. Temporal scale limitations of the SPE products was not an issue for this study, as the rain gauge data applied for this study were of daily resolution.

The available rain gauge and landslide data for Bangladesh were sources of high uncertainties both considering validation of data and for establishing thresholds, due to the limited numbers of gauges and the short length of rainfall series available. The one rain gauge series available for a considerable length of time, ended at the point where a new version of the SPE product applied in this thesis became functional. Because of this, the most resent and best performing products performance could not be validated for Bangladesh.

Only few SPE products were applied in this study, other products may perform better for the studied area and improve the results. The number of applied SPE products was limited by the time consuming process of developing scripts for handling the relatively complex data sets representing the SPE products; as most products are developed by different institutions, the file structures are different for different products. Additionally the processing of the global data sets is also time-consuming.

1.4 Access to data

The initial landslide and rainfall data used in this thesis were provided by Mr. Reshad Ekram, director of the Geological Survey of Bangladesh) and the Asian Disaster and Preparedness Centre (ADPC) in collaboration with the Norwegian Geotechnical Institute (NGI). Some of these data are presented in appendix A. Requests regarding access to these data may be addressed to Mr. Reshad Ekram (reshadekram@gmail.com).

2 RAINFALL INDUCED LANDSLIDES

Landslides represent a major threat to human lives in most mountainous and hilly regions of the world. According to statistics from The Centre of Research on the Epidemiology of Disasters (CRED) landslides are responsible for at least 17 % of the world's fatalities due to natural hazards(Lacasse et al. 2010). This figure is probably underestimated as landslides may occur as a secondary event e.g. when triggered by another natural hazard, such as earthquakes. In these cases landslides fatalities are often accounted to the main triggering event, not to the landslide(s) (Guzzetti et al. 2007, Lacasse et al. 2010). Landslides may also pose an economical risk as properties, roads and other infrastructure and supply lines may be destroyed. Increasing population and urbanization lead to increased landslide vulnerability; as the population becomes higher, urbanisation expansion reaches unsafe areas, as these may be more readily available (Highland and Bobrowsky 2008).

Intense rainfall is probably the most important triggering mechanism of fatal landslides, and represents the main focus of the literature study. Petley et al. (2007) accounts 92 % of Nepal's landslide fatalities to intense rainfall, SAARC (2007) reports 89.2 % for 2007 worldwide landslide fatalities. Intense rainfall may trigger most slide- and flow type landslides, some which may cause severe impact on people and infrastructure; e.g. debris flows. The high impact of debris flows is due to the high velocity and high density resulting in severe destruction, and the high mobility and long runout resulting in impact over large areas.

2.1 Landslide types and classification

2.1.1 Classification

People from different areas of work, e.g. geologists, hydrologists, engineers, may work together when assessing landslides. As different terms tend to be used in different disciplines, the need of a universal classification arrives. Varnes (1978) classified landslides based on type of movement and type of material. He proposed the use of three different material types, rock, earth and debris, where earth is used for material of grain size sand or finer, while debris were used for coarser material. The types of movement in Varnes' classification were fall, topple, slide, spread and flow. The classification of Varnes (1978) has been widely adopted, at least in the English speaking part of landslide research. Hutchinson (1988) represent another recognised classification. The classifications of Cruden and Varnes (1996) and Highland and Bobrowsky (2008) is based on Varnes (1978), but with some convergence between the two classification branches of Varnes and Hutchinson.

There are many different landslide types classified, and many different triggering factors may apply to the different types. As the focus of this theoretical part is rainfall induced landslides, only the types most commonly related to rainfall triggering will be included. The following landslide types will be sorted by movement type, and there are generally two types most generally known for triggering by rainfall; slides and flows.

2.1.2 Slides

A slide is the down slope movement of rock or soil mass occurring on a surface of rupture or on a relatively thin zone of intense shear strain (Highland and Bobrowsky 2008).

- ***Rotational slides***: In rotational slides the surface of rupture is curved in a circular or a spoon-like shape. This kind of slides occurs in homogeneous material and may be relatively deep. Rotational slides may be caused by instant rainfall, a rise of groundwater level or a combination of the two. Rise of groundwater is caused by prolonged rainfall or snowmelt. Undercutting the foot of a slope by water erosion or human activity is also common factors in causing this kind of landslides.

- ***Translational slides***: In translational slides the surface of rupture is planar or undulating. The depth of the failure surface is relatively shallow compared to rotational slides. Translational slides commonly fail along geological discontinuities or in the rock-soil contact. The primary failure mechanism is intense rainfall, groundwater rise, snowmelt, flooding or other natural or human causes for inundation of water. These slides may disintegrate into debris flows at higher velocities.

2.1.3 Flows

Flows are spatially continuous movements of viscous behaviour where the shear surfaces are short lived and usually not preserved. The lower boundary of the flow may have differential movements or distributed shear through a thicker part of the bottom material. There is often a gradation between slides and flows depending on water content and evolution of the movement (Highland and Bobrowsky 2008).

- ***Debris flows*****:** Debris flows are rapid mass movements of soil and rock combined with very high water content. The flow may be of low viscosity, almost like water, or thick in sediments and highly viscous. In some literature they may be referred to as *mudslides* because of the high content of fine grains. Debris flows is commonly generated in easily eroded material, due to heavy rainfall or rapid snowmelt causing surface erosion. When this material adds up in streams and channels, possible from large areas, a debris flow is formed. The debris flows may have very high water content, and are often limited to gullies and canyons. Some debris flows are thus mistaken for floods. Wild fires may increase the susceptibility to debris flows, as the erodibility of the soils increase (Turner and Schuster 1996).

  Debris flows may also develop from nearly saturated translational or rotational slides if they gain water or disintegrates when accelerating. It is also found that debris flows may develop from sudden and rapid collapse of highly porous saturated material in steep slopes, due to a undrained failure mechanism of static liquefaction (Olivares and Damian 2007). Other causes of such flow can be dynamic liquefaction or impact collapse (Cruden and Varnes 1996).

  Natural dam breach may create debris flows, as the dam material may mix with the water flow and the rapid flow may erode the surface downstream increasing its content of granular material (Breien et al. 2008).

- ***Lahars*:** Lahars are debris flows that originate from the slopes of a volcano. They occur as volcanic airfall deposits, as tephra become mobilised. As for other debris flows, lahars may have various content of water and debris, resulting in differences in viscosity. Lahars can extend over very large areas (up to hundreds of square kilometres). Water is the primary trigger mechanism of lahars; it can originate from rainfall, crater lakes, and condensation of erupted steam or rapid snowmelt from volcanic venting.

  Lahars may be released almost instantly when subjected to heavy rainfall, in the matter of tens of minutes. The intensity required to initiate a lahar varies with slope conditions; volcanic activity may result in accumulation of airfall deposits. These loose deposits are easily eroded so lower rainfall intensity is required in triggering of lahars (van Westen and Daag 2005).

- ***Debris avalanches***: These flows may be larger flows that may be extremely rapid. They can involve as much as several tens of millions cubic meters of mass, reach velocities close to 100 m/s and a run-out of several kilometres from the source area. These avalanche occurs when an unstable slope collapse and the material disintegrates into a rapid flow. The slope instability can be created e.g. by weathering in steep slopes, or by sub-surface springs in karst environments. Slide type landslides and debris avalanches may transform into a debris avalanche if they disintegrates and gain velocity.

- ***Earthflows*:** As the term indicates this type of landslide occurs in fine grained material (*earth*), commonly clay and silt. It may also occur in clay-bearing strongly weathered bedrock. The movement of such flows is plastic or viscous with strong internal deformation. The rate of movement may be slow (creep) to very rapid, depending on material properties and water saturation. The depth of failure may be shallow or up to several tens of meters deep. Head scarp retrogression is common for these flows, causing the size of the affected area to increase with time. Earthflows are mainly caused by water, intense rainfall or snowmelt, rapid lowering of groundwater level, stream erosion in bottom of slope. Other causes related to initiation of earthflows may be ground vibrations and excessive loading of slope, both natural and anthropogenic, or other human activities changing the slope properties.

  A special kind of earthflow occurs in quick clay, only present in subaerial marine clays in North America and Scandinavia. These events are often very rapid and catastrophic, as the material loses all friction and suddenly liquefies. Earthflows in quick clay may destroy large areas and flow for several kilometres.

- ***Slow earthflows (creep)***: This kind of flow may be extremely slow. The movement is caused by internal shear stress insufficient to cause failure and result in internal deformation of the moving mass. Earthflows may be seasonal, continuous or progressive, where the progressive results in a failure, it can transform into other type of landslide. The velocity of slow earthflows is usually less than 1 m per decade, and will thus probably not be perceived as a landslide by most people.

2.2 Triggering mechanisms

2.2.1 Pore pressure and groundwater levels

Storms producing intense rainfall for a short period of time, or medium intensity rainfall a longer period of time (e.g. several days), are known to cause a large amount of landslides (Turner and Schuster 1996). The triggering of landslides succeeding heavy rainfall is caused by infiltration of water into the ground. The infiltration of water reduces soil suction (negative pore pressure) in unsaturated soil. High rainfall intensities and corresponding infiltration rates may even result in positive pore pressures (Iverson 2000, Tsaparas et al. 2002, Damiano and Olivares 2010). Increasing pore pressures reduce the effective soil strength, and a failure may occur. During intense rainfalls the infiltration of water will appear as a wetting front, or temporary perched aquifer, percolating into the ground(Wieczorek and Glade 2005). If the pore pressures created by this wetting front are insufficient to create failure, water will migrate down and add up to the groundwater. In thin soils the whole soil cover may become almost fully saturated for long duration and high intensity rainfall (Damiano and Olivares 2010).

Low intensity rainfall usually does not result in pore pressures high enough to create failure in shallow soils (Wieczorek and Glade 2005). On the other hand, prolonged infiltration will increase groundwater levels. Higher ground water levels increase the water pressure down through the soil. Increased pore pressure and corresponding lowering of effective strength may be crucial in deeper sediments; a deep seated failure may occur and trigger a landslide.

The intensity and duration of rainfall required to trigger a landslide depends on the soil properties like porosity, permeability and the total thickness of the soil. Layering may also be important if there are different properties in the different layers; as transition from higher to lower permeability may create a longer lasting temporary perched groundwater above the low permeability layer.

As triggering of landslides are generally based on build-up of pore water pressures, a landslide may not necessarily be triggered during a storm, but shortly after; this can be days but even month (Cepeda 2009).

2.2.2 Instant rainfall

Because of this dependency of soil properties to rainfall duration and rainfall intensity, it is useful to discriminate between instant and antecedent precipitation. Since deep seated failures are more dependent on antecedent precipitation than instant precipitation, it is also useful to separate between shallow and deep landslides. Rainfall induced shallow landslides include shallow translational slides, debris flows, lahars and shallow earth flows. Deep seated landslides dependent on antecedent rainfall will then be deeper translational slides, rotational slides and deeper earth flows. Debris avalanches and slow earthflows are generally not caused by rainfall.

Short duration high intensity rainfalls are a recognized cause of shallow landslides. Landslide events are known to occur in the relatively near future after such an instant rainfall. Shallow landslides are

often generated in steep slopes of soil or weathered rock during the most intense part of a storm. In the San Francisco Bay area in 1982, a 32 hours long intense rainstorm released 18.000 predominantly shallow landslides in soil and weathered rock (Turner and Schuster 1996). Loose and weak soils are especially susceptible to shallow landslides, as the soil may get eroded by surface water or fail due to increased pore pressures. Storms of very high intensity and short duration (e.g. 1 hour) may create such high surface runoffs and result in erosion and possible generation of a debris flows (Turner and Schuster 1996). Instant rainfalls of very high intensity will often not create sufficiently high pore pressures to initiate a landslide if the duration is short; the result may will only be a small water front migrating down the soil profile and result in a small increase in the groundwater level (Wieczorek and Glade 2005).

Some soil properties may also allow rainfall from intense storms to percolate quickly into the soil, resulting in fast saturation and rapid rise in groundwater levels. Temporary perched groundwater and corresponding high pore pressures may result in landslide triggering. High groundwater levels and almost full saturation of soils, even in steep slopes, may be reached. One such example is presented by Olivares and Damian (2007) from steep slope pyroclastic airfall deposits in Italy. Another example is from a study from New Zealand where no correlation where found between landslide initiation and rainfall duration longer than 2 days. The ground in the area consisted of coarse grained soils on volcanic ash, with shallow rooted vegetation; such conditions allow for rapid infiltration (Wieczorek and Glade 2005).

2.2.3 Antecedent precipitation

Instant rainfall alone is usually not enough to cause a landslide; triggering is also dependent on the antecedent rainfall i.e. groundwater and soil saturation conditions (Turner and Schuster 1996). The pre-storm rainfall conditions are thus important to identify the amounts of rainfall needed to trigger landslides. Wieczorek (1987) and Wieczorek and Sarmiento (1988) identified this significance analysing the antecedent rainfall preceding rainstorms causing debris flows in Northern California. It was found that antecedent rainfall were significantly higher for high intensity storms triggering debris flows, than for the storms not triggering debris flows. Storms with lower antecedent rainfall did not produce any debris flows, even with rainfall intensities higher than for the debris flow triggering storms. Storms of lower intensity did not result in any debris flows regardless of the amount of antecedent precipitation (Wieczorek and Glade 2005).

The antecedent rainfalls contribution to the triggering of debris flows is widely recognized, but studies do not agree upon the time period significant for establishing the critical pre storm conditions. Time periods ranging from 2 days to more than 2 week has been reported. This may be explained by seasonal variations affecting temperature and rainfall, i.e. affecting evapotranspiration. In colder periods the ground will stay saturated for a longer period of time after a rainfall, compared to warmer periods; this may have a significant effect on the amount of antecedent rainfall needed to cause a landslide. Additionally the time of year of occurrence of rainstorms may differ among climate regions. The significance of antecedent rainfall in an area may thus depend on both season and the climate region (Wieczorek and Glade 2005).

2.2.4 Other causes related to precipitation

Infiltration of water to a slope may also have other sources than rainfall. Additional sources of water infiltration may decrease the significance of instant and antecedent precipitation to triggering of landslides.

- ***Snowmelt***: Snowmelt is an important source of water in areas seasonally covered by snow; this may also apply to high altitude areas in warmer regions. Rapid melting of a snowpack due to sudden temperature increase is an important factor for supply of water to a slope, especially as increased water level as snowmelt may occur at high rates over longer periods of time. Rapid melting of snow from volcanoes is one of the factors causing Lahars. Snowmelt may also be a result of rainfall on the snowpack and thus result in additional water to the measured precipitation (Turner and Schuster 1996).

- ***Water-level changes***: Sudden lowering of the water level, such as in a river or in the sea by tidal changes, may result in destabilization of adjacent slopes. If the water level drops fast and the corresponding change in the slopes groundwater level is slow, this may create slope instability and increased shear in the soil. An increase in slope groundwater levels by prolonged water infiltration may cause the same kind of instability (Turner and Schuster 1996). This kind of water level changes favour deep seated failures (Iverson 2000). Thick uniform layers of soil or weathered rock are especially susceptible, and earthflows and rotational slides are typical results. (Turner and Schuster 1996). Increase in groundwater level can also accelerate this kind of landslides of slow velocity, as in the Super-Sauze earthflow (Malet et al. 2003).

- ***Flooding***: Flooding and landslides are closely correlated, since both are related to rainfall (or snowmelt). Floods are a secondary factor of rainfall as a result of increased surface- and groundwater-runoff from intense and prolonged rainstorms. Flooding may result in landslides due to erosion at the foot of adjacent slopes. Erosion of large volumes of soil from slopes and river banks may result in high amount of debris in the water flow and successive debris flows. Small steep channels are especially susceptible to develop debris flows (Highland and Bobrowsky 2008).

  Flooding from dam breaches may result in high erosion of soil and bedrock because of high intensity turbulent currents of water. For example, a glacial lake outburst in Norway eroded material along its path increasing the volume of debris with a factor of 10 from the dam breach to deposition (Breien et al. 2008).

2.2.5 Other causes increasing the occurrence of landslides

The landslide susceptibility of an area may change due to natural or human influence. Usually such an influence increase the landslide susceptibility with successive higher landslide occurrence, but severe events triggering great numbers of landslides may also reduce the landslide susceptibility for successive time periods:.

- ***Wild fires***: In addition to the loss of vegetation, wild fires may create a water repellent soil layer parallel to the surface. This results in surface flow during rainfall events that can erode loose grains of soil and other loose material at the surface. As all material carried by water

tend to gather in ravines channels and rivers debris flows may develop (Turner and Schuster 1996).

- ***Volcanic activity***: Accumulation of loose volcanic airfall deposits from volcanic activity increases slopes susceptibility to erosion and debris flows. A study of lahars by van Westen and Daag (2005) illustrates a poor correlation of intensity-duration rainfall thresholds for release of lahars. This was explained by the spatial and temporal variations in slope conditions due to deposition of new loose tephra deposits in the studied area. Other sources of water related to volcanic activity may increase the soil saturation prior to, or simultaneously to a rainfall event. Sources of water related to volcanic activity are given in the description of lahars.

- ***Earthquakes***: Strong earthquakes are a known trigger of many types of landslides; rock falls, soil slides and rock slides in steep slopes and e.g. earth avalanches on gentler slopes. Earthquakes are also known for its ability to liquefy loose, saturated, cohesionless soil. Landslides commonly occur by this process in low to moderate slopes, by the temporarily increased pore water pressures and decreased soil strength created by the ground shaking. Since saturation is one of the factors determining a soil susceptibility to earthquake induced liquefaction, rainfall preceding the earthquake event is of importance. High antecedent precipitation and i.e. high water content in the ground will increase the risk of landslides triggered by an earthquake. Timing of earthquake to climatic influence is thus important (Turner and Schuster 1996).

  Wieczorek and Glade (2005) point out that this climatic influence on earthquake-triggered landslide may be demonstrated by a comparison of events in the same region. A comparison of two events in the area of San Francisco indicate the importance; one event with 1-, 3-, and 6- month antecedent rainfall 50-100 % above the normal resulted many deep seated failures and debris flows, the other event triggered many shallow events but no debris flows.

- ***Human activity***: There are many factors of human activity that may increase the possible release of a landslide. Slopes may become destabilized by excessive loading of the slope or by undercutting of the slope foot, resulting in instability by oversteepening of the slope. Drainage patterns may be disturbed or changed e.g. by building roads (Fiorillo et al. 2001, Guadagno et al. 2003). Removal of vegetation resulting in increased erosion and change in infiltration. Introduction of more water to a slope through irrigation or leaking water pipes are other possibilities (Highland and Bobrowsky 2008).

3 RAINFALL THRESHOLDS

A threshold defines a minimum value or condition on which exceeded an event is likely to occur. For rainfall-induces landslides the threshold condition may be rainfall, soil moisture or other hydrological conditions, resulting in one or several landslides. Thresholds can be defined by either as physically- or empirically-based models. Caine (1980) presented the concept of rainfall thresholds, using a rainfall intensity-duration (I-D) relationship to assess the rainfall needed to trigger shallow landslides and debris flows on global scale. Caine (1980) used 73 different landslide triggering storm events to define the empirical threshold with the general form

$$1\ I = 14.82\ D^{-0.39} \quad \textbf{(1)}$$

where *I* am the rainfall intensity and *D* is the rainfall duration in hours. Later rainfall thresholds have been widely adopted; the three last decades numerous rainfall thresholds have been presented both at local, regional and global scale – different empirical methods and equations has been proposed (Guzzetti et al. 2007).

Description of rainfall thresholds are generally based on Guzzetti et al. (2007).

3.1 Physically based thresholds

Physically-based models use the concept of slope stability in models for an extended area – traditionally single slope models have been used. For good predictability of slope failures the models have to consider and include the spatial and temporal variability in the soil conditions. Processes-based infiltration models and rainfall patterns are used to predict slope stability and the soil conditions needed for develop a slope failure. Two different physically-based threshold models have been used; The "*leaky barrel*"-model (Wilson 1989) and the *antecedent soil water status* (ASWS) model (e.g. Glade et al. 2000). The *leaky barrel*-model use numerical modelling on the simple concept of a leaky barrel which receive water at one rate and loose water at another rate. ASWS is a simple conceptual water balance model for daily estimation of soil moisture conditions. The receiving rate is a combination of instant and antecedent precipitation and the "*leakage*"-function is a drainage function based on storm discharge hydrographs. Both models have been used successfully to predict shallow landslides.

As physically-based models include the spatial and temporal variability in soil properties and rainfall patterns, they should theoretically be able to predict both the time and location for a landslide. These are good conditions for incorporation of physically-based models in early warning systems, but there are limitations. The physically-based models coupled with information like soil properties and land-use may be applied in GIT-systems. There is though limitation using these models: To be able to make good landslide forecasts the models require high detailed soil, geology and surface information of high spatial resolution. This kind of information is hard to collect over larger areas, suggesting that these models are best suited at local scale. The predictability of the models also depends on the spatial and temporal resolution of precipitation data. To increase predictability, the physical-based models are calibrated against landslide event with good corresponding precipitation data and known

location and time. Because of the uncertainty of soil conditions the physically-based models are less efficient for deep seated slides (Guzzetti et al. 2007).

3.2 Empirically based thresholds

Empirically-based models define thresholds for landslide initiation by the study of rainfall events that have already resulted in landslides. Physical conditions like slope, soil properties, and evapotranspiration are not considered. For different sets of conditions, the amount of rainfall needed to trigger a landslide will vary; thus is the homogeneity of the area assessed with an empirically based model important. The advantage not considering the ground conditions is avoidance of the need of high detailed soil condition and property data. As for physically-based thresholds the quality of rainfall data is important, as rainfall rates may have high spatial and temporal variability. The position of a rainfall measurement compared to a triggered landslide is important for the reliability of the rainfall data. Empirically-based thresholds can, as the physically-based thresholds, be used in landslide early warning systems.

Guzzetti et al. (2007) suggest that empirical thresholds can be grouped in three categories: (a) thresholds combining precipitation measurements from specific rainfall events, (b) thresholds that consider antecedent conditions, and (c) other thresholds, including hydrological thresholds. Properties of the different groups and comparison of the published thresholds of each group will be reviewed later.

3.2.1 Best fit of empirical thresholds

Empirical thresholds are usually visualised as a line separating the triggering conditions from the non-triggering conditions (Guzzetti et al. 2007). The reliability of this line depends on the type of data used to define the thresholds. Generally three different kinds of data may be used:

(1) inventory of landslide triggering rainfall events, as used by Caine (1980)
(2) inventory only of non-triggering rainfall events
(3) inventories using both triggering and non-triggering events, e.g. by Giannecchini (2005).

For the use of (1), the threshold will give the lowest value known to trigger in landslides in the area studied. Threshold of type (2) give the highest value where landslides did not occur. (3) include both triggering and non-triggering events, and often show that lower triggering events may have lower values than the highest non-triggering events – referred to as false negatives and false positives. This indicates the inadequacy of (1) and (2), but if available data only enable use of type (1) or (2) thresholds, these thresholds are better than not having thresholds at all. The thresholds will still indicate what kind of rainfall may trigger landslides in the current area.

As empirically-based thresholds generally do not account for properties such as soil and climate conditions, the area used for defining the threshold should be relatively homogeneous. The use of type (3) thresholds can indirectly account for different conditions by introducing different susceptibility levels for combination of different soil, climate and other conditions (Cepeda et al. 2010).

3.2.2 Classification of thresholds

The empirically-based thresholds represent a large group of rainfall thresholds; more than 100 thresholds of different geographical extent have been published. In a review of all rainfall thresholds published until 2005, Guzzetti et al. (2007) reveal that no unique set of measurements exist to characterise rainfall conditions for triggering of landslides. Guzzetti et al. (2007) point out that the definition of rainfall intensity is a key factor for the empirical studies, especially for rainfall of longer durations. They further stress that peaks of rainfall intensity are often important for triggering of shallow rainfalls; this may be recorded by "instantaneous" measure of rainfall rate. This correlates with the influence of rainfall and water infiltration on soil properties and triggering conditions, as explained in section 2.2.1. If rainfall intensity measurements are averaged over a longer period of time the "peaks" will disappear from the data. The difference between short and long duration rainfall (instant and antecedent rainfall) is thus very important (Guzzetti et al. 2007). Many different rainfall measurements have been used for empirical rainfall thresholds: Rainfall event duration, daily rainfall, critical hourly rainfall and antecedent rainfall to mention some. A complete list with explanation of the different variables may be found in Guzzetti et al. (2007, 2008).

Rainfall thresholds may be classified according to the geographical extent of the area they have been defined for:

- ***Global thresholds***: Global thresholds, as first proposed by Caine (1980) and lately updated by Guzzetti et al. (2008), try to define a worldwide minimum level at wish landslides do not occur, regardless of geology, soil cover, climate conditions, etc. Local or regional rainfall patterns or historical rainfall data may be used. Lately the use of rainfall estimates from remote sensing techniques, e.g. satellite based, has become an option (Shen et al. 2010).
- ***Regional thresholds***: These are considered for areas at the size of a few to several thousands of km³. Regional thresholds are considered for areas of similar climatic, physiographical, and meteorological and soil characteristics. As long as the characteristics of the area are approximately the same, defined thresholds may be suitable for an early warning system. Rainfall can be based on quantitative rainfall measurements, estimates or forecasts.
- ***Local thresholds***: The local thresholds are defined for areas extending from some few to some hundreds of km³, and consider local climatic regimes and geomorphological settings. Local thresholds may even be applicable for single landslides or small groups of landslides.

3.2.3 Thresholds using event rainfall measurements

The empirically-based rainfall thresholds that use event rainfall measurements use single or multiple landslide triggering (or non-triggering) events. Intensity-Duration thresholds (as used by Caine (1980)) are the most frequently used type of threshold, but others have also been used. Including all used types, the event rainfall thresholds may be subdivided into: (a) intensity-duration (ID) thresholds, (b) thresholds based on total event rainfall (E), (c) rainfall event-duration (ED) thresholds, and (d) rainfall event intensity (EI) thresholds.

Except from the E-thresholds, all thresholds using rainfall measurements are defined generally for shallow landslides or debris flows, including some few for lahars.

- ***ID thresholds***: The equation of the ID-thresholds proposed so far, presented by Guzzetti et al. ((2007), has the general form

2 $$I = c + \alpha \, x \, D^{\beta} \quad \textbf{(2)}$$

where *I* is (mean) rainfall intensity, $c \geq 0$, *D* is rainfall duration, α and β are parameters. For most of the ID thresholds proposed so far $c = 0$, resulting in a simple power law function for equation similar to *eq. (1).* The duration (*D*) rage from 1 to 100 hours for most thresholds, and the intensities (*I*) from 1 to 200 mm/h.

Comparison of all ID thresholds reveals some interesting characteristics (Guzzetti et al. 2007): the local thresholds are slightly higher than the regional thresholds, and the regional thresholds slightly higher than the global. Guzzetti et al. (2007) attribute this to the different geographical scales and the sampling resolution. At larger scales the observational data is often acquired by regional averaging and often have lower sampling resolution. In these processes some intensity peaks may disappear from the data. Thus higher sampling resolution results in higher, and probably more realistic (Guzzetti et al. 2007), rainfall intensities for release of landslides. Differences have also been discovered for regional and local thresholds in same and similar areas. This may be due to differences in properties like geology and soil characteristics, variability in rainfall conditions or differences in the completeness of the data used.

Thus are the local thresholds generally providing the most reliable ID-thresholds for landslide initiation - followed by the regional thresholds. There are still limitations; the models cannot easily be exported from one area to another because of dissimilarities like meteorological conditions (Jakob and Weatherly 2003). Morphological and lithological differences make exporting even harder. Limitations like this are expected as e.g. infiltration rates depend on soil properties and soil cover, and local rainfall on meteorological conditions.

Solving the problem of extraction of thresholds from one area to another has been attempted by normalisation of rainfall data. Two climatic indexes for normalisation have been introduced; *mean annual precipitation* (MAP) and the *rainy day normal* (RDN) (Wilson 1997). Significant differences still remain after normalisation of data. The rainy day normal is established dividing the MAP by the average number of rainy days (RDs).

- ***E-thresholds***: Some authors have tried to establish empirical rainfall thresholds based on the total amount of rainfall occurring during one rainfall event (*E*). Different rainfall variables have been used (Guzzetti et al. 2007). One of the variable used are mean annual precipitation (MAP), and thresholds based on a storm events total amount of rainfall in percentage of the MAP. Govi and Sorzana (1980) found that a higher amount of rainfall were needed to trigger landslides in areas with high MAP, compared with areas of low MAP. This indicates that slopes "*adjust*" to the "*normal*" rainfall conditions in the area situated. Other

conditions like geology and soil properties also change from one area to another, explaining the difficulties with transferring thresholds with the help of MAP-normalisation.

- ***ED-thresholds***: The rainfall event-duration thresholds link the rainfall event duration to the cumulative event precipitation, including the final critical rainfall. Normalised variables have been applied to some ED-thresholds. Most published ED-thresholds are comparable regarding slope trends, but the cumulative event rainfall differs by a factor of 2. This may be explained by the difference between the published thresholds, e.g. in geographical extent (locally to globally) and physiographic conditions.

- ***EI-thresholds***: The event intensity thresholds link the event rainfall (*E*) to the average rainfall intensity (*I*); this then gives the event intensity (*EI*). Normalised rainfall values have also been applied for EI-thresholds. Different sets of data have been used: e.g. hourly rainfall replacing event rainfall. EI-thresholds have also been determined using only the final stage of rainstorms preceding the landslide trigger.

3.2.4 Thresholds that consider antecedent conditions

As explained in section 2.2.1 groundwater level and soil moisture are factors that predispose slope failure and triggering of landslides. As antecedent rainfall influences these conditions the amount of antecedent rainfall can be used as an indirect measure of the slope conditions, and corresponding slope stability conditions. The duration of the antecedent rainfall influencing the soil conditions depends on e.g. soil properties and infiltration rates. Different methods of how to use antecedent rainfall in the purpose of determining rainfall threshold have been proposed.

As the soil conditions dependency to antecedent rainfall depends on soil properties, the duration of antecedent rainfall influencing landslide triggering will vary in different areas. The key difficulty is thus to establish the duration of antecedent rainfall to consider. Most thresholds considering the antecedent rainfall is defined plotting instant rainfall against antecedent rainfall. Many different rainfall durations have been used, and most published thresholds use several durations of antecedent rainfall to find the duration with best correlation to landslide occurrence:

Durations of antecedent rainfall investigated rage from 1 to 120 days, and have been plotted against 1, 2 or 3 days of intense rainfall (instant rainfall). The periods found to correspond best with landslide occurrence also vary; Pasuto and Silvano (1998) found the best corresponding antecedent rainfall to be 15-days, while Cardinali et al. (2006) found best correlation to 3- and 4-month antecedent rainfall. One exception is a study of lahars where the 1 hour intense rainfall where plotted against antecedent rainfall of 4 days. Normalised values of instant and antecedent rainfall have also been used. As discussed earlier these differences may be attributed to different factors like soil and surface characteristics, climate and meteorological variations in addition to heterogeneities or incompleteness in the data used for determining the thresholds.

3.2.5 Other thresholds

A few other types of thresholds have also been suggested, including several rainfall related factors not used in the other types of thresholds. Measure of evapotranspiration, rainfall normalised by expected value of a storm of 5-years return period, mean daily discharge data are some of the parameters introduced (Guzzetti et al. 2007).

Jakob and Weatherly (2003) established a threshold combining the duration of discharge data exceeding 1 m^3/s, cumulative 4 weeks antecedent precipitation and cumulative 6 hour precipitation. This kind of threshold show it is possible to include more than just two parameters in a rainfall threshold. Increased number of variable does though increase uncertainties, and more complicated methods of analysis are required. Here a huge amount of hydrological and rainfall variables were analysed; one multivariate statistical method, *discriminant analysis*, was used to select the variables.

3.3 Operation of thresholds in early warning systems

3.3.1 Remote sensing rainfall products

Areas where establishing early warning systems may be useful may not necessarily be covered by rain gauges. Establishing the rain gauges and the systems for collecting data - physically or electronically - may be quite costly and maybe not achievable, especially for larger areas. An alternative to provide precipitation data for large geographical area is the use of precipitation estimates, acquired by remote sensing techniques. Such estimates are usually acquired using radar technology, either ground based or satellite based.

The use of satellites has increased rapidly the last few decades. The large number of satellites available today gives the possibility to gather data covering most of the earth surface, in near-real time and in relatively high resolution. There is a wide range of applications for global and regional satellite data in e.g. the science of hydrology or climatology (Kidd et al. 2003). Precipitation is one factor hard to estimate over larger areas using ground-based tools, due to its small scale variability in space and time. Satellite imagery is playing a key role in precipitation estimates, especially on a regional and global scale. Satellite precipitation estimates (SPEs) provide the possibility to gather precipitation data in areas not covered by rain gauges or other ground based tools. Today SPEs provide almost a full global coverage of both temporal and spatial high resolution.

High resolution SPE-data is commonly produced in a global grid where the pixel size is 0.25° x 0.25° or less, with a temporal resolution between 30 min and 6 hours (Kidd et al. 2003). As the pixels size is given in degrees, the earth surface pixel size will vary with latitude bands, and decrease with distance from equator. Both spatial and temporal resolution depends on the type of satellite data used, and vary between different SPE products.

3.3.2 Satellite precipitation estimates (SPE)

The focus of this thesis has been to use high-resolution global rainfall estimates to establish thresholds for landslide initiation. Use of satellite precipitation estimates (SPEs) provides the

possibility to collect rainfall data for areas not covered by rain gauges, or where the rain gauge network is sparse. Global precipitation estimates are obtained using imagery from generally to types of satellites; (1) geostationary or geosynchronous (GEO) satellites, and (2) polar-orbiting satellites known as low-Earth-orbit (LEO) satellites. The GEO satellites generally provide the highest temporal resolution, as they do not move significantly in relation to the earth surface. In some areas data may be provided as often as every 15 min at a spatial resolution of 1-4 km (Scofield and Kuligowski 2003).

3.3.3 Available SPEs

There are a huge number of different SPEs available today using IR and passive microwave imagery. Most are based on the same principles and techniques, but use different algorithms and methods for estimating the precipitation – thus resulting in different numerical precipitation estimates (Shen et al. 2010, Sohn et al. 2010). Several criteria had to apply for the SPE(s) selected for this study; (1) The precipitation product(s) had to cover the selected study area. For the possibility of applying the methodology also for other areas, this basically resulted that the precipitation product(s) should provide global or semi-global data. (2) The data should be freely available, and possible to handle with common programs like Matlab, FORTRAN, Octave, R, etc. This is important for application of the methodology by other people and institutions with limited funding. (3) The SPE should provide data at high temporal and spatial resolution, as landslide triggering depends on local differences in rainfall and soil properties. Data frequency of 1 hour or less is desirable as rainfall frequency may change drastically from one hour to another. (4) The precipitation estimates should be available in longer time series, to provide rainfall data from as many landslide events as possible.

Several satellite precipitation estimates are available from different research organisations:

- **Hydro-Estimator (H-E):** The Centre for Satellite Applications and Research (STAR) of the National Oceanic and Atmospheric Administration (NOAA) use different algorithms for SPEs (NOAA STAR 2011). The algorithms create near real time estimates presented graphically in the STAR Satellite Rainfall Estimate webpage. The Hydro-Estimator algorithm produces global data archived in digital files publicly available in a ftp-server (ftp://ftp.orbit.nesdis.noaa.gov/pub/smcd/emb/f_f/hydroest/world/world/).
- **TMPA-RT:** The Real-Time TRMM Multi-satellite Precipitation Analysis. Precipitation estimates based on NASAs Tropical Rainfall Measuring Mission (TRMM). Several precipitation products are provided, and three semi-global high-resolution products are publically available, 3B40RT, 3B41RT and 3B42RT (Huffman and Bolvin 2010). These may be accessed in digital files from a ftp-server at ftp://trmmopen.gsfc.nasa.gov/pub/merged/.
- **CMORPH:** NOAAs Climate Prediction Centre (CPC) uses an algorithm known as the CPC Morphing Technique (CMORPH) (NOAA CPC 2011). Precipitation data are estimated in data-sets of different spatial and temporal resolutions, but only one data set is fully available at ftp://ftp.cpc.ncep.noaa.gov/precip/global_CMORPH/3-hourly_025deg/.
- **PERSIANN:** The Centre for Hydrometeorology and Remote Sensing (CHRS) at the University of California, Irvine (UC Irvine) creates satellite based precipitation estimates based on the use of artificial neural network; PERSIANN - Precipitation Estimation from Remote Sensing Information using Artificial Neural Network (http://chrs.web.uci.edu/persiann/).

- **NRL-blended:** The SPE product of National Research Laboratory (NRL) called NRL-blended (http://www.isac.cnr.it/~ipwg/algorithms/inventory/NRL.pdf) also covers all criteria. These data are available at a spatial scale of down to 0.1° and a temporal scale of 3 hours. The ftp-server providing the NRL-digital datasets (ftp://ftp.nrlmry.navy.mil/pub/receive/turk/global_rain) were not found accessible during the period of work on this thesis, spring 2011.

4 THE STUDY AREA - BANGLADESH

Bangladesh is a country at high risk of e.g. natural hazards like floods, earthquakes and landslides (Gov. Bangladesh 2010). The country is situated in a belt of tropical climate (Peel et al. 2007) neighbouring the Bay of Bengal in the south. As a result Bangladesh is suffering from intense prolonged rainfall in the monsoon season, occasionally causing landslides and flooding. Extreme rainfall events may occur the monsoon season and due to cyclones.

During these events several landslides may occur in the area of the rainstorm, often with fatal consequences. The occurrence of fatal landslides in Bangladesh is frequent, and especially the landslide events during the last few years have resulted in high number of victims (IFRC 2007, 2010a). Two recent and devastating events are worth noticing; the landslide events of 11 June 2007 in Chittagong city and surroundings and of 14-15 June 2010 in the districts of Cox's Bazar and Bandarban. These events caused at least 127 and 55 deaths respectively (Ekram et al. 2007, IFRC 2007).

4.1 Physical Geography

Bangladesh is situated in South Asia bordering to India, Burma and the Bay of Bengal (Figure 4-2). Most of Bangladesh consists of low altitude areas of resent plains. These plain are mostly flood plains, delta- and tidal plains and may thus be inundated by flood or tidal water (Figure 4-1). Only 18 % of the land area are terrace or hilly areas (Mahmood and Khan 2008). The hilly areas are located mainly in southeast, but some hills of same origin are also situated in northeast (Figure 4-1), known as the Chittagong hill tracts and the hill ranges of Sylhet respectively (Mahmood and Khan 2008). The hilly areas consist of rocks that have been uplifted and folded into a series of sync- and anticlines (Mahmood and Khan 2008) that are tight, plunged and faulted (Ekram et al. 2007). The alignments of the faults are NNW-SSE, a trend also visible for the ridged in the elevation map in Figure 4-1.

The geology in the hilly areas is dominated by unconsolidated sedimentary rocks. The higher hills in the Chittagong area and the hills of Sylhet are of Oligocene to mid-Miocene age; while Chittagong's lower hills are of late-Miocene. These hills are of sedimentary rocks are mostly sandstone, siltstone and shale, but also limestone and conglomerate. The lower hills are of less consolidated rocks, mainly sandstone and shale (Mahmood and Khan 2008). Geotechnical investigations of landslide sites following the devastating 2007 Chittagong landslide event confirmed this: Most samples were found to be of poorly consolidated, loose, friable sandstone of different coarse to fine sand ratios and with little clay (Ekram et al. 2007). The report from this investigation also characterizes the geomorphic setting in the Chittagong area, and is here used as an example on how the landslide prone areas of

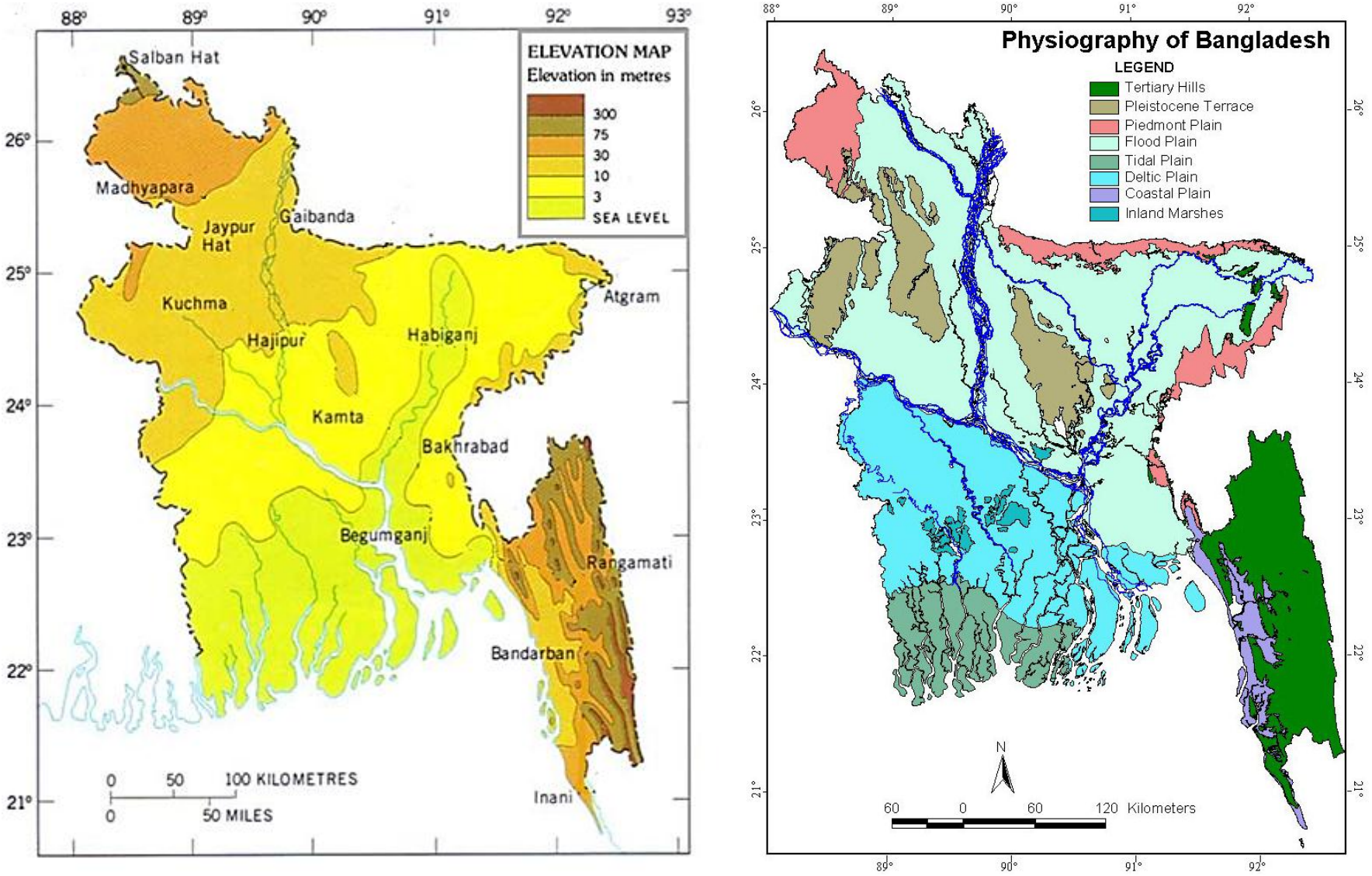


Figure 4-1 Left: Elevation map of Bangladesh (Sarker et al. 2010). Right: Physiographic map of Bangladesh (Mahmood and Khan 2008).

the hilly areas of Bangladesh may look like: There are low dome shaped hills in the area with flat tops and short flanks of only 15-50 m in length. Most natural slopes are of higher slope angle than the internal friction angle of the materials and vary from 36° to 83°. The slopes are covered with grass and shrub type vegetation.

4.2 Climate

Bangladesh is located in Southern Asia on the eastern part of the Indian subcontinent (Landsberg 1981). This area is dominated by seasonal changes referred to as four different periods (Lockwood 1974, Landsberg 1981):

- *Winter period*: January – February
- *Hot weather period*: March – May
- *Monsoon period*: June – September
- *Post-monsoon period*: October - December

For the area of the Indian subcontinent and the surrounding oceans July and August are usually the months of most rainfall. During these month two tropical storms or cyclones form on average each year, lasting for 4-5 days; most of them form in the Bay of Bengal (Lockwood 1974). The dominating monsoon wind direction from SW result in high amounts of rainfall over Bangladesh during the whole monsoon period; more than 75 % of the annual rainfall occurs during this period (Lockwood 1974). The eastern part of the country gets the highest amount of rainfall and also the highest number of rainy days (Landsberg 1981). This is probably caused by the presence of the hilly areas of Chittagong and Sylhet due to uplift of humid air and subsequent orographic rainfall.

As visible in Figure 6-1to Figure 6-3 relatively high amounts of rainfall may also occur in May and October. For May this may be explained by an early onset of the monsoon season, as the normal onset for Bangladesh is between 25 May and 5 June. The normal monsoon withdrawal is 10-15 October, explaining the rainfall numbers of October (Lockwood 1974). The post-monsoon period normally get relatively low amounts of rainfall, but cyclones may occur even in this period resulting in rainfall intensities equivalent to the monsoon period (Landsberg 1981); this is demonstrated by rain gauge data from Cox's Bazar located in the far south east of Bangladesh (Table 4-1).

4.2.1 Köppen-Geiger climate classification

The Köppen-Geiger climate classification is a more than 100 years old method that is widely used and regularly modified. Based on a set of precipitation and temperature criteria it uses a large set of rain gauges and climate station, spread around the world, to classify areas into different climate groups. There are 5 different main groups; A - tropical, B - Arid, C - Temperate, D - Cold and E – Polar, each with 30 corresponding subgroups further explained in Peel et al. (2007).

According to Peel et al. the climate of Bangladesh is mainly tropical with some temperate areas in the north and west. The temperate areas are classified as Cwa – main group *C* and subgroup *w* and *a* –an area with dry winters and hot summers. One such area is the area of Sylhet in northwest including the hills in the area. Areas in Bangladesh classified as tropical are classified as either Am – tropical

monsoon climate or Aw – Tropical Savanna. The Chittagong hilly tracts are classified as Am together with an area in the north west of the hilly areas of Sylhet.

Looking at the precipitation data from the areas of Bangladesh where landslides are known to occur (Table 4-2) the hilly areas in southeast and northeast, the southeast piedmont plains and the western plains – there seems to be a stronger correlation between physiography and rainfall amounts (MAP –

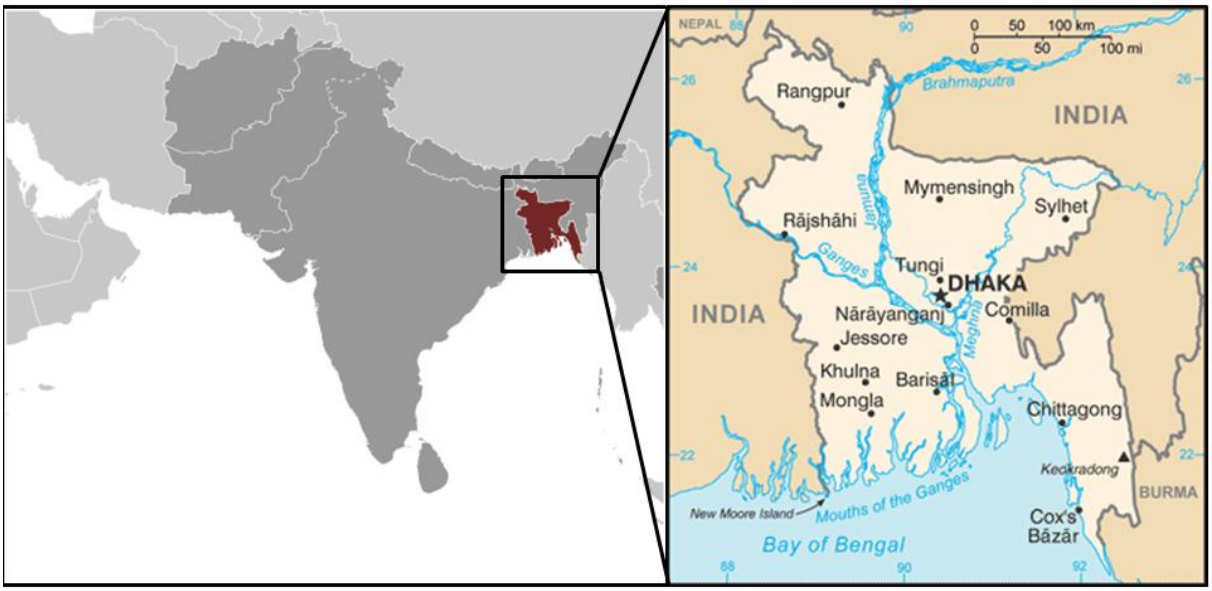


Figure 4-2 Map of Bangladesh and surrounding countries (CIA 2011).

Table 4-1 Rainfall data from Chittagong and Cox's Bazar located in the southeast of Bangladesh, extracted and modified from (Landsberg 1981).

	Rainfall Chittagong		Rainfall Cox's Bazar	
	mean	max 24h	mean	max 24h
Jan	10,40	32,60	10,70	56,10
Feb	7,60	26,70	12,20	55,90
Mar	88,90	152,70	32,30	162,80
Apr	67,80	87,90	80,00	206,00
May	283,70	150,10	292,60	194,80
Jun	569,20	230,60	770,60	279,40
Jul	624,10	417,10	933,40	317,50
Aug	564,60	256,30	780,00	289,80
Sep	305,80	375,40	443,20	238,00
Oct	290,80	389,40	275,10	398,30
Nov	50,00	77,00	63,20	329,40
Des	10,40	6,40	32,80	385,10
	2 873,30	417,10	3 726,10	398,30

Mean Annual Precipitation) than of physiography and the Köppen-Geiger climate classification: The hilly areas receive the highest MAPs and the highest amounts of rainfall during one month, the western plains get least rainfall for MAP and one month average maximum, while the piedmont plains get rainfall values in between. As the Köppen-Geiger climate classification uses temperatures in addition to precipitation, this may account for the level of evaporation and thus the soil conditions preceding a rainfall. Soil conditions are also dependent on vegetation and the corresponding transpiration. Evapotranspiration (evaporation + transpirations) is only partially accounted for in the mentioned climate classification, there is thus considerable uncertainty in how soil conditions in Bangladesh are represented; further study would be needed.

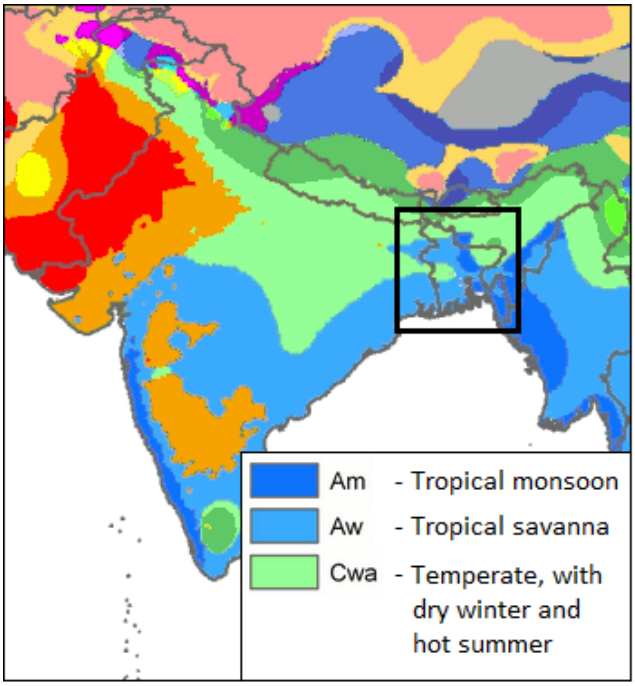


Figure 4-3 Climatology of Southern Asia with Bangladesh in black square and labels for the climate sub-groups of Bangladesh, extracted and modified from (Peel et al. 2007)

Table 4-2 Rainfall values from rain gauges around areas known for occurrence of landslides

Name	Lat.	Lon.	Elev-ation	Series length (yrs)	MAP (mm)	Mean max. month (mm/month)	Corresponding Physiographic Area
SYLHET	24,9	91,9	35	29	4080,48	812,38	Sylhet hill area
SRIMANGAL	24,3	91,7	23	30	2420,33	502,13	SE piedmont plains
COMILLA	23,4	91,2	9	37	2380,16	472,63	SE piedmont plains
JESSORE	23,2	89,2	12	28	1643,86	309,65	W plains
SATKHIRA	22,7	89,1	4	36	1713,08	353,44	W plains
KHULNA	22,8	89,5	3	33	1729,03	350,89	W plains
CHITTAGONG A.	22,4	91,8	14	42	2867,95	746,47	Chittagong hill tracts
COX'S BAZAR	21,4	92	4	38	3753,84	983,61	Chittagong hill tracts

4.3 Social geography and landslide issues

Bangladesh is a relatively small country of only ~130.000 km^2 land cover inhabited by a population of more than 158 million, making it one of the most densely populated areas of the world. The main capital is Dakha with over 14 million inhabitants, Chittagong is the next biggest city populated by almost 5 million an d the main city of the south east and the hilly regions (CIA 2011). The level of poverty is high; as 40 % of the people are estimated to be below the poverty-line (CIA 2011). Combined with the high population density and urbanization, this creates problems related to the need of housing areas. As the pressure on establishing new areas for housing increase, residential areas expand into unsafe areas beneath or on top of steep hills. Illegal, uncritical hill cutting has become common to make room for simple houses, and foothill slums are established (IRIN 2008, Mahmood and Khan 2008). It is claimed that 70,000 people live at risk of landslides due to illegal slope cutting, despite the governments ban of the practice (IRIN 2008, Mahmood and Khan 2008).

4.4 Landslides in Bangladesh

During the last few years more than 300 people have lost their lives in Bangladesh because of landslides. The landslide record of Bangladesh is relatively short; as most landslides are registered during the last 10 years. On average there are registered approximately 2 fatal or potentially fatal landslide events each year in Bangladesh and most of them occurred in the Chittagong hill tracts; this is also where most people have lost their lives. The most severe landslide event in the history of Bangladesh is the landslide event of 11 June 2007 hitting the city of Chittagong and the surrounding areas, causing several landslides at different locations. At least 127 people lost their lives in landslides during this event (Ekram et al. 2007, IFRC 2007).

Most landslides in Bangladesh seem to be triggered by intense rainfall, mostly related to the monsoon season as explained in chapter 4.2. Ali and Kahn (2008) found that in addition to intense rainfall there are two main causes of landslides in the area Chittagong; hill cutting and deforestation. Hills where found to be cut at an angle of 70-80 degrees, causing slopes to fail during prolonged and intense rainfall. As roots from vegetation stabilize loose soil, deforestations also cause increased landslide susceptibility. The investigation also found "*a correlation (…) between the "landslide victims" and the "poverty*"". Several slum areas where found to be situated in landslide prone foothill areas; houses in these areas cannot sustain are not able to withstand mass movements, causing these areas to be very vulnerable to landslides.

Earthquakes may also be a trigger of landslides. The seismic hazard in Bangladesh is caused by the proximity to the seismically active region of the Himalayas in the north (Kamal 2008, Sarker et al. 2010). Strong earthquakes are known to affect Bangladesh, less than 10 events each 100 years (Banglapedia 2011a, USGA 2011). No records were found of landslides occurring close to the dates of earthquake records.

There is a trend of an increasing number of landslide victims in Bangladesh. As the landslide record is short, this cannot be established certainly, even though the most severe landslide events have occurred the last few years. Landslide event of the same magnitude may have occurred earlier, but without any records. If no lives were lost during an event, the event may not have been recorded. There seems to be an increasing trend for keeping records of events the last few years generally

around the world, partly because of higher awareness and increased accessibility to data through the use of computers. Increasing presence of humanity organisations may also result in more records of events. An increasing amount of casualties in landslide events may also be explained by the urbanization and the expansion of residential areas into unsafe areas.

Table 4-3 List of known landslide events in Bangladesh in recent years

Year	Date	Area	Casualties	Sources*
1968		Chittagong hill tracts		Banglapedia
1970		Chittagong hill tracts		Banglapedia
1990	30 May	Chittagong hill tracts		Banglapedia
1997	July	Chittagong hill tracts		Banglapedia
1999	11-13 August	Chittagong hill tracts	17	Banglapedia
2000	24 June	Chittagong hill tracts	13	Banglapedia
2003	5 May	SE piedmont plains	31	NASA
2003	17 June	Chittagong hill tracts	6	NASA
2003	29 June	Chittagong hill tracts	4	NASA
2003	30 July	Hill ranges of Sylhet	6	NASA
2007	11 June	Chittagong hill tracts	127	NASA/GSB[3]/GSB[1]
2007	10 September	SE piedmont plains	2	NASA
2007	15 October	SW plains	3	NASA
2007	19 October	Chittagong hill tracts	Unsure	NASA
2008	3 July	Chittagong hill tracts	9	NASA/ GSB[3]/GSB[2]
2008	6 July	Chittagong hill tracts	4	ADPC/GSB[2]
2008	1 August	Hill ranges of Sylhet		NASA
2008	18 August	Chittagong hill tracts	11	NASA/ GSB[3]
2008	23 August	Chittagong hill tracts		NASA
2009	18 May	Hill ranges of Sylhet	6	NASA
2009	31 July	Chittagong hill tracts	11	NASA
2010	15 June	Chittagong hill tracts	55	GSB[3]/Red Cross
Nr of events:	***22***	***Sum casualties:***	***312***	

**Sources: Banglapedia (2011b), NASA (2011), Red Cross (2010a) and Geological Survey of Bangladesh; GSB1 – Ali et al. (2007), GSB2 - Ali and Khan (2008) and GSB3 (Table A 1 of appendix A).*

Area of landslides in Table 4-3 are given according to the physiography of Bangladesh (see Figure 4-2) based on location names or coordinates given in the different sources. For some of the landslide events listed in Banglapedia the dates are unknown, and casualties are for many events either not mentioned or unknown. For the NASA data casualties and location are based on news-reports and may thus not be accurate.

5 METHODOLOGY

The aim of this work is to develop a methodology for use of high-resolution global satellite precipitation estimates to establish rainfall thresholds in regions with no or sparse ground-based precipitation data coverage. A methodology for using NASAs TMPA-RT rainfall products on a global threshold has been proposed by Hong et al. (2006), and are currently used in the TRMM *Current Heavy Rainfall, Flood and Landslide Estimates* (http://trmm.gsfc.nasa.gov/publications_dir/potential_flood_hydro.html). The accuracy of thresholds tends to decrease with increasing spatial coverage. For a single country like Bangladesh, a small-scale regional threshold (or possibly local threshold) will be appropriate to increase the predictability of the threshold. Such a threshold will be developed using the same TMPA-RT data of global coverage.

There are several SPEs of global coverage available today. The SPE rainfall products to be used in this study had to meet certain criteria as presented in section 3.3.3 - Available SPEs. These rainfall products are stored in digital data files that may be accessed and handled using common computer languages like i.e. Matlab, R, Octave or FORTRAN. Handling of binary data may also be required for some rainfall products. The software Matlab was used in this study, including the Matlab Statistical Toolbox. The methodology used in this study is thus based on the programming language of Matlab, but application of the methodology using other software and programming language should be possible developing new scripts.

An analysis including all SPE-products would be preferable, but due to time limitations of the work on this thesis this was not possible to achieve: Most rainfall products are created by different organisations and institutions, using different data structures (in additions to different methods and algorithms). As a result separate scripts have to be developed to access and analyse each rainfall products data files. This is a time consuming process, in addition the computing time may also be very long. For instance, accessing and analysing 5 years of rainfall data when each file consist rainfall data for the duration of between 30 minutes and 3 hours require reading and handling of between 14600 and 87600 files. As each file consist of rainfall data for each point in a global grid the amount of data are huge; the smallest TMPA-RT data consist of a 1440 x 480 points grid of rainfall data (~700,000 point) covering most of the world. The NOAA H-E data, as another example, consist of a global grid of 8001 x 3111 data points (~25 million points of rainfall data) (STAR 2011); these files of hourly data take more than 150 Mb each when unzipped also creating storage issues, hence the data has to be handled one rainfall file at the time. Because of this limitation only two quite similar rainfall products were selected for the study; NASAs TMPA-RT products 3B41RT and 3B42RT.

5.1 Landslide and rain gauge data

The initial landslide and rainfall data used for this thesis where provided by Mr. Reshad Ekram, the director of the Geological Survey of Bangladesh (GSB), and the Asian Disaster and Preparedness Centre (ADPC) through a collaboration project between ADPC and the Norwegian Geotechnical Institute (NGI). These presented location and details about the most recent landslide events in Bangladesh and eight days of daily rainfall data related to each event. No coordinates were given for the landslides or rain gauges, something that would be needed for using corresponding satellite precipitation estimates. Coordinates of the landslides were found based on area names in the

landslide event descriptions, and the rain gauge coordinates were extracted from attachments in Peel et al. (2007). Additional rain gauge data where provided by GDB; these contained daily rainfall data for the area of Chittagong city during the period 1 June 1996 to 31 December 2008. The most relevant parts of these data are presented in Appendix A.

5.1.1 Landslide inventory

Additional landslide data was gathered from Banglapedia , landslide reports from GSB (Ekram et al. 2007, Ekram and Khan 2008), a global landslide catalogue (NASA 2011) and by own searches for new articles and online reports. Banglapedia (2011b) were found to describe some historical landslides in Bangladesh, some including quite specific location names. The GSB reports included assessment of the fatal landslide events of Chittagong 2007 and Teknaf 2008, and included accurate coordinate position of individual slides as well as geotechnical investigations related to these. The landslide catalogue is compiled based on the gathering of landslide data from online news media and different hazard databases, as described by Kirschbaum et al. (2010). Due to the original source of this data, the exact location of the events the landslide coordinates were given including a confidence radius. Many landslides related news articles and disaster reports were found for Bangladesh, but covered mostly landslides available in the other sources. One new landslide position was found, but this was related to a landslide event covered by the other sources.

All landslide data were gathered in a single inventory. The position of each landslide was described using location name and the corresponding district and region as well as latitude and longitude decimal degrees. Missing coordinates were estimated finding for the location names in Google Maps and Google Earth. The highest accuracy of position would not be required, due to the spatial resolution of the satellite rainfall estimated to be applied in the following analysis. All positions where given an estimated confidence radius for possibilities of numerical evaluation of the position accuracy. The data samples were labelled as single landslides (single LS) and events of multiple landslides (LS events), as some data were site and landslide specific and others were related to the occurrence of a landslide triggering rainfall event. One example; Ekram et al. (2007) reports 19 landslides including their coordinates from the 11 June 2007 landslides in Chittagong, the ADPC data and the global landslide catalogue referred to this as one landslide event in Chittagong, providing one coordinate. Landslide type was often not described in the available data, but where available this was included as an inventory data classifier, as the landslide type may be important for the triggering condition. Dates of landslide occurrence were added to be able to correlate landslide and rainfall data. The full inventory is presented in Table A 3 in appendixes.

5.1.2 The rainfall data

Use of rain gauge data for analyses purposes was limited in duration and extent. The one rain gauge series of long duration, available for Chittagong, would represent a possibility to do some assessments for the area of Chittagong. The rainfall data for this thesis was thus mainly based on satellite based rainfall estimates; TMPA-RT data. Accessing and using the data would be a quite complex process, and will be explained in the following chapter 5.2.

5.2 TMPA-RT data

NASA's Tropical Rainfall Measuring Mission (TRMM) is a satellite-based measuring campaign collecting near-real-time high-resolution quasi-global data for different scientific purposes. One of the data provided is the TRMM-based rainfall estimates, Real-Time TRMM Multi-Satellite Precipitation Analysis (TMPA-RT). TMPA products are available at a resolution of 0.25° x 0.25° in a global grid extending over the latitude bands 60° N to 60° S. The temporal resolution is 1 or 3 hours, depending on the product type. Each global rainfall estimate is stored in digital files and may be accessed from a FTP-server. TMPA uses infra-red (IR) and passive microwave (PMW) satellite images; from the international constellation of geosynchronous (GEO) satellites and low-earth-orbit (LEO) satellites respectively. The limited extent of latitude bands of TMPA was chosen as IR and PMW-data tend to lose skill at higher latitudes (Huffman et al. 2009).

The formal name of the TMPA-RT data is "Version 6 TRMM Real-Time Multi-Satellite Precipitation Analysis." (Huffman and Bolvin 2010), and three different products are available; combined PMW estimates at 3 hour resolution (3B40RT), PMW-calibrated IR estimates at 1 hour resolution (3B41RT) and merged PMW and IR estimates at 3 hour resolution (3B42RT). The spatial resolution was chosen for the grid size to be somewhat larger than the "footprint" of the passive microwave precipitation estimates (Huffman et al. 2009). Temporal resolution was chosen from a combination of satellite coverage frequencies and fitting to the diurnal circle.

Table 5-1 Basic information for the different TMPA-RT data sets

	3B40RT	*3B41RT*	*3B42RT*
Source type	Micro wave	IR	MW/IR-combined
Spatial resolution	0.25° x 0.25°	0.25° x 0.25°	0.25° x 0.25°
Temporal resolution	3 h	1 h	3 h
Global grid size	1440 x 720	1440 x 480	1440 x 480
Coverage	90° N-S, 0-360° E	60° N-S, 0-360° E	60° N-S, 0-360° E
First grid-box centre	0.125° x 89.875°	0.125° x 59.875°	0.125° x 59.875°

In addition to the real time (RT)-products, gauge adjusted TMPA products are available, hereby referred to as TMPA-GA or GA products. These are rescaled to monthly rain gauge analyses and are thus not available in real time (Huffman et al. 2007). An increased performance of the TMPA-GA data is confirmed e.g. by Su et al. (2008) and Shen et al. (2010). TMPA-RT estimates have been operational since late January 2002, and the current version (V6) running since March 2005 and the lasts update since October 2008. Versions are updated occasionally as results improvements in e.g. data processing techniques, from assessment of performance data and experiences from using the data. Data processing errors may also occur and are corrected successively as they are discovered, and listed in the TMPA-RT data set documentation (Huffman and Bolvin 2010). Improved performance in the new version is verified, e.g. for the GA product 3B42 over Bangladesh (Islam and Uyeda 2008) and over Thailand (Chokngamwong and Chiu 2008).

5.2.1 Selecting TMPA-RT product types

A mentioned in chapter 3.3.3, three different TMPA-RT products are available; 3B40RT (PMW data), 3B41RT (IR data) and 3B42RT (merged IR and PMW data). 3B40RT is limited to 3 hours accumulated rainfall due to the lower temporal resolution of PMW-data. 3B41RT provides the best temporal coverage at intervals of 1 hour. The 3B42RT product provides the best spatial data coverage of the TMPA-RT products, by merging IR and PMW data.

3B41RT and 3B42RT were selected for this study, as these represent the best high temporal resolution and the best spatial resolution respectively. Both features believed to represent a potential advantage in landslide studies; successful application of hourly data would be an advantage for landslide studies by picking up rainfall intensity peaks, PMW-data are believed to predict rainfall more accurately and combining these with IR-data increase the satellite coverage of an area. The merged TMPA product also seems to be most widely studied TMPA product (both the RT and the GA product), e.g. by (Chokngamwong and Chiu 2008, Islam and Uyeda 2008, Shen et al. 2010, Yong et al. 2010) - most studies are however of the gauge adjusted TMPA version.

5.2.2 File names and observation periods

The TMPA-RT digital data files are given name according to rainfall product, and date and time of the measurements; e.g. *3B42RT.2010010115.bin.gz,* where the first part represent the TMPA product type, *2010010115* represent time and date (data format 'yyyymmddHH' – 4 digit year, 2 digit month, 2 digit day and 2 digit hours). *bin.gz* represent the file format, indicating these files are binary files compressed in the format zipped. V6 files are indicated with ".6" before the file extensions.

The time and date stamp of each file indicates the time of satellite observations used as basis for the estimated rainfall data, each at interval of 1 and 3 hours for 3B41RT and 3B42RT respectively and is set in UTC (Coordinated Universal Time). It is important to note this is the time and date denotes the mean time of the satellite observational period related to each file; this is called the nominal time (Huffman and Bolvin 2010). The file name above, given as an example would thus represent the rainfall at 10 January 2010 from 13:30 to 16:30.

5.2.3 Structure and format of files and data

Understanding the file structure and format is important to be able to extract the correct data and correct values. Without guidance this may be especially hard when dealing with binary data and/or complex file structures. All TMPA-RT files are stored in binary format: The first part of each file is a header record of ASCII characters followed by blocks of gridded 2-byte integer precipitation estimates at grid (as presented in Table 5-2). The header is 2880 bytes, the same size as one row of the 2-byte integer data. Each data point of the gridded data represents one grid cell (pixel) of a global latitude/longitude grid of 1440x480 cells. The data points increment to the east from the prime meridian and then to the south, as demonstrated by Table 5-3. The rainfall estimates in the TMPA-RT files are in mm/hr, but scaled by hundred; this scaling must be accounted for when extracting the data (Huffman and Bolvin, 2010).

As listed in Table 5-2, the TMPA-RT data files also contain other blocks of gridded data; error estimates, satellite source and uncalibrated precipitation data. The most interesting of these data would be the gridded precipitation estimate error data (*error*, Table 5-2). The precipitation error data were examined for a random selection of files, as only "no data" cells where found these data where not used in further analyses. Information on all data registered in TMPA-RT files may be found in Huffman and Bolvin (2010).

Table 5-2 Overview of data structure in the TMPA-RT files, adapted from (Huffman and Bolvin 2010)

	3B41RT		*3B42RT*	
Block	*Byte Count*	*Field*	*Byte Count*	*Field*
1	2880	header	2880	header
2	1382400	precipitation	1382400	precipitation
3	1382400	error	1382400	error
4	691200	number of pixels	691200	source
5			1382400	uncal. precipitation

Table 5-3 Structure of the 1440x480 gridded data in the TMPA-RT data files. Each value represents one grid box centre of a latitude/longitude semi-global grid, adapted from (Huffman and Bolvin 2010)

	3B41RT		*3B42RT*	
Location	Latitude	Longitude	Latitude	Longitude
1st point centre	59.875ºN	0.125ºE	59.875ºN	0.125ºE
2nd point centre	59.875ºN	0.375ºE	59.875ºN	0.375ºE
Last point centre	59.875ºN	0.125ºW	59.875ºN	0.125ºW

5.3 Creating a TMPA-RT rainfall inventory

5.3.1 Acquiring the rainfall data

To acquire the different TMPA-RT rainfall product data several Matlab-scripts where developed (see 0). The scripts were developed to work for both TMPA-RT rainfall products. First all 3B41RT- and 3B42RT-files where downloaded from the server onto a local hard disk drive (HDD), in a folder-structure sorted by year and month. The folder structure was chosen to assure a limited number of files in each folder, as this proved to increase the Matlab-scripts rainfall data processing speed. The complete set of TMPA-RT data included 3B41RT- and 3B42RT files from 1 March 2005 until 9 April 2011, comprising a total of almost 73000 zipped files of global rainfall data and 11 GB of storage space.

The rainfall inventory was to include rainfall data from the whole period of available TMPA-RT data. To limit the amount of data and simplify further data analysis, only TMPA rainfall data corresponding to known landslide positions where extracted. This was achieved creating a TMPA-coordinate grid

matrix in Matlab according to the coordinates described in Table 5-3, where each position represented a grid box centre. Using a numerical search for all landslides positions in the landslide inventory, the closest TMPA grid position was found. Using these positions, TMPA rainfall data corresponding to all landslide position could be extracted from the global gridded rainfall data sets. Only one grid where selected for each file, not discriminating between if a landslide where represented in the corner of a TMPA grid box or in the centre. The full process was acquired through running a series of Matlab-scripts, which are presented in 0. The basic principles of the process can be described like this:

1. Insert input data (product name, landslide inventory file name, time zone, etc.).
2. Load landslide inventory.
3. Locate TMPA-RT global grid positions representing surface position of landslides in landslide inventory.
   i) Create 1440x480 global grid with all TMPA-RT grid box centre positions of the same grid size and configuration as the rainfall data in the TMPA-RT rainfall files.
   ii) Locate grid boxes centre positions which are closest the landslide positions in the landslide inventory, and save these positions.
4. Create vector containing filenames of all files to be processed for extraction of rainfall data.
   i) Unzip file one file at a time.
   ii) Read and save header info as text string and rainfall data as numerical data.
   iii) Extract rainfall data from landslide positions located by step 3.ii, and save data into one variable that is used for all files, such creating one continuous rainfall data series for each TMPA position.
   iv) Delete unzipped file after processing, to avoid storage space issues.
5. Save position data and rainfall data series into M-file (Matlab-file used for storage).
6. Create a series of daily accumulated rainfall data from each TMPA-RT product and save in M-file, accounting for local time and adjusting for when daily rain gauge data are gathered in the area.

Figure 5-1 illustrates the principle of TMPA data extraction from a simplified grid. The small dots represent a known landslide position, and the black squares a TMPA grid box of rainfall estimates representing the landslide position. For each file data from the known positions are extracted into a line in a rainfall data series, where the line represents the time of rainfall. Data from each position are stored in different columns, creating one rainfall series for each position. The hourly/3-hourly data was accumulated daily into a final series of daily data.

5.4 Rain gauge- and TMPA-RT data: a comparison

The reliability of an established threshold depends highly on the quality of the data used for the analysis. Satellite based precipitation estimates (SPEs) are associated with some range of uncertainty as the precipitation are measured indirectly using IR and PMW imagery. The method is believed to make fairly good estimates, especially for intense rainstorms of long durations. A limitation is the lack of spatial and temporal coverage; uncertainties exist in the methods ability to represent the small scale spatial variations in rainfall (Shen et al. 2010). Rain gauge data take these local conditions into account measuring the exact amount of rainfall reaching the ground. Still rain gauges are neither a

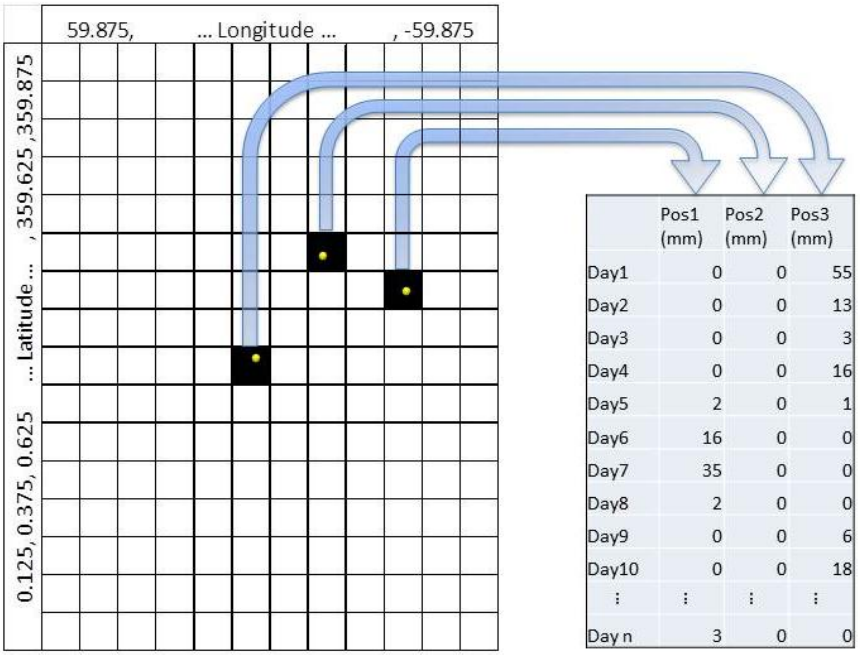


Figure 5-1 - Principal of TMPA-RT rainfall data extraction from global grid rainfall files (grid is simplified and the data are not actual data).

good method for landslide studies, as the limitations in spatial coverage is significant. Unless a rain gauge is present in almost the exact location of a triggered landslide, the small scale spatial variations of rainfall amount and intensity may cause uncertainty in rain gauge measurements as well.

Despite the uncertainties in both SPE and gauge data, regarding spatial resolution and coverage, a comparison of these rainfall data sources will be necessary to give an impression of the SPEs accuracy in the geographical area of study. As a SPE grid cells cover some 10s of square kilometres (a TMPA-RT grid cell cover 0.25° x 0.25° equal to more than 25 x 25 km), rainfall in this area are expected to be heterogeneous. A good comparison would thus require long time series of continuous precipitation data compared for several rain gauges located inside the same SPE-grid cell. The available rain gauge data for this study were limited, and the length of time series from SPEs are limited; such thorough correlation was thus not possible. The accuracy of the SPE data for the Bangladesh area thus had to be based on a less complex data correlation, combined with earlier studies of similar climatological areas.

5.4.1 Comparing the rainfall data sources on local scale

The rainfall data used for comparison against the satellite estimated rainfall data contained of several years of data, including some years of overlap with the satellite based rainfall products. As rainfall data may vary annually, an overlap of some years were important for the statistical significance of the methods to be used. Both of NASAs satellite estimates were included in the comparison, only using data from the grid cell corresponding to the position of the rain gauge. Accumulated daily TMPA-RT data were used, as this corresponds to the temporal resolution of the rain gauge data.

The TMPA-RT data and rain gauge data mentioned above were compared and analysed using exploratory data analysis (EDA). EDA is an approach for analysing and characterizing data. This includes several techniques used to describe and compare data in order to get information about e.g. similarity, correlation and data outliers. EDA is a good way of presenting and visualising data as many EDA techniques are graphical; scatter plots for correlation, histograms, distribution curves, etc. (NIST/SEMATECH 2011). Monthly rainfall, daily rainfall, number of rainy days, rainfall distribution and correlation of the different data were assessed; the analyses where computed using Matlab-scripts as presented in Appendix C. The graphical analyses were supported by evaluation numerical variables like mean, standard deviation and correlation coefficients.

As SPE-data relies in the satellite coverage in the area of study, a final key element to the TMPA-data was to which extent rainfall data were missing; the extent of missing TMPA-data were evaluated and rainfall data at these time examined. The best overall performing TMPA-RT rainfall product where selected for the threshold study.

It was expected to demonstrate how the TMPA-RT data performed on a local scale compared to the gauge data. At the same time, due to the limited amount of rain gauge data, this would hopefully also reflect any problems related to the assessment of rainfall thresholds by using TMPA-RT data in areas with inadequate coverage of gauge data.

5.4.2 Previous studies of TMPA and gauge data

NASAs TMPA precipitation products are widely studied. Many studies investigates different SPEs ability to reproduce the rainfall observed in rain gauge networks (Sapiano and Arkin 2009, Shen et al. 2010, Sohn et al. 2010, Yong et al. 2010), studies are conducted at different spatial scale. Some have been related to e.g. studies of global landslide hazard assessment (Hong et al. 2006) and hydrological modelling of stream flow (Yong et al. 2010). Of TMPA data especially the products 3B42RT and 3B42 have been used, as these use merged IR and passive microwave data.

Different studies have shown inconsistent results regarding correlation of TMPA rainfall products and rain gauge data: It has been found that the gauge adjusted rainfall product generally performs better than the RT-product (Shen et al. 2010, Yong et al. 2010, Behrangi et al. 2011). Improvements in estimates when adjusted to local data is expected, but a study in the Korean Peninsula by Sohn at al. (2010) found surprisingly found smaller RMS errors for the RT product. Different studies have found that the TMPA products both underestimate (Shen et al. 2010) and overestimate (Sohn et al. 2010) rainfall compared to gauge data. Yong et al. (2010) found that especially rainfall of intensities above 30 mm/day were overestimated.

Yong et al. (2010) also found systematic errors for TMPA data compared to gauge data dependent on latitude and altitude, where the error rates decreased with decreasing latitudes and altitudes. Sohn et al. found that the TMPA data showed similar patterns to the climatology of the studied area. These discoveries indicate that contradictory results between studies of TMPA and gauge data may be caused by a different TMPA performance for different climate types and a performance changing with latitude and altitude. It may be concluded that the TMPA data (RT and gauge adjusted products) perform differently in different areas, suggesting that application of data require adjustments to local conditions. This may impede reliable application of these data for non-instrumented regions.

Conversely, climatology dependent data patterns found by Sohn et al. (2010) and recent discoveries that estimates of uncertainty may be transferred from instrumented to non-instrumented areas (Tang et al. 2010), do indicate that improvements may be expected in the future in this matter; although it was so far concluded that transfer show high error rates event at low temporal scale. Theoretically it was found applicable dawn to daily scale.

5.5 Establishing thresholds

There are many methods available for establishing a threshold, and as described in chapter 0 the type of threshold is also depending on what kind of data are available: To establish a rainfall threshold it is requires some sort of rainfall series. The most commonly used threshold type, I-D thresholds evaluate the relationship between rainfall intensity and duration, requiring continuous rainfall series for a number of different rainfall durations. Other thresholds consider different kind of storm event data, like total storm rainfall or storm duration and intensity. For this study, the focus is on using continuous rainfall series of different duration, like for the I-D thresholds.

A landslide inventory was created for this study, thus combining landslide data and rainfall data both triggering and non-triggering conditions could be established. Obtaining a threshold for these data would then require establishing a criterion for separating these two types of data. Such criteria were earlier established estimating the best fit of a threshold equation purely graphically from an I-D plot. In recent years such threshold criteria has been established using different statistical methods like Discriminant Analysis, e.g. Jakob and Weatherly (2003), and Neural Networks (Lee et al. 2003). These multivariate exploratory techniques are suitable for handling multiple factors (StatSoft Inc 2011), and are also commonly used in other landslide and hydrology related areas; e.g. for evaluation of landslide susceptibility and hazard (Baeza and Corominas 2001, Lee et al. 2003, Santacana et al. 2003, Guzzetti et al. 2006, Baeza et al. 2010) and for hydrology studies related to correlation of satellite estimates and rain gauge data (Bellerby et al. 2000, Rossi et al. 2010).

Statistical methods for establishing thresholds would also be tried and evaluated for the rainfall and landslide data in Bangladesh. Two multivariate statistical methods will be used: The widely adopted method discriminant analysis, e.g. (Jakob and Weatherly 2003, Cepeda et al. 2009, Rossi et al. 2010), will be used with the purpose of apply a second rainfall variable I-D threshold as by (Cepeda et al. 2009). One method that cannot be found to have been used in threshold studies will also be attempted used, as it proves to be a relatively easy method for assessing multiple variables; classification tree analysis. Correlation coefficients will be considered when combining variables, and a simple form of ROC application will be used for evaluating results numerically.

5.5.1 Data preparation

5.5.1.1 Antecedent precipitation - predictors and outcome variable

The first step of creating the thresholds was to prepare the rainfall inventory for a threshold analysis. Accumulated rainfall of 13 different durations (1d, 2d, 3d, 4d, 5d, 6d, 7d, 9d, 11d, 13d, 15d, 30d and 60 days) were calculated for the whole rainfall series acquired for the TMPA-RT data. The values of

accumulated rainfall (antecedent rainfall) were given in rainfall intensity per day, calculated according to

$$A_{D,i} = \frac{\sum_{i-(D-1)}^{i} RF_i}{D} \quad (3)$$

where *i* is a day in the rainfall inventory, *RF* is the rainfall of day *i* in mm and $A_{D,i}$ is the antecedent rainfall of duration *D* (including day *i*) where *D* = [1d, 2d, 3d, 4d, 5d, 6d, 7d, 9d, 11d, 13d, 15d, 30d, 60d]. Consequently, denomination of A_D is mm/day.

A_D was calculated for each day throughout the rainfall inventory resulting in an almost 4 year's long rainfall series including antecedent rainfall of 13 different durations - 13 predictors to use in the threshold analysis. As antecedent rainfall could not be calculated for the first D number of days, these days where given "no actual value", NaN, as presented in Table A 5 in appendixes.

In addition to dates and corresponding values of 13 predictors, an outcome variable was added to the inventory for threshold analysis. This variable was set to '*y*' for days landslide(s) occurred and '*n*' for when no landslides occurred. The outcome variable was dependent on both date and location; similar inventories were prepared for each TMPA grid position of extracted rainfall estimates. A part of this inventory is presented in appendix A, Table A 5, for illustration of the inventory structure.

Table 5-4 Predictors used in this study

Predictor	1	2	3	4	5	6	7	8	9	10	11	12	13
Duration	1d	2d	3d	4d	5d	6d	7d	9d	11d	13d	15d	30d	60d
Variable	A_{1d}	A_{2d}	A_{3d}	A_{4d}	A_{5d}	A_{6d}	A_{7d}	A_{9d}	A_{11d}	A_{13d}	A_{15d}	A_{30d}	A_{60d}

The NaN data of rainfall accumulation were removed from the data set as the discriminant analysis function in Matlab `classify` only deals with existing data. All first 60 days of the threshold inventory were removed, but should not represent a problem as these 60 days represent the months January and February of 2007. As explained in chapter 4.2 these months are characterized as a dry period with only minor amounts of rainfall, besides has no landslides been found to occur in Bangladesh during these months.

5.5.1.2 Predictor correlation

The data set of different variables of antecedent rainfall (predictors), prepared as explained above in chapter 5.5.1.1, were to be applied in different statistical methods where high variable dependency would be undesirable. The goal of the different methods was to establish rainfall thresholds based on at least two different rainfall variables; establishing a threshold using two variable that are highly dependent, would not necessarily differ considerably from using one of the variables. For such a combination of rainfall variables to be effective, their dependency must be limited. This may also be reflected by an example: For describing the rainfall occurring during a period of rainfall using two

rainfall variables a combination of 5- and 6-days accumulation of rainfall would not necessarily be suitable, as these often tend to be quite similar. Using e.g. 1- and 5-days accumulated rainfall would on the other hand make more sense, as the daily data would pick up the daily extremes and the 5d data would represent a measure of the short time average rainfall if measured in mm/day.

For assessing the dependency between the different predictors, an analysis of correlation was performed using the function `corr` in Matlab; this function combines (a set of) two different variables at a time using the correlation method of Pearson's correlation coefficient as a default (The Mathworks Inc. 2011). Pearson's correlation coefficient is defined as the covariance of two different variables, X and Y, divided by the product of their standard deviation.

Calculated for all different combinations of predictors, the method produce a correlation matrix, as presented in Table 6-6. The correlation coefficient ρ is a number between 1 and 0 as an output for each combination of variables X and Y (here: combination of predictors), where 1 equals a perfect correlation and 0 equals no correlation. A high correlation was for this study defined as $\rho \geq 0.8$. Highly correlated variables are as mentioned not desirable, and predictors of high correlation would not be combined in a threshold equation.

5.5.2 Classification tree analysis

Classification tree analysis (CTA) is a method used on one or more sets of data (predictor variables) to evaluate their membership to different classes (StatSoft Inc 2011). Such classes will in this study be related to the outcome classes "*y*" and "*n".* A classification tree analysis is searching to find the predictors, and their values, that separate the classes in the best manner from a clustered set of data. This may be comparable with the search of a best fit condition between two or more parameters, representing the same principles as establishing a threshold. CTA may be very simple to perform, given the appropriate computational- or statistical software, as i.e. Matlab or programs developed for the purpose of CTA only. In Matlab the CTA is performed very easily using the function `classregtree`, which then include all predictors in the process of finding the best way for separating the classes testing one at a time.

CTA is a hierarchy based system where several criteria for separation may be combined. One predictor is assessed at the time to find the best linear criterion for separating data between two classes. Ideally this separation would result in two data clusters; each with membership of all data points to only one class of data. The equivalent to this study would be to separate all landslide triggering conditions from the non-triggering conditions. Such a separation would be a 100 % success. As clustered data are usually not this easy to separate CTA search to find the best criterion for spitting the data into two groups; where most landslide triggering conditions are in one group and most non-triggering conditions are in the other. The best criterion (a value of one of the predictors) is saved in the classification tree.

The split data, represented by two new clusters of data, are visualized in the classification tree as two new nodes branched out from the original data (which is also visualized as a node). The two new nodes make out a new level in the classification tree. Each group at this level is then assessed again, one predictor at the time, to find a new separation criterion for each. This separation may be

continued until each group contain only one data class; often resulting in many branch levels and a complex three (see Figure 5-2).

For the CTA method used in this thesis the measurements are on ordinal scale, meaning that the criteria are defined by if the data are "higher than" or "lower than" a certain value; the linear splitting criteria are thus represented by a constant (StatSoft Inc 2011). This way the classes represent "landslide triggering" and "non-triggering" data, and the splitting criterion would be a horizontal or vertical line.
A group on a lower branch-level is a result of all splits above, and is then defined by the combination of all the above separation criteria. This way, a group of data at level *n* would be defined by the criteria C_1, C_2, C_3, …, C_n, where *C* is the split criteria. Each criteria may be defined by the value of any which predictor (P_1, P_2, P_3, …, P_i) providing the best data separation.

Despite that the method of CTA is very simple to use, it is not found to be widely used in landslide studies. Some have used the method for landslide studies in recent years (Wan and Lei 2009, Wan et al. 2010, Yeon et al. 2010), but not for establishing thresholds for triggering of landslides. There are no explanation for this found in the literature, only that it is widely used in fields like medicine and computer sciences. This may though be because more stringent data are believed to require more traditional methods like regression or discriminate analysis (StatSoft Inc 2011).

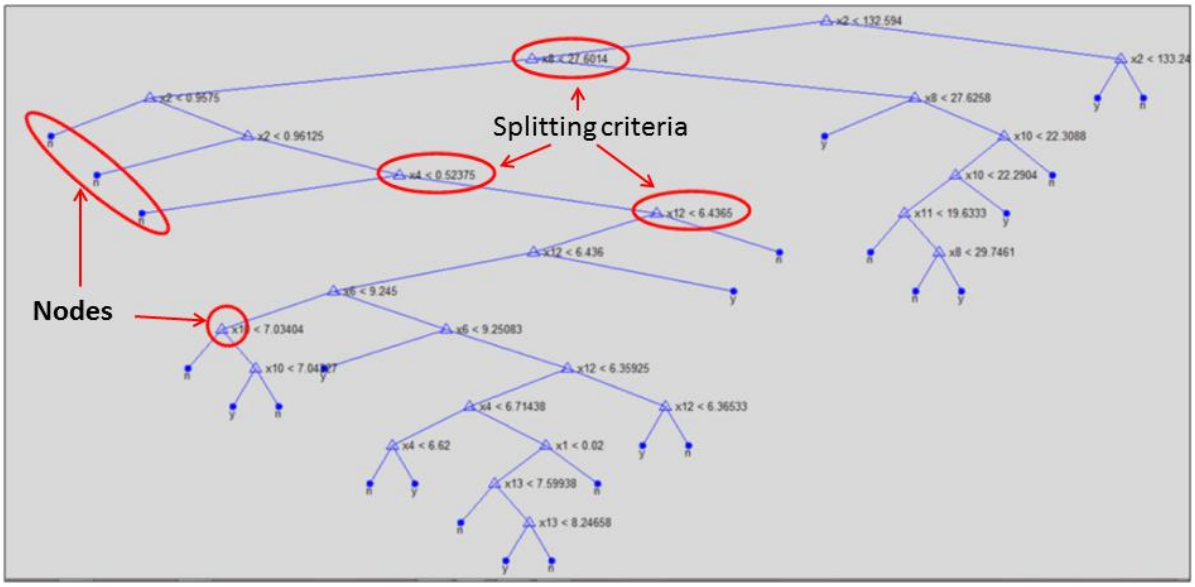


Figure 5-2 Example of a classification tree.

5.5.2.1 CTA: Carrying out the analysis

Classification tree analysis (CTA) was used to analyse data from the inventory prepared for the threshold analysis (see Table A 5). As CTA could easily be carried out using the full set of predictors, the first CTA was conducted without selecting specific predictors based on the previously performed predictor correlation (chapter 5.5.1.2). Each splitting criteria resulting from the CTA would represent a potential threshold variable. *N* different criteria using *N* different predictors would result in *N* variables for the final threshold; e.g. C_1, C_2, C_3 comprising the predictors P_i, P_j, P_k would result in a threshold of the variables V_1, V_2, V_3. If some criteria consist of values from the same predictor this

would only change the resulting threshold variable, resulting in fewer threshold values; e.g. C_1, C_2, C_3 comprising the predictors P_{i1}, P_j, P_{i2} would result in a threshold of the variables V_1 and V_2, where the value of V_1 is related to P_{i2} and V_2 to P_j.

Thresholds where established one classification tree level at a time; the best performing node at each level where selected and the corresponding splitting criteria (C_i) where applied in a threshold. This way, first, a single variable threshold would be established, then a two variable threshold using nodes from two branch levels, then a threshold utilizing three nodes, etc. Each nodes - and corresponding criterions - performance where checked using *Class membership* in the graphical interphase *Classification tree Viewer* in Matlab; on this basis the best performing criteria of each branch level where selected. Each established threshold were evaluated regarding their performance, assessed by calculating and evaluating parameters used in Receiver Operating Characteristics, as will be explained later.

The classification tree analyses where performed in three stages to investigate how predictor correlation affected the results. The first CTA were performed combining all predictors, without considering correlation coefficients. In stage two, correlation coefficients were partially considered; based on one predictor all highly correlated predictors where left out of the CTA. Several stage two CTAs were conducted. In the final stage correlation coefficients between all predictors were taken into account, making sure no predictors used for establishing a threshold were highly correlated. It was assumed that the best threshold would only consist of predictors of $\rho < 0.8$.

5.5.2.2 CTA: Testing threshold performance

The performance of a threshold is highly related to how it would perform if implemented in an early warning system (EWS). It is crucial for an EWS for landslide mitigation purposes that landslide triggering conditions are discovered and that the number of false alarms is limited. Such variables may be assessed by calculating and evaluating parameters used in Receiver Operating Characteristics (ROC). ROC are usually applied using ROC curves to assess the rate between successfully predicted triggering conditions and false alarms (Fawcett 2006), and has earlier been used successfully for optimization of rainfall thresholds (Cepeda 2009).

The principle of threshold optimisation for separating to groups of data becomes clear when visualized as in a figure by Beguería (2006): Figure 5-4 shows the optimized criteria (in this example a score) for a threshold between negative and positive observations of equal groups (a) and unequal groups (b); the optimization search to put most positive observations above the threshold and most negative observations below the threshold. The positive observations above the threshold are referred to as *true positives* (a) while the negative values above are *false positives* (b). Negative observations below the threshold are *true negatives* (d) and the positive observations below are *false negatives* (c). False positives are in common statistics literature called error type I and false negatives, error type II (Beguería 2006).

Relations between these four different data types, true positives (TP), false positives (FP), false negatives (FN) and true negatives (TN), may be used to test the threshold performance. These relations are evaluated using equations as presented by Fawcett (2006) using a confusion matrix

(Figure 5-3), where true class represent the actual class of an observation and the hypothesized class represent the classification of an observation based on if it is above or below the established threshold.
Other parameters that may be explained from the confusion matrix are; *P* - the total number of true class positive in the data set (TP + FN), *N* – the total number of true class negatives in the data set (FP + TN), *false positive rate* (FPR) – the rate of false positives to the total of true class negative values (fp rate) and *true positive rate* (TPR) – the rate of false negatives to the total of true class positive values (tp rate) (Beguería 2006, Fawcett 2006). The following relations may also be established using the confusion matrix; FN = P – TP and FP = N – TN.

The Roc curve is defined by plotting the TPR (y-axis) by the FPR (x-axis) with the result that the top left position in the plot (1, 0) represents a perfect classification (Fawcett 2006). This way, the best performing threshold would be the data plot closest to (1, 0) in the ROC space. See Figure 6-7 for an example of a ROC curve.

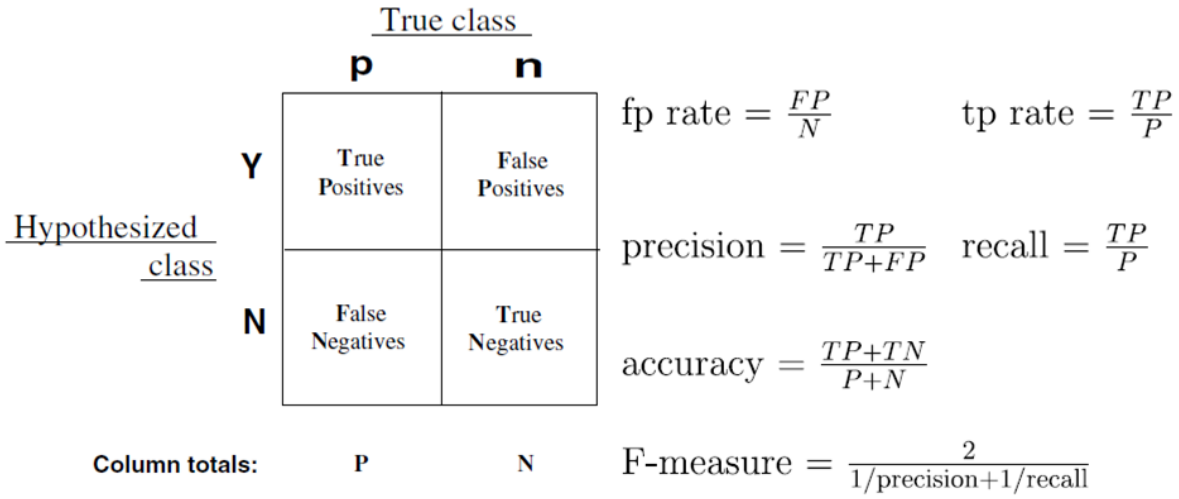


Figure 5-3 - Confusion matrix and variables for assessing threshold performance (Fawcett 2006).

For application of the ROC curve in an early warning system for rainfall induced landslides, the variables of the confusion matrix would represent the following:

- TP - rainfall condition correctly predicted to trigger one or more landslides
- FP - rainfall condition predicted to trigger landslide(s) when no landslides occurred (potential false alarm)
- FN - rainfall condition where no landslide is predicted, but landslide occurs (missed landslide event)
- TN - rainfall condition correctly predicted not to trigger any landslides

This way the ROC curve demonstrates the rate of predicted landslides (TPR) by the rate of false alarms (FPR), important factors for EWS's.

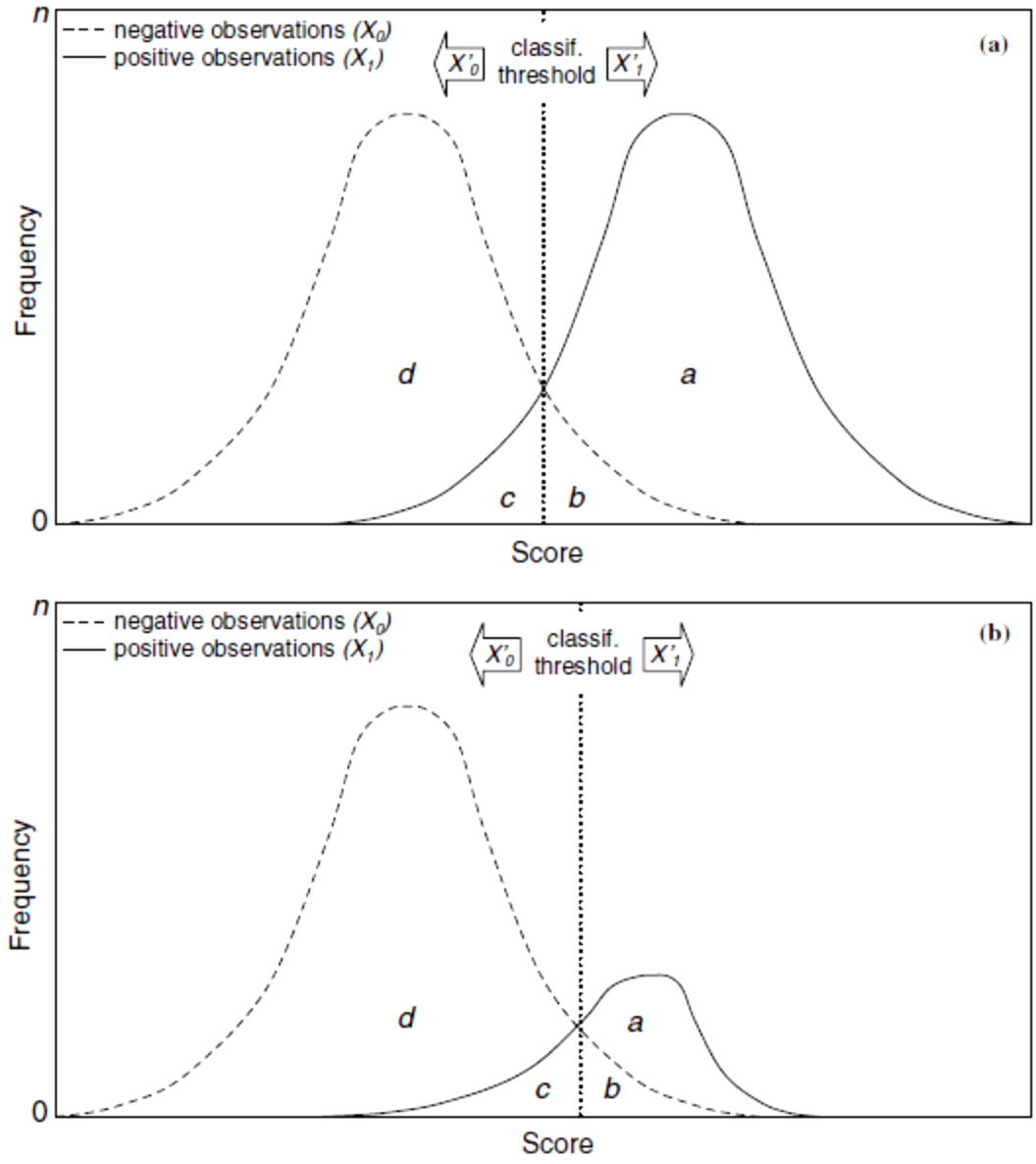


Figure 5-4 - The principle of classifying data of two classes into two groups of data with an optimizing a threshold (Beguería 2006).

5.5.3 Discriminant analysis

5.5.3.1 Introduction to discriminant analysis

Discriminant analysis is a statistical method for multivariate analysis, used for obtaining a function which provides the maximum separation between groups of data (Lachenbruch 1968). As for other multivariate techniques, discriminant analysis may be used to evaluate one or several groups of data containing different data classes. The aim is to separate the different classes - ideally into discrete groups, but as this may often not be achieved for complex or clustered data the optimum criterion for separation is searched. This criterion is described by the discriminant function, the outcome of the analysis, and its ability to successfully separate the different classes of data may be described

with an error rate; the criterions rate of misclassified samples. There are different types of discriminant analyses, dependent on how the group relationship is evaluated and how the discriminant function is defined; e.g. linear- and quadratic discriminant analysis (Lachenbruch 1968, The Mathworks Inc. 2011).

The simplest case of a discriminant analysis is performing a linear discriminant analysis (LDA) on only two groups of data. This may be thought of as using Multiple Regression; a linear equation of the form

$$y = a + b_1 * x_1 + b_2 * x_2 + \cdots + b_n * x_n \quad \textbf{(4)}$$

, where a is a constant and b_1-b_n are regression coefficients, could be fitted to the groups (StatSoft Inc 2011). However, when dealing with more groups in a single analysis the numerical background become more complex. Still, the method is founded on simple principles for evaluation of mean and covariance of the different groups and theirs samples (Lachenbruch 1968, StatSoft Inc 2011).

One important element of the discriminant analysis is the possibility of adding cost to the different data classes (Lachenbruch and Goldstein 1979). This increases the methods complexity, but is of high relevance for its analytical strength, as different classes may be given different weight. Consequently, the different classes of data may be prioritised by importance and affect the resulting discriminant function, changing the boundary condition i.e. the threshold it represents.

In the case of rainfall thresholds data the discriminant function represents the numerical description of the threshold – separating landslide triggering conditions form non-triggering conditions. Thus may the classes of data be defined as outcomes; landslide triggered or no landslide triggered, 'y' or 'n'. Discriminant analysis has previously been used in the study of threshold landslides the mathematical background of discriminant analysis may be found in e.g. (Lachenbruch 1968, Lachenbruch and Goldstein 1979).

5.5.3.2 Discriminant analysis using Matlab

Using Matlab, a discriminant analysis is executed using the function `classify` available in the Statistical Toolbox. Classify may estimate both linear and quadratic discriminant analyses (LDA and QDA respectively) on two sets of variables (predictors), using a third variable as a classifier (outcome variable). As in the statistical theory of discriminant analysis the different classes may be given a cost, in the program titled priority. The priority is given in a 1-by-2 matrix where each value is between 0 and 1, for this study priority is set to give a total priority equal to 1 (100%). For this study, as the (landslide) outcome variable is *y* or *n*, the level of priority is distributed between triggering and non-triggering rainfall conditions (The Mathworks Inc. 2011).

As only two predictors, resembling rainfall variables of daily data, are evaluated at once using a discriminant analysis the data could be presented in a scatter plot as shown in Figure 5-5. All daily data of one predictor is plotted against all daily data of the other; in the example 3 days by 6 days of accumulated antecedent rainfall (A_{3d} by A_{6d}) in mm/day. The non-triggering rainfall conditions are plotted as green dots and the triggering conditions as read circles, classified *n* and *y* respectively by the outcome variable, the blue line represents the discriminant function, i.e. the threshold.

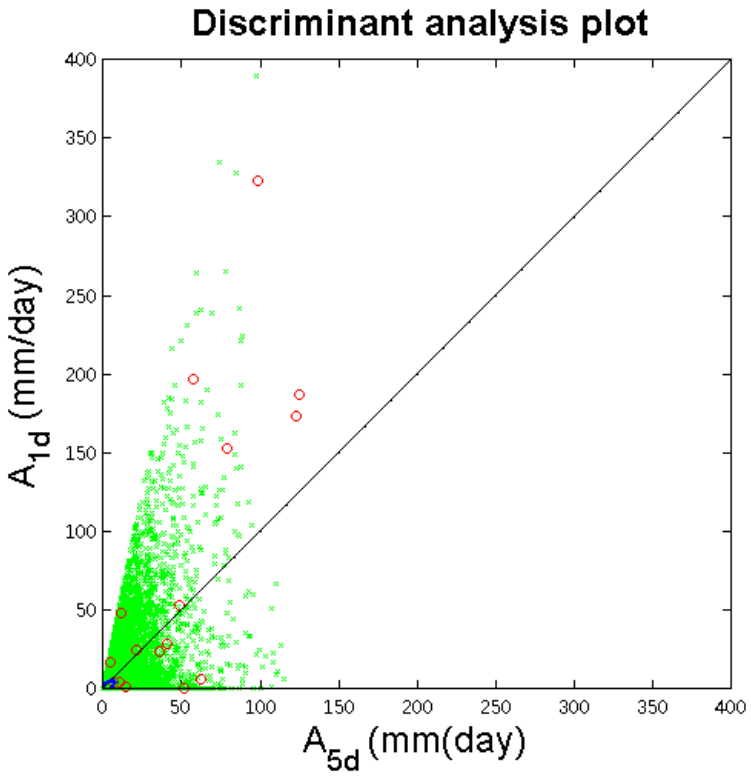


Figure 5-5 Example of DA scatter plot, without established discriminant function

5.5.3.3 Establishing a LDA methodology

A methodology for analysing data using discriminant analysis in Matlab where developed, with corresponding scripts presented in Appendix C. The simplest type of discriminant analysis where selected; that is linear discriminant analysis (LDA) on a set of two groups (predictors). Due to the limitation in number of predictors per analysis, conduction a LDA for each predictor combinations, one at a time, was not an alternative. One possible solution was to apply the results from the preceding classification tree analyses, using the predictor combinations suggested there. This would be a simple, and maybe good, solution, but to be sure to get the best result of using a discriminant analysis a different approach where chosen: A loop where created in Matlab where a LDA were executed for all predictor combinations successively, creating a multiple LDA. All combinations of predictors, equal to those combined in the correlation matrix of Table 6-6, where analysed. To assure the best quality of the results, only predictor combinations with a correlation coefficient (ρ) less than 0.8 where used, as for the classification tree analysis; the highly correlated combinations where ignored in further analysis.

The linear discriminant function was defined by coefficients of one constant and two linear terms as presented in equation (5), where K, L_1 and L_2 represent the constant and linear coefficients. x and y represents the analysed predictors equal to the x- and y-axis when plotted in a Cartesian coordinate plane, as in Figure 5-5.

$$0 = K + L_1 x + L_2 y \quad \textbf{(5)}$$

For this equation to be valid as a rainfall threshold there are some criteria that has to be met: We define the rainfall variables of which the predictors x and y corresponds, as

$$x = A_i \text{ and } y = A_j, \text{ where } i < j \text{ and } K > A_i > A_j$$

$$-L_2 A_j = K + L_1 A_i \quad \textbf{(6)}$$

,where $i < j$ and represent the duration of accumulated antecedent rainfall in days; note that antecedent rainfall is given in intensity (mm/day). It is assumes that A_i represent accumulated rainfall of relatively short duration compared to A_j, as highly correlated rainfall variables are removed from the analysed data; thus will A_j indirectly reflect soil moisture conditions and A_i short duration high intensity conditions. Increased A_j will then increase landslide susceptibility and decrease the amount of A_i required to trigger landslides. It is also reasonable to assume that the absolute value of the constant K is higher than the linear coefficients.

As L_1 and L_2 are numerical coefficients, equation (6) can be transformed to a more common form;

$$A_j = \frac{K + L_1 A_i}{-L_2} \quad \textbf{(7)}$$

$$A_j = a - bA_i\,, \quad a = \frac{\mathrm{K}}{-\mathrm{L}_2}\,, b = \frac{L_1}{-L_2} \quad \textbf{(8)}$$

,where a and b respectively are constant and linear coefficients describing the threshold curve.

Taking the mentioned criteria into consideration, equation (6) proves that the linear coefficients and the constant must have opposite sign and that the linear coefficients must have the same sign. As these criteria had to be met for producing valid rainfall thresholds, they were also included as criteria used when selecting the best performing threshold. Linear discriminant function which did not meet these criteria where disregarded.

Due to the high number of discriminant analyses a numerical evaluation of the linear discriminant function performance where needed, as a graphical evaluation of each LDA would be unrealistic. Such a numerical assessment of a threshold performance may be done by calculating a misclassification error rate (ERR). The ERR is a measure of the linear discriminant functions ability to successfully classify samples in a set of data. Error rates may be calculated for a particular sample, or for the full set of samples. As described by Lachenbruch (1968), the statistical theory of error rates in discriminant analysis is very complex. Simplified, it is a measure of the ratio between misclassified sampled and total number of samples. Performing discriminant analysis in Matlab, ERR is calculated and may be saved as a variable. Here ERR is automatically weighted by the priority applied to the discriminant analysis (The Mathworks Inc. 2011).

5.5.3.4 LDA methodology – prior probabilities

Compared to the non-triggering rainfall conditions in the landslide inventory, the landslide triggering conditions (class 'y') where highly outnumbered. As the samples of triggering conditions were of high importance, prior probability was applied to the linear discriminant analysis. Best fitted linear discriminant function, representing a threshold, were expected to be defined when applying high prior probability to the 'y' class, but to what extent was not known. A stepwise exploratory analysis was based on values of prior probability and applied to the data set: Each multiple predictor LDA, as explained in previous chapter were performed on a series of different prior probabilities, i.e. a second loop was applied. This methodology was used to investigate how prior probability affected the resulting linear threshold, and to locate the prior probability providing the best threshold performance. These steps where applied successively:

Table 5-5 Example on misclassification error matrix output from a multiple LDA. The red fields represent predictor combinations excluded due to high correlation coefficient, and remaining blank fields are due to invalid threshold equation

Pred	A_{1d}	A_{2d}	A_{3d}	A_{4d}	A_{5d}	A_{6d}	A_{7d}	A_{9d}	A_{11d}	A_{13d}	A_{15d}	A_{30d}	A_{60d}
A_{1d}	-	-	0,301	0,242	0,301	0,242	0,301	0,361	0,420	0,479	0,479	0,479	0,479
A_{2d}	-	-	-	-	0,419	0,419	0,419	0,419	-	-	-	-	-
A_{3d}	0,301	-	-	-	-	-	-	-	-	-	-	-	-
A_{4d}	0,242	-	-	-	-	-	-	-	-	-	-	-	-
A_{5d}	0,301	0,419	-	-	-	-	-	-	-	-	-	-	-
A_{6d}	0,242	0,419	-	-	-	-	-	-	-	-	-	-	-
A_{7d}	0,301	0,419	-	-	-	-	-	-	-	-	-	-	-
A_{9d}	0,361	0,419	-	-	-	-	-	-	-	-	-	-	-
A_{11d}	0,420	-	-	-	-	-	-	-	-	-	-	-	-
A_{13d}	0,479	-	-	-	-	-	-	-	-	-	-	-	-
A_{15d}	0,479	-	-	-	-	-	-	-	-	-	-	-	-
A_{30d}	0,479	-	-	-	-	-	-	-	-	-	-	-	-
A_{60d}	0,479	-	-	-	-	-	-	-	-	-	-	-	-

- First, the full range of possible prior probabilities (prior) were assessed at 0.05 intervals from [0.05 0.95] to [0.95 0.05], where prior was given by [(1-y) y] and y described the prior probability of the landslide triggering conditions. The best performing linear discriminant function from analysis of each prior were evaluated, and the prior probability providing the best conditions where basis for next step of the exploratory analysis.
- The same principles where used for the next step, only at a lower scale. Multiple predictor LDAs where executed 20 times for different prior of 0.001 intervals. The prior probability

established as best fit in step 1 where used as centre point (Sp) for the range of priors, i.e. analysing a range of Sp+0.01 to Sp-0.01. Based on misclassification error rates (ERR), the best performing linear discriminant function were selected for each prior interval, and the all total lowest ERR decided which prior value to use for the final step of the analysis.

- In a third and final step, the prior probability interval scale where further lowered 5:1, resulting in an interval scale of 0.002. An analysis where conducted using same method as in earlier steps; 20 intervals with result of step 2 as centre.

Based on the result of step 3, the best performing linear discriminant function of the full exploratory analysis could be established. The analysis would establish the best level of prior for 'y' class rainfall conditions, and which predictors to apply in a final LDA to produce the best performing rainfall threshold.

6 RESULTS

6.1 Inventories and data collection

The amount of rainfall and landslide data available for this thesis was limited. Different sources were used, resulting in high variety in which descriptions where available. Data provided by GSB and ADPC provided relatively detailed information considering landslide type, landslide material and coordinates. The TRMM based global satellite precipitation estimates, TMPA-RT products, were available back to 1 March 2005, limiting the landslide inventory to landslide occurrence from that day to present. The resulting full landslide inventory is presented in appendixes' Table A 3. Rainfall inventories were created in several stages: Extracted data from the initial TMPA-RT rainfall inventory series are presented in Table A 4 in appendixes. The inventory of different predictors, including an outcome variable, used for the threshold analyses is presented in Table A 5. This is presented by a sample on the data structure, as the full inventory of 4 years daily data could not to be attached to this document.

6.2 Comparison of TMPA-RT and rain gauge data

The only series of continuous rain gauge data were available for the Chittagong rain gauge; the data extended from 1st June 1996 to 31st December 2008. As TMPA-RT data were only available from 1 March 2005 until present, the compared period was only of almost 4 years; 1402 continuous days from March 2005 to December 2008. The satellite rainfall products included NASAs TMPA-RT products 3B41RT and 3B42RT. The missing TMPA-RT data for January and February 2005 was not of high importance as this is a part of the dry period with low rainfall intensity and no landslide records.

The position of the Chittagong rain gauge and the corresponding TMPA grid cell centre position were highly correlated with a latitude and longitude decimal degree position of 22.4° 91.8° and 22.375° 91.875° respectively, making a good basis for a local correlation of the different sources of rainfall data as the difference in position was small.

A small amount of the techniques of EDA were used to compare the rainfall data from TMPA-RT and the Chittagong rain gauge. Several simple techniques were used, as illustrated and explained in the following sections, comparing rain gauge data and the two TMPA rainfall products:

6.2.1 Comparing the rainfall series

The first part of exploratory data analysis is usually visualizing the data. A linear plot of the different data sources for the full period (Feb 2005 – Dec 2008) shows a high level of discrepancy between all three rainfall products, as shown in Figure 6-1 and Figure 6-2. Most peaks appearing in the rain gauge data (black line) seem to be picked up by the TMPA-RT products, but the magnitude is rarely in agreement. The merged TMPA-RT product seems to be most often overestimating extreme peaks, except from in 2005 when both TMPA-RT products tend to overestimate some peaks. This difference should not be due to TMPA-RT version numbers, as the current V6 was introduced in February 2005, but a change in the used IR satellite types (Huffman and Bolvin 2010) in the end of 2005 may be an

explanation. IR data seem to have a tendency to sometimes estimate smaller peaks of rainfall when no rainfall occurs, as may be seen in march and May 2005, Figure 6-1.

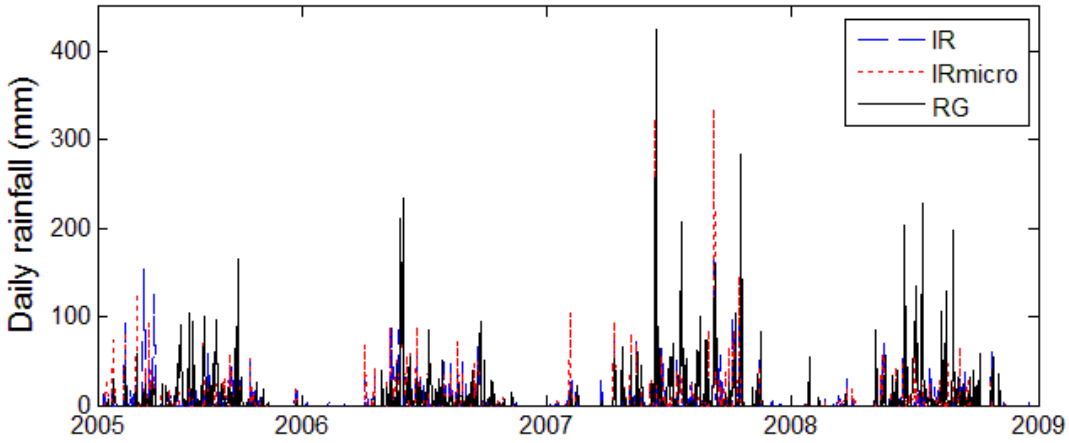


Figure 6-1 Plot of the complete TMPA-RT rainfall series (daily data) compared to rain gauge data

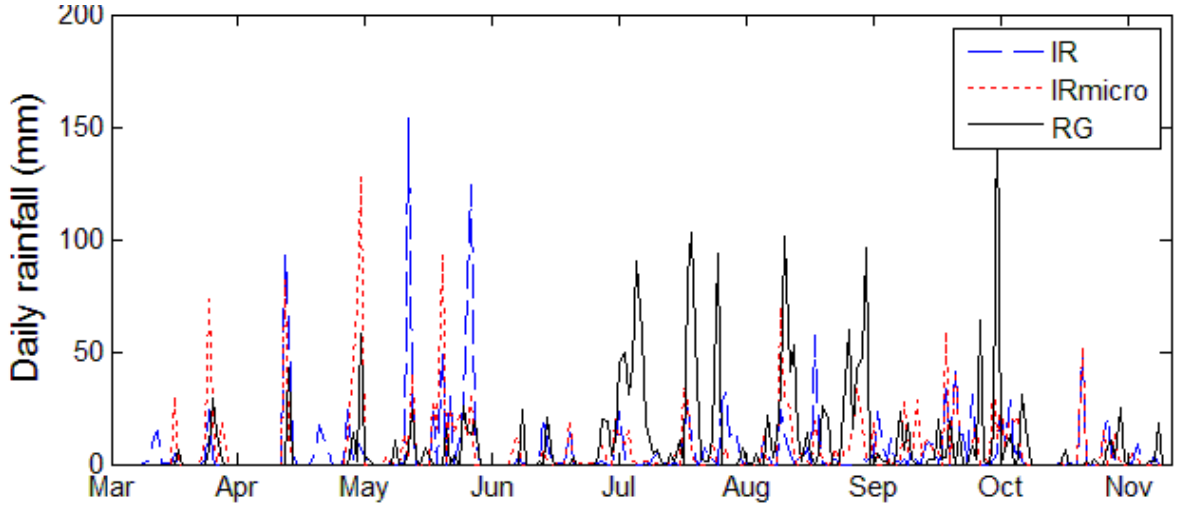


Figure 6-2 Plot of TMPA-RT rainfall estimates against gauge data; daily values in the period March 2005-November 2005

Table 6-1 Statistical variables for the full reifall series of IR, IRmicro and rain gauges

	mean	std	max
Rain gauge	8,78	26,9	425,0
IRmicro	6,39	20,5	334,4
IR	5,63	16,67	224,6

Comparing mean, standard deviation and maximum daily rainfall intensity (Figure 6-1) indicate that IRmicro produce rainfall estimates that are overall closer to the rain gauge data than the IR data. It is important to notice that the highest rainfall intensity measured by the IRmicro product is highly overestimated, but conversely is the series absolute maximum rainfall intensity measured by the rain gauge relatively well estimated, but slightly underestimated.

6.2.2 Comparison of monthly rainfall

Histograms of monthly average rainfall were used to compare the total amount of rainfall between the different TMPA-RT products and the rain gauge data (Figure 6-3). The full 4-year correlation period were used to compare these data. Monthly rainfall is not of importance for the short high-intensity rainfalls most likely to trigger landslides, but is still important for understanding the data and for observing at the annual variations. This comparison may also indicate the accuracy of antecedent rainfall of longer durations, an indicator for the long term soil water conditions and such a possible variable for rainfall thresholds. Comparing these data for rainy season, May to September-October is most important, as this is the period where landslide triggering has been observed (see Figure 6-3).

Figure 6-3 a) shows the average monthly rainfall for the studied 4-year period, comparing TMPA-RT and gauge data. It also compare the rainfall data to the distribution of landslides registered in the rainfall inventory; one line for landslides in the study area for comparing rainfall data – Chittagong (CG), and one for the full landslide inventory of Bangladesh. The figure shows that the TMPA products generally overestimate rainfall during the winter and hot weather period (January-May), and underestimates rainfall during the monsoon and post-monsoon period (June-December). Still, TMPA-RT data follow the annual rainfall trend except for the months June – August. For this period, of highest amounts of rainfall and highest occurrence of landslides, rainfall is extremely underestimated for both TMPA-RT products.

Due to the high discrepancies in extreme rainfall values, as illustrated in the previous section, monthly rainfall values where compared only including intensities up to a certain level; < 50 mm and < 30 mm as shown in Figure 6-3 b) and c) respectively. The tendency of underestimated TMPA-RT rainfall estimates were still there, but almost removed for rainfall intensities below 30 mm. This indicates that TMPA-RT data predict rainfall poorly at higher intensities, which could be bad for establishing a rainfall threshold for the area.

6.2.3 Number of rainy days per month

A comparison of the number of rainy days per month was also conducted, as this is not important for a rainfall threshold, but is used for evaluating how the TMPA-RT estimates perform generally compared to the rain gauge. It was found that the satellite rainfall estimates registered more rainfall days than the rain gauge in Chittagong (se Figure 6-4). From the distribution of rainfall days throughout the year there are no sign of discrepancy in specific months; the TMPA-RT data have generally more registered rainy days throughout the year. The higher number of rainy days in the TMPA-RT data may be caused by rainfall in areas outside of Chittagong City, as one TMPA rainfall grid covers an area of 0.25 x 0.25 degrees - equal to an area greater than 25 x 25 km. As landslide

triggering are commonly related to high intensity rainfall and most daily rainfall values are expected to be of low intensity, this may not pose a problem for a threshold using one of the TMPA-RT rainfall products. On the other hand if some discrepancy is caused by high intensity rainfall, this may cause errors in the rainfall thresholds. A plot of rainfall intensity distribution and correlation plots of daily rainfall had to be made to look into this (see Figure 6-4 and Figure 6-5).

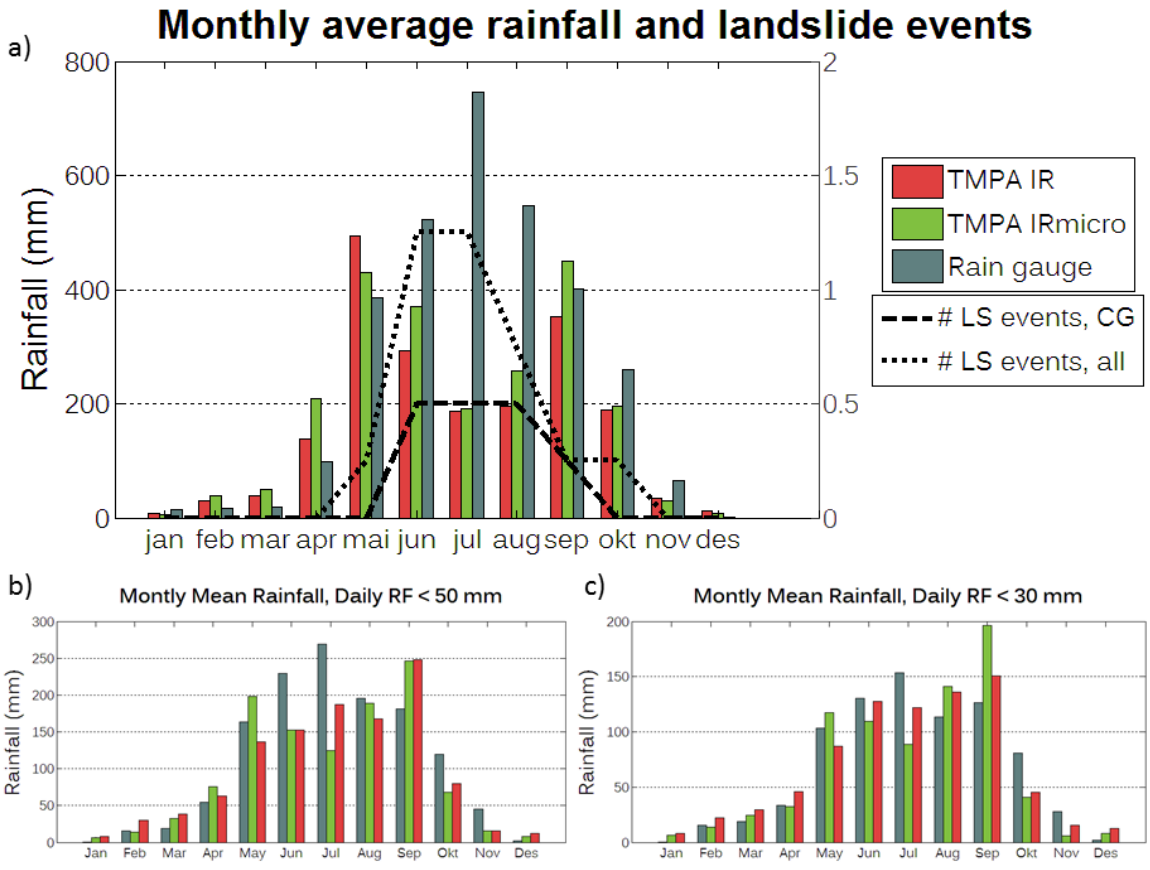


Figure 6-3 Monthly average rainfall (2005-2008) for both TMPA-RT products and Chittagong rain gauge, presenting seasonal variations: a) monthly rainfall data compared to occurrence of landslides in Chittagong and Bangladesh. b) and c) monthly accumulation of rainfall when extreme daily rainfall events are removed from the data.

Table 6-2 Annual mean rainfall for de different rainfall sources

Rainfall source	TMPA IR	TMPA IRmicro	Rain gauge
Sum yearly average rainfall	1974 mm/year	2239 mm/year	3078 mm/year

6.2.4 Rainfall intensity distribution

As mentioned earlier extreme rainfall intensities were expected to be most important for establishing the rainfall threshold, so deviations in low intensity rainfall could be tolerated to higher extent than for high intensity rainfall. Comparing the occurrence of high daily rainfall values was thus important. A rainfall intensity distribution was created, plotted as a histogram, comparing both TMPA-RT products and gauge data at different rainfall intensities (see fig Figure 6-5). Because of the high discrepancy in monthly rainfall values for the summer months June-August, a similar plot was also created for these months. This graphical method is commonly used in exploratory data analysis (EDA) to show how the data are distributed; using different distributions and distribution functions (PDF) to describe the data (Ang and Tang 2007). Instead of probabilities the data are here presented in number of rainy days.

The majority of rainy days are of low rainfall intensity, and there are only few days of high or extreme rainfall intensity. The daily rainfall data may be described as exponentially distributed, with a high positive skewness and high kurtosis (Ang and Tang 2007). Comparing the different rainfall sources intensity distribution it becomes clear that the TMPA-RT rainfall estimates tend to overestimate the number of low intensity rainy days (≤ 20 mm/day). For rainfall values above 20 mm/day the number of rainy days are underestimated, compared to gauge data. Especially for daily rainfall values of more than 160 mm the satellite rainfall estimates were almost absent.

6.2.5 Rainfall data correlation

Scatter plots are probably the most important visual technique for determining how good the TMPA-RT data will be for a threshold analysis, as a good correlation is depending on timing if the rainfall as well as the intensity. The rainfall intensity of each day from one data source is plotted against the rainfall intensity of each day from another. In case of a bad correlation, a trend in the scatter plot may be hard to find.

Each TMPA-RT product were plotted against gauge data in scatter plots; one scatterplot discriminating between triggering and non-triggering conditions for landslide events in Chittagong (Figure 6-6), and one including an evaluation of when in the season the conditions occurred (Figure 6-7). Both TMPA-RT data presented bad correlation to the rain gauge data, but the combined IR and microwave product revealed a slightly better correlation. This was also supported by the correlation coefficients; 0.248 for IR and gauge, 0.214 for gauge and IRmicro. Accounting only landslide triggering events the IRmicro estimates also preform best, but given correlation coefficients the performance of either TMPA-RT data are desirable. By Figure 6-7 it may be noted that the highest extremes of daily rainfall occurs during the monsoon period, and some in the month before or in the post-monsoon period; for normal high to low values no trend is clear.

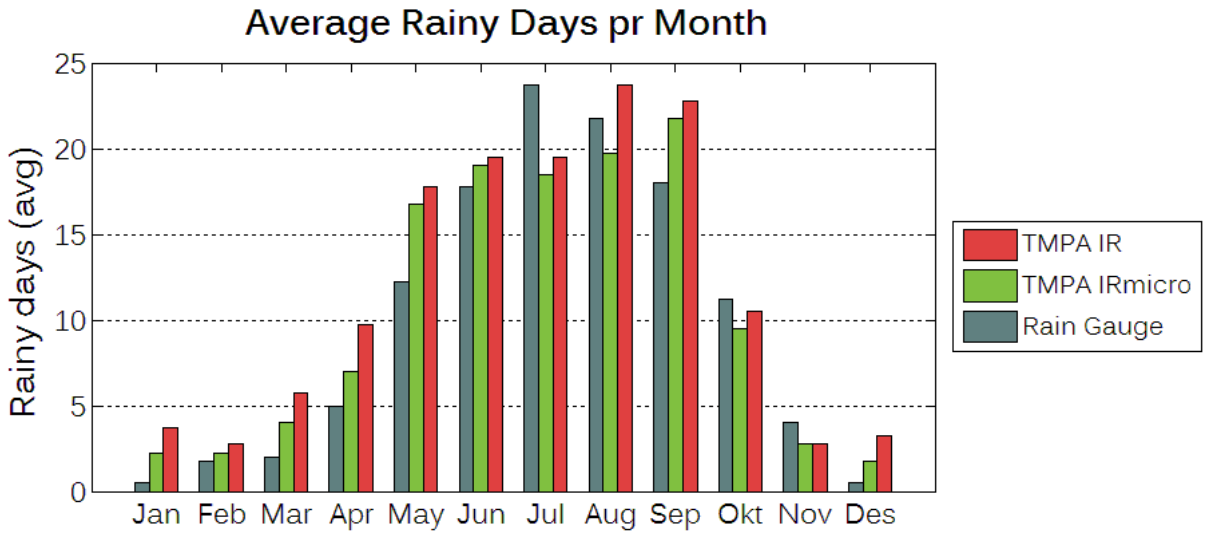


Figure 6-4 Number of rainy days per month registered by the TMPA-RT products and rain gauge (averaged 2005-2008)

Table 6-3. Total annual averages rainy days for rain gauge and TMPA-RT rainfall estimates

Rainfall source	TMPA IR	TMPA IRmicro	Rain gauge
Average rainfall days per year	142 days	125 days	119 days

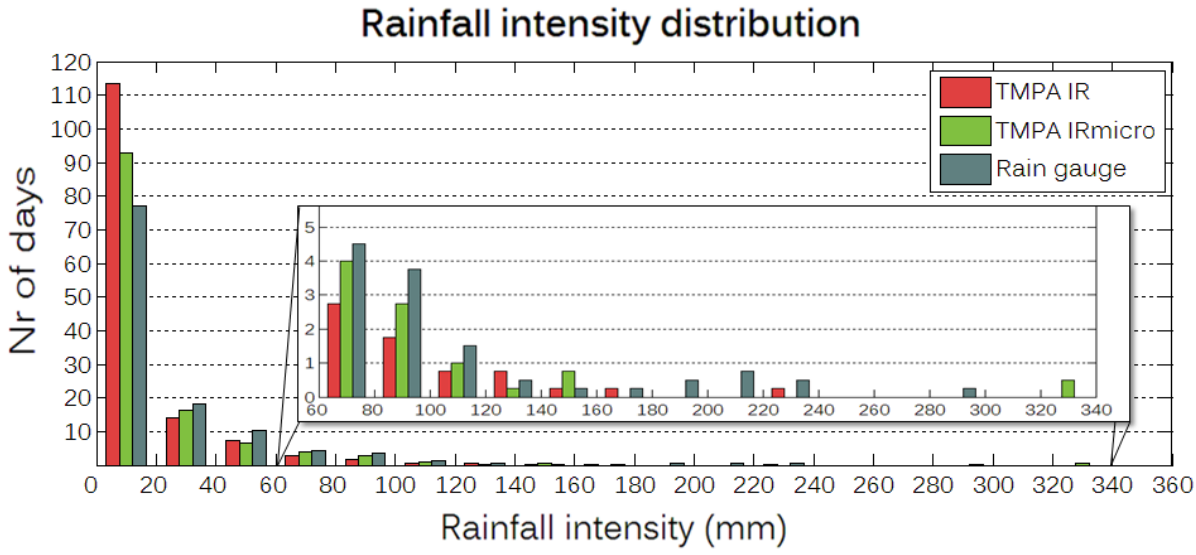


Figure 6-5 Rainfall intensity distribution of TMPA-RT and rain gauge data averaged for 2005-2008.

Table 6-4 Correlation coefficients for TMPA-RT data plotted against rain gauge data

Correlation coefficients	IR (3B41RT)	IRmicro (3B42RT)
Rain gauge	0.248	0.214

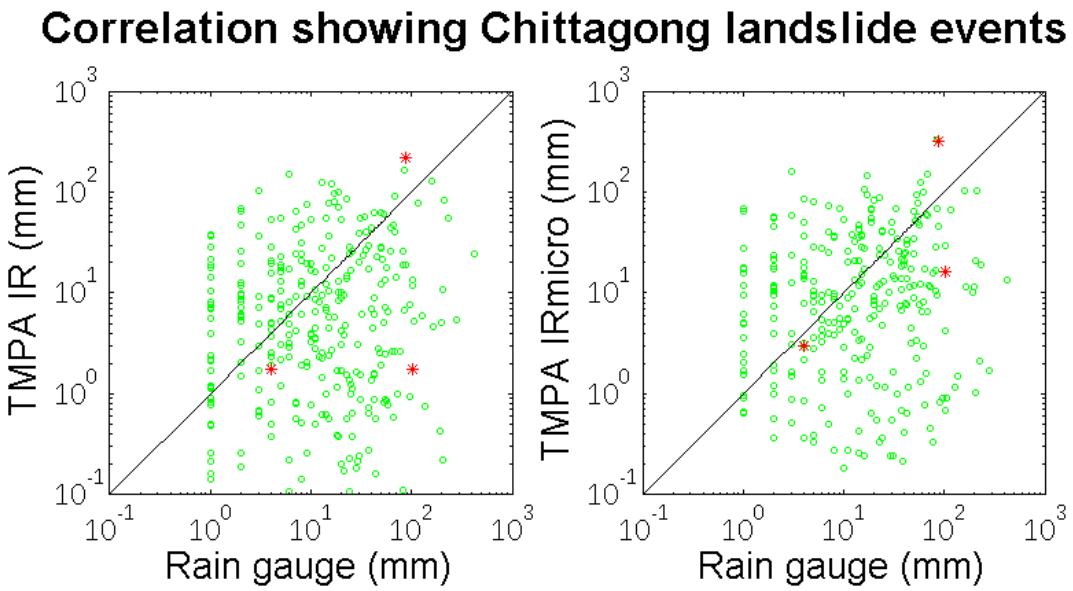


Figure 6-6. Correlation plots of the full series of daily data for each TMPA-RT rainfall estimate against rainfall data (red stars represent days where landslide events occurred in the Chittagong area)

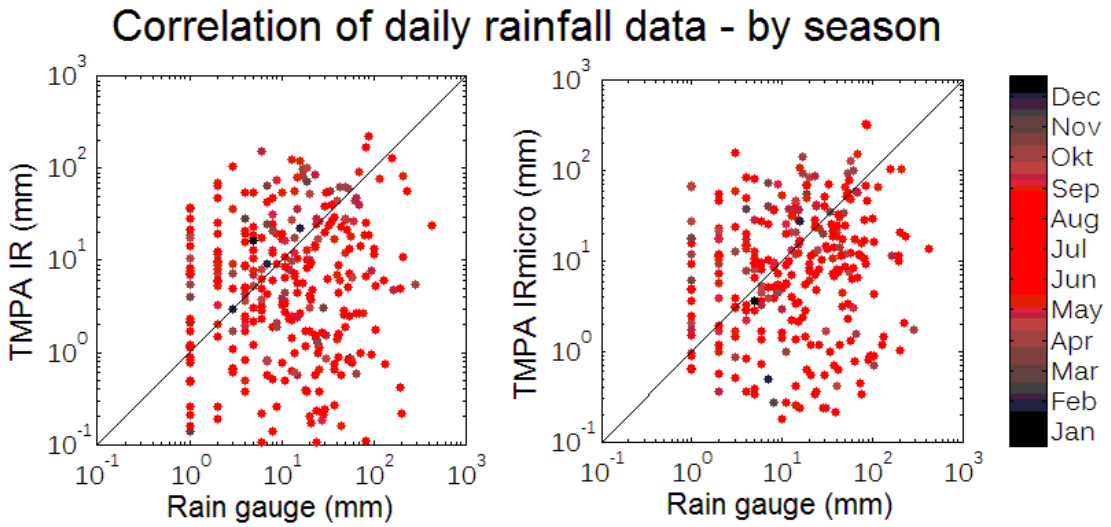


Figure 6-7 Correlation plots of same data as Figure 6-6 with respect to seasonal rainfall variations; red represent rainy period and black period represent relatively dry periods, transfer areas represent periods of medium rainfall amounts

6.2.6 Missing data – TMPA-RT

Discrepancies in the analysis above, between TMPA-RT products and Rain Gauge data, could have been affected by missing TMPA-RT data due to temporal lack of satellite coverage. It was found that a total of a total of 227 and 106 days were missing data in the period of study, for IR (3B41RT) and IRmicro (3B42RT) respectively. Table 6-5 shows the number of days where some TMPA-RT data were missed, sorted by satellite data type and by how many hours were missing per day; the data is given so the first column represent all days of missing data, the second represent only days with more than 6 hours of missing data, and so on. The merged TMPA-RT product has fewer days off missing data, as expected due to combining the IR- and PMW- satellite data. On the other, more days have a higher number of lost hours; this is probably as the merged data has a resolution of 3 hours, while the IR data has 1 hour resolution and such a lower loss of hours for each file missing.

The average rainfall of the days of missing data is relatively low, far below 10 mm/day, meaning that the total loss of rainfall data is not high considering that landslides occur in periods of intense rainfall. What is most important are if rainfall data where missed during high rainfall intensities resulting in loss of potentially landslide triggering rainfall conditions. A loss of more than 12 hours of rainfall during a rainy day have not been recorded here, but during up to 12 hours occurred during days of rainfall intensities of more than 70 mm/day for both TMPA-RT products. This has probably caused some lower TMPA-RT rainfall estimates. It is important to be aware of these missing data, but considering the other uncertainty and discrepancies found in the previous analysis this cannot be said to make a huge difference.

Table 6-5 Missing TMPA-RT coverage and comparison to rain gauge data in the "lost" periods; a missed day represent a day where at least one hour of data were missing, NaN represent a missed hour of rainfall data.

Source		NaN > 0	NaN > 6	NaN > 12	NaN > 18
IR	Missed days	227	13	6	3
	Rainfall average (missed days)	4,45 mm	7,42 mm	0 mm	0 mm
	Max (missed day)	75,83 mm	75,83 mm	0 mm	0 mm
IRmicro	Missed days	106	21	16	15
	Rainfall average (missed days)	4,71 mm	5,42 mm	0 mm	0 mm
	Max (missed day)	80,55 mm	70,19 mm	0 mm	0 mm

6.3 Threshold analysis

Based on results of the TMPA-RT and rain gauge data comparison it was concluded to do the threshold analysis based on the merged IR and PWM product, 3B42RT; this was the product estimating rainfall closest rain gauge data, both regarding intensities and intensity distribution, total rainfall amounts, correlation and the least number of missing days. Some intensity peaks of the merged product were severely overestimated, but this was acceptable as underestimation of peak intensities generally seemed less than for the IR product.

After preparing the landslide data for threshold analysis, some further comparison of gauge and TMPA-RT data were possible. Accounting each day of the analysed period (2005-2010) as one data sample, a rainfall condition, and each variable of antecedent rainfall of each day as a data sample, the available data consisted of a total of 16824 samples equally spread over 12 different TMPA-locations (grid boxes); all these samples where to be assessed in the threshold analysis.

6.3.1 Comparing landslide triggering rainfall data

As the outcome variable distinguished between triggering and non-triggering rainfall days, the triggering landslide conditions of the satellite rainfall estimates could be compared against rain gauge data: Rainfall data from landslide triggering days where extracted from the 3B42RT rainfall inventory and are presented in Table A 6 in appendix A. Plotting these data in a semi-logarithmic intensity-duration space reveals no certain trends for all data in general (Figure 6-8), put it can be pointed out that the events of the highest intensities (> 100 mm/day in day of landslide) showed similar trends; these also represented the recent most fatal landslide events in Bangladesh.

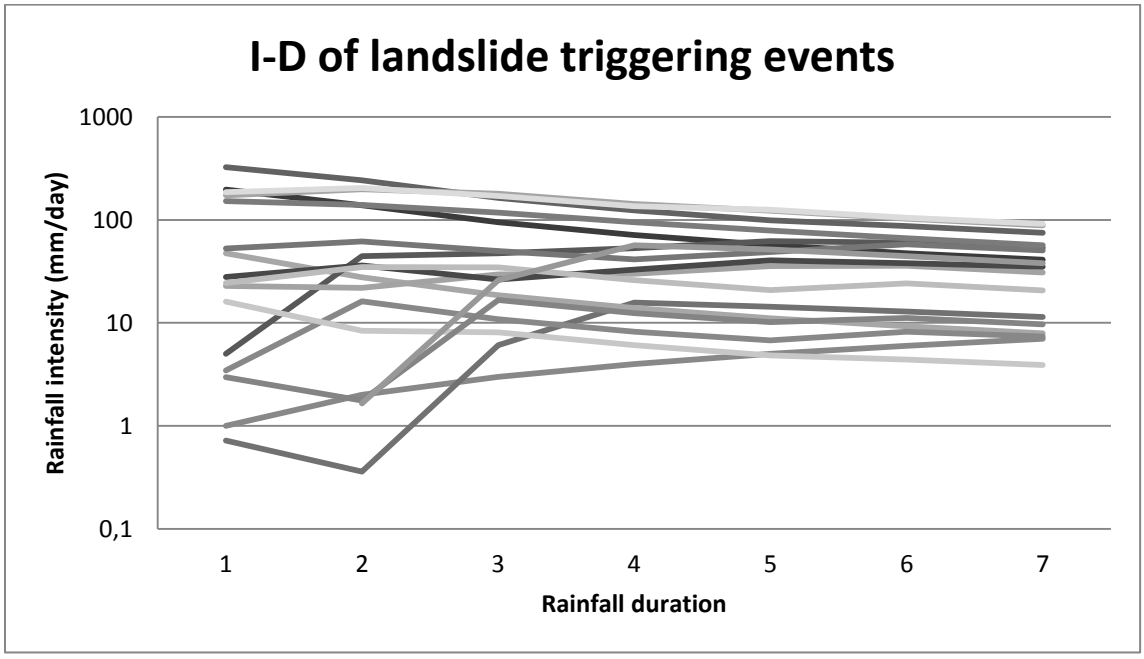


Figure 6-8 I-D plot of all landslide triggering events in Bangladesh for 3B42RT rainfall estimates

Rainfall data from five TMPA-RT grid boxes represented areas where the 11 June 2007 Chittagong landslides event and the 15 June 2010 event in Cox's Bazar and Bandarban. These data where compared against rain gauge data provided by ADPC as presented in appendixes' Table A 7 and in Figure 6-9 and Figure 6-10. This comparison stressed what was already indicated from the complete rainfall series comparison, that the extreme rainfall intensities are not properly estimated by the TMPA-RT satellite based rainfall estimates. On the other hand it revealed similarities in the I-D curves of rain gauge and 3B42RT data. It must be noted that the 3B42RT data overestimated the rainfall intensities in the Cox's Bazar area, and underestimated the rainfall in Chittagong. For the Cox' Bazar

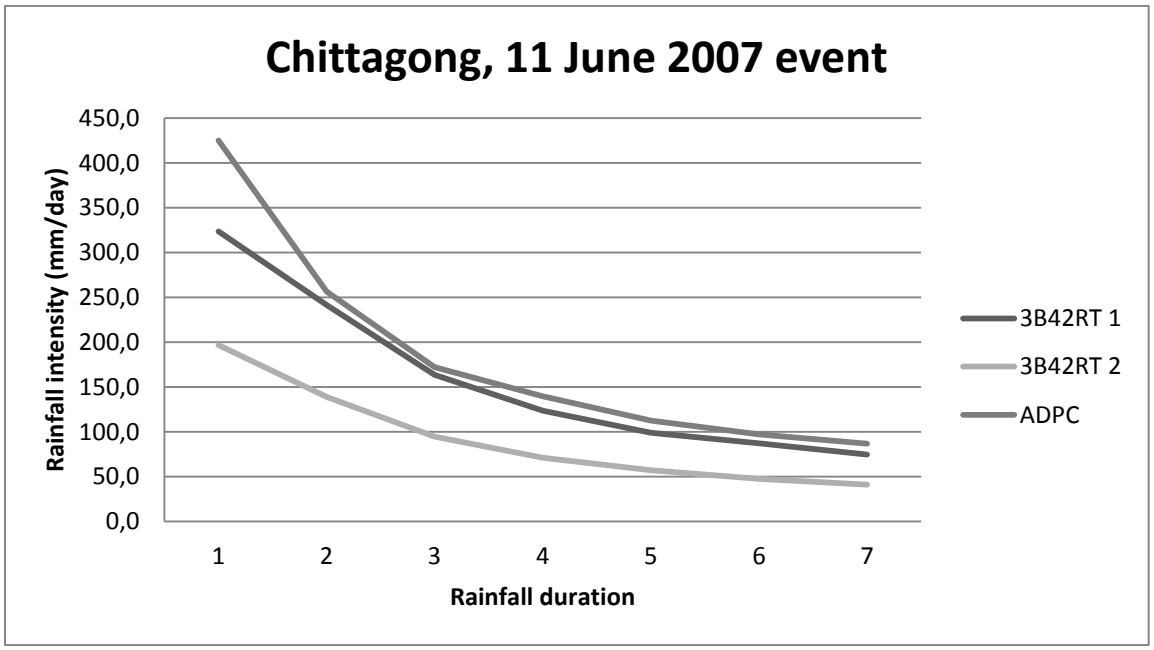


Figure 6-9 I-D plot of gauge and 3B42RT data of the fatal event in Chittagong, 11 June 2007 (3B42RT data represented by two different grid box positions). Coordinated may be found in appendixes' Table A 7

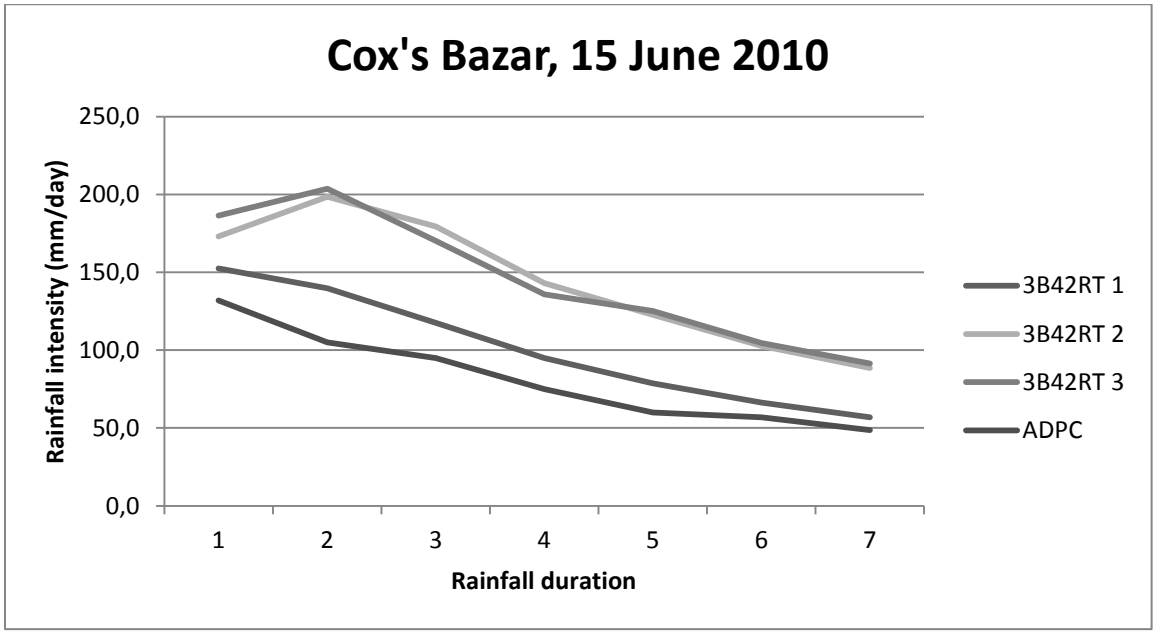


Figure 6-10 I-D plot of gauge and 3B42RT data of the fatal event in Cox's Bazar and Bandarban district of 15 June 2010; 3B42RT data represented by three different grid box positions (coordinated may be found in appendixes' Table A 7)

event it should be mentioned that the rainfall intensities may have been stronger in other parts of the exposed area than what was measured by this one rain gauge; a IFCR (International Federation of Red Cross and Red Crescent Societies) DREF (Disaster Relief Emergency Found) operation report (IFRC 2010a) of the event in Cox' Bazar referred to a rainfall of 18 inches (461 mm) in the period 12 –

15 June. This 4-days average rainfall intensity of more than 115 mm/day is more comparable to the 3B42RT rainfall values.

6.3.2 Predictor correlation

The data set prepared for establishing thresholds using statistical multivariate analyses consisted of 13 different predictors. Each predictor corresponded to a variable of antecedent rainfall of mm/day; each with series of rainfall intensity values for different duration of accumulated rainfall. This is further explained in section 5.5.1.1. As all predictors were to be combined in different multivariate analysed, and high dependence between variables might affect the results, all these predictors were assessed regarding their correlation coefficients. The result of the assessment was a correlation matrix as shown in Figure 6-6. A high dependency, correlation coefficients (ρ) ≥ 0.8 is shaded red in the figure; these combinations were not desired for the final established thresholds.

Table 6-6 Correlation coefficient matrix for 3B42RT (highly correlated predictors (ρ > 0.8) are shaded red)

Pred	A_{1d}	A_{2d}	A_{3d}	A_{4d}	A_{5d}	A_{6d}	A_{7d}	A_{9d}	A_{11d}	A_{13d}	A_{15d}	A_{30d}	A_{60d}
A_{1d}	1,000	0,827	0,705	0,632	0,574	0,537	0,509	0,472	0,447	0,426	0,414	0,368	0,306
A_{2d}	0,827	1,000	0,904	0,813	0,746	0,693	0,655	0,604	0,569	0,541	0,524	0,459	0,379
A_{3d}	0,705	0,904	1,000	0,936	0,865	0,809	0,763	0,701	0,659	0,627	0,604	0,525	0,434
A_{4d}	0,632	0,813	0,936	1,000	0,953	0,896	0,849	0,778	0,731	0,695	0,668	0,578	0,478
A_{5d}	0,574	0,746	0,865	0,953	1,000	0,963	0,917	0,842	0,791	0,753	0,723	0,623	0,516
A_{6d}	0,537	0,693	0,809	0,896	0,963	1,000	0,970	0,898	0,842	0,801	0,770	0,662	0,550
A_{7d}	0,509	0,655	0,763	0,849	0,917	0,970	1,000	0,943	0,886	0,843	0,811	0,696	0,579
A_{9d}	0,472	0,604	0,701	0,778	0,842	0,898	0,943	1,000	0,958	0,913	0,878	0,750	0,630
A_{11d}	0,447	0,569	0,659	0,731	0,791	0,842	0,886	0,958	1,000	0,967	0,931	0,794	0,671
A_{13d}	0,426	0,541	0,627	0,695	0,753	0,801	0,843	0,913	0,967	1,000	0,974	0,832	0,706
A_{15d}	0,414	0,524	0,604	0,668	0,723	0,770	0,811	0,878	0,931	0,974	1,000	0,864	0,737
A_{30d}	0,368	0,459	0,525	0,578	0,623	0,662	0,696	0,750	0,794	0,832	0,864	1,000	0,882
A_{60d}	0,306	0,379	0,434	0,478	0,516	0,550	0,579	0,630	0,671	0,706	0,737	0,882	1,000

6.4 Classification tree analysis

All classification tree analyses (CTA) were performed according to the methodology described in chapter 5.5.2, on the merged TPMA-RT rainfall estimates, and the results were analysed using receiver operation characteristics (ROC). Several CTAs were completed, each resulting in a classification tree; these are fully presented in Appendix B. The classification tree analyses where performed in three stages to investigate how predictor correlation affected the results, as described section 5.5.2.1:

- Initial stage – all predictors: a CTA were conducted without concern of predictor correlation coefficients.

- Stage 2 – most predictors: all predictors of high correlation to one selected predictor were excluded from the CTA. Several CTAs were conducted, using different predictor as exclusion criterion each time.
- Final stage – some predictors: with basis of two predictors, removing all predictors highly correlated to these two, a final CTA where conducted.

Some CTAs proved not to deliver any valuable results; a full analysis will not be presented for these, but all resulting classification trees are presented in 88Appendix B. For the CTAs providing valuable results to some degree the most important segments of the classification trees and the corresponding data and ROC curves presented.

For classification by CTA in Matlab it became clear that the different predictors must be handled with caution, as predictor names cannot be inserted to the applied data set. The output predictors in the classification tree are given the variables x_1, x_2, x_3, …, x_n corresponding to each set of data (each column corresponding to the data in figure XXX). This way, when one predictor is removed, all succeeding variables will be assigned new predictors. The relation between classification tree variables and the predictors corresponding rainfall variables are listed in figures and tables to avoid confusion.

6.4.1 CTA – stage 1

The first stage of CTA utilized all 13 predictors. Threshold criteria were selected one variable at the time, one classification tree branch level at the time. For each level the best preforming criteria were selected, as illustrated in Figure 6-11, and for each variable applied to the threshold the thresholds performance where assessed using ROC variables true positive rate (TPR), false positive rate (FPR) and accuracy (ACC), as defined in section 5.5.2.2. For each added threshold variable (or lowered variable value) the number of false positives increased. The increase was strong when adding the 5th threshold criteria, resulting in a high false positive rate.

Assessing these data using the ROC curve it was found that the threshold resulting in FPR = 0.625 and FNR = 0.034 provided the best performance. This threshold includes 3 variables, as presented in Table 6-7 below, utilizing the predictors x2, x5 and x6 representing the rainfall variables A_{2d}, A_{5d}, and A_{6d} respectively. The variable A_{2d} appears in two criteria, (a) A_{2d} = 231 and (b) A_{2d} = 138, corresponding to the CTA criteria x2 < 231, and x2 < 138. The existence of criterion *b* only lowers the value of the already existing criterion.

Using the suggested variables from the combination of CTA and ROC then result in the following threshold criteria, at which exceeded would result in a high probability of triggering of landslides:

$$\begin{aligned} \mathrm{A_{2d}} &= 230.8\ mm/d \\ \mathrm{A_{5d}} &= 119.1\ mm/d \\ \mathrm{A_{6d}} &= \ \ 35.9\ mm/d \end{aligned} \qquad \textbf{(9)}$$

These threshold criteria represent a challenge, concerning its reliability: The large difference between $\mathrm{A_{5d}}$ and $\mathrm{A_{6d}}$ is challenging; as this indicates the two variables are not highly dependent on each other, but the predictor correlation coefficient show high correlation, ρ = 0.953; high

correlation where defined as ρ ≥ 0.8. Additionally, as A_{2d} and A_{3d} are highly correlated, ρ = 0.904, these are not ideal to be combined in the threshold. Summarized, these things indicate that this threshold cannot be used.

Table 6-7. Thresholds selected in initial CTA and corresponding ROC values

CTA pred-ictor	Equiv. rainfall variable	Threshold values (mm/day)				Threshold value (mm/hr)	# of threshold variables	TP	FP	FPR	TPR
		A_{2d}	A_{3d}	A_{5d}	A_{6d}						
x5	A_{5d}			119,1		119,1	1	2	0	-	0,125
x2	A_{2d}	230,8		119,1		230,8	2	3	1	5,95E-05	0,188
x2	A_{2d}	138,1		119,1		138,1	2	5	19	1,13E-03	0,313
x6	**A_{6d}**	**138,1**		**119,1**	**35,9**	**35,9**	**3**	**10**	**571**	**0,0340**	**0,625**
x3	A_{3d}	138,1	6,1	119,1	35,9	6,1	4	*11*	*12682*	0,755	*0,688*

6.4.2 CTA – stage 2

Based on which variables of antecedent precipitation where selected in the initial CTA threshold, the second stage were initiated performing CTAs with a selection of variables. In four separate CTAs the single rainfall variables, A_{2d}, A_{3d}, A_{5d} and A_{6d}, were used as basis for excluding all predictors of ρ ≥ 0.8. As an example, according to the correlation coefficients in Table 6-6 using the rainfall variable A_{5d} would exclude the variables A_{3d}, A_{4d}, A_{6d}, A_{7d} and A_{9d}. Correlation between the other parameters selected of the CTA was not considered.

It was found that the ROC curve again suggested (see Figure 6-12) a 3-variable threshold. Similar problems appeared to be present in the thresholds of these CTAs as of the initial CTA. The CTAs using A_{2d} and A_{5d} as basis for parameter selection, where the only without highly correlated predictors in the suggested threshold; the resulting thresholds were the same and are thus presented as one. The other CTAs were disregarded and are not presented here.

Predictors of high correlation where present in the suggested thresholds as well; A_{5d} and A_{6d}. These two variables are highly correlated, but the suggested intensities for the threshold suggested something different with values of 119.1 mm/day and 35.9 mm/day respectively. The resulting threshold of this stage did not change compared to of the initial CTA threshold. All ROC parameters also stayed the same, including TP, FP, TPR, FPR and the accuracy ACC. The ROC curve thus also stayed the same (Figure 6-12). The new threshold was defined by the following 3 criteria, according to the classification tree in presented in Figure B 2 in appendixes and the following variables and parameters:

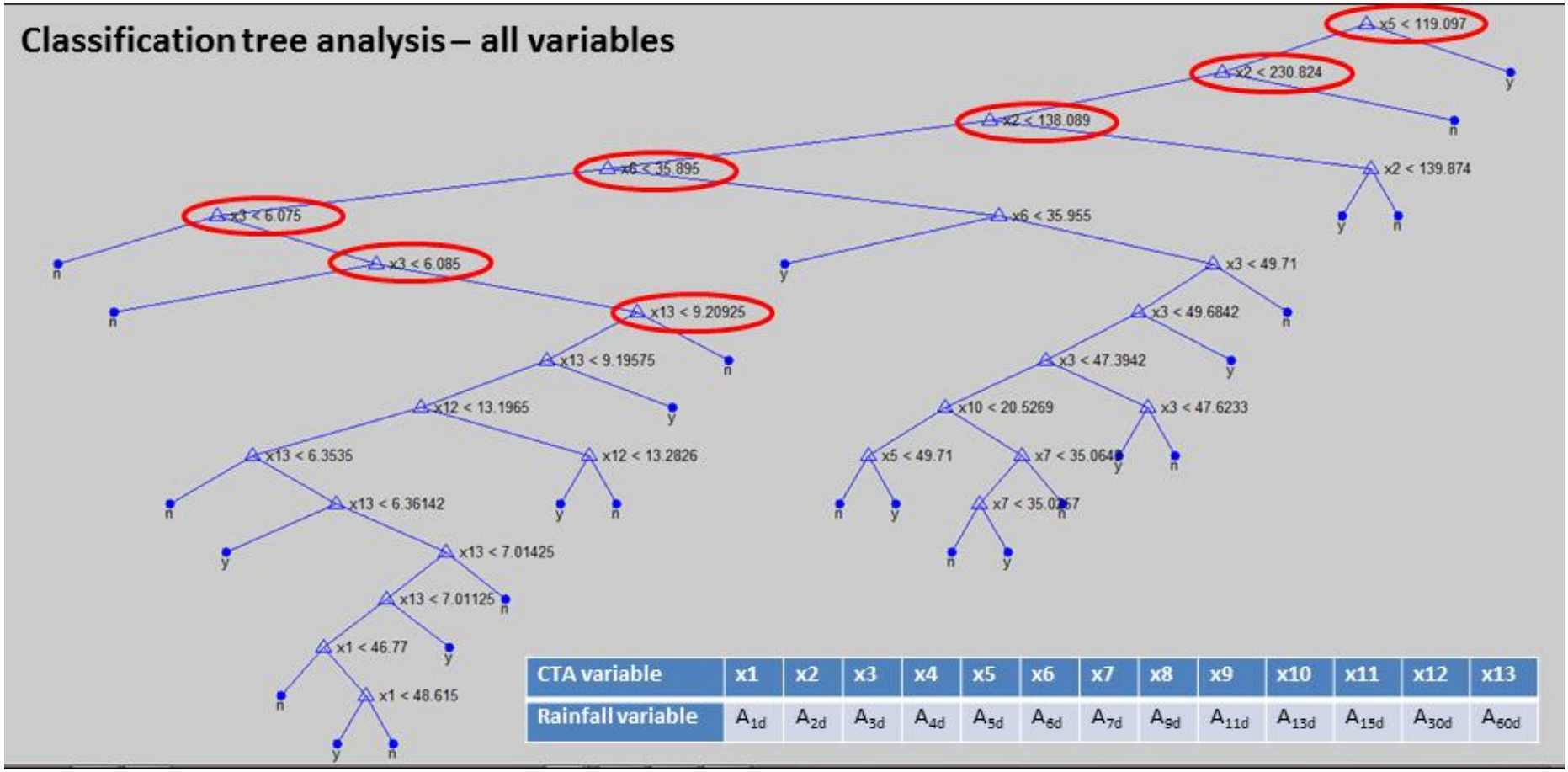


Figure 6-11 Initial CTA; selected criteria for evaluation of potential thresholds by ROC

$$\begin{aligned} A_{2d} &= 230.8\ mm/d \\ A_{5d} &= 119.1\ mm/d \\ A_{6d} &= 35.9\ mm/d \end{aligned} \quad \textbf{(10)}$$

Table 6-8. Thresholds selected in CTA stage 2 and corresponding ROC values

CTA pred-ictor	Equiv. rainfall variable	Threshold values (mm/day) A_{2d}	A_{5d}	A_{6d}	A_{11d}	Threshold value (mm/hr)	# of threshold variables	TP	FP	FPR	TPR
x2	A_{5d}		119,1			119,1	1	2	0	-	0,125
x1	A_{2d}	230,8	119,1			230,8	2	3	1	5,95E-05	0,188
x1	A_{2d}	138,1	119,1			138,1	2	5	19	1,13E-03	0,313
x3	**A_{6d}**	**138,1**	**119,1**	**35,9**		**35,9**	**3**	**10**	**571**	**0,0340**	**0,625**
x6	A_{11d}	138,1	119,1	35,9	7,05	7,05	4	11	*11632*	0,692	*0,688*

6.4.3 CTA - stage 3

The final stage of the CTA approach for establishing a rainfall threshold was using a combination of several predictors as basis for excluding others from the analysis. The two rainfall variables appearing at the top two CT branches in all previous CTAs were selected; A_{2d} and A_{5d}. This excluded approximately half of the rainfall variables from the analysis; A_{1d}, A_{3d}, A_{4d}, A_{6d}, A_{7d} and A_{9d}. ROC curve and variables (see Figure 6-12 and Table 6-9) related to this final CTA suggested a threshold of only 2 variables; none of high correlation (see Table 6-6). The final CTA's suggested threshold criteria and ROC parameters are presented below, and resulted in the following threshold criteria:

$$\begin{aligned} A_{2d} &= 138.1\ mm/d \\ A_{5d} &= 35.5\ mm/d \end{aligned} \quad \textbf{(11)}$$

Table 6-9. Thresholds selected in the final stage in CTA approach and the corresponding ROC values

CTA pred-ictor	Equiv. rainfall variable	Threshold values (mm/day) A_{2d}	A_{5d}	A_{11d}	A_{60d}	Threshold value (mm/hr)	# of threshold variables	TP	FP	FPR	TPR
x2	A_{5d}		119,1			119,1	1	2	0	-	0,125
x1	A_{2d}	230,8	119,1			230,8	2	3	1	5,95E-05	0,188
x1	A_{2d}	138,1	119,1			138,1	2	5	19	1,13E-03	0,313
x2	**A_{5d}**	**138,1**	**35,5**			**35,5**	**2**	**10**	**625**	***0,037***	***0,625***
x3	A_{11d}	138,1	35,5	7,05		7,05	3	*11*	*11687*	*0,695*	*0,688*

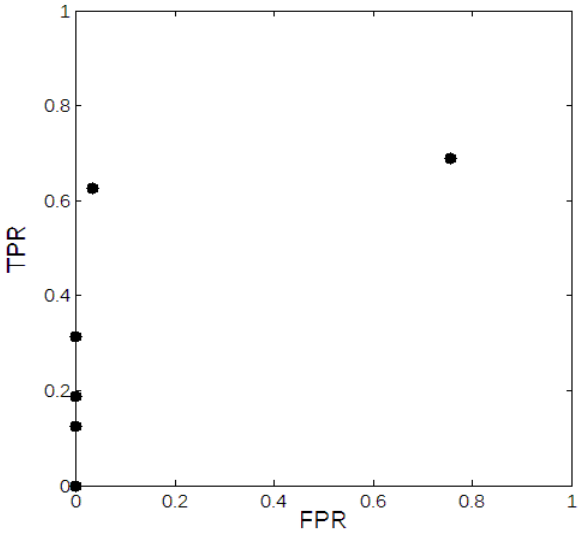


Figure 6-12 ROC curve of the threshold suggested by the above mentioned CTAs. Note that this curve include plots from all stages of CTA analysis; the performance is equal to such degree they cannot be separated visually.

6.4.4 CTA threshold evaluation and final results

The true positive rate (TPR) of the suggested threshold stayed the same for all stages of classification tree analyses (CTAs), meaning that all predicted the same number of triggering rainfall conditions correctly. A small increase in false positive rate (FPR) from 0.0340 to 0.0372, were present in the ROC of stage 3 compared to the other stages. This results in a slightly decreased rainfall threshold performance, as application in an early warning system (EWS) would result in a ~10% increase in false warnings.

The decrease from four to three threshold variables on the other hand is to be considered as an advantage; as all established variables hold uncertainty, fewer variables are desirable. Using highly correlated variables could also pose an increased uncertainty. As the change in FPR is relatively small, the decrease in uncertainty of utilizing fewer variables may compensate for the increase in FPR.

6.5 Discriminant analysis

Linear disccriminant anslysis (LDA) was performed for combinations of a series of 13 different predictors, where each predictor represented a variable of anteceedent precepitation specified in rainfall intencity (mm/day), originated from the merged TMPA-RT rainfall estimates. Such an analysis is hereby called an multiple linear discriminant anaysis (multiple LDA), as many LDAs are executed at once. The predictor combinations correlation coefficent matrix, as presented in Table 6-6, were taken into account not performing any LDAs on highly correlated predictors $\rho \geq 0.8$. Additional criteria were applied to get a correct match of the discriminat function, as listed in relation to equation (6), page 44. Performance of each defined discriminant function where evaluated using missclassification error rates (ERR). The predictor combination producing the lowes ERR was then

selected producing a single LDA output, as presented in Figure 6-13 and Figure 6-14 . This variable proved to work good for finding the best performing predictor combination and was used in an exploratory analysis, as explained in section 5.5.3.4, adjusting ERR for finding the best possible prior probability for optimal fit of a linear discriminant function (linear discriminant function). The results of the LDA methodology will be presented stepwise, as explained in above mentioned section:

Before applying any prior probabilities on multiple LDA analysis, the results of one single multiple DA where assessed. As visible in Figure 6-14 a), the plotted discriminant fuction allready showed a tendecy to sucsessfully discriminate between different classes of data represented by triggering and non-triggering conditions. A problem regarding this initial result is that the discrimanant function seems to preferre having the highes amount of false positives (FP) to true positives(TP), se section 5.5.2.2 for a description of these values. This threhold represented by the discriminant function is not optimal as most landslides are missed, i.e. occuring below the threshold.

6.5.1 Step one of multiple LDA methodology

To establish the best possible threshold represented by a linear discriminant function, a multiple LDA were performed for different ranges of prior probabilities, the best performance for each range were found using misclassification error estimates (ERR). Prior probability may be set to between 0 and 1 for each class of data the LDA try to discriminate. The prior probabilities were always set so that the total prior of the system was 1; e.g. [0.7 0.3], where 0.7 is prior probability of triggering conditions and 0.3 of non-triggering conditions.

Initially exploratory ERR based analysis were conducted at prior intervals of 0.05, from [0.95 0.05] to [0.05 0.95]. The results of this analysis is presented in Figure 6-13 a) and show that the lowest error estimates are expected when the prior of triggering events are close to zero. High priori probability for landslide triggering events produce ERR values a lot higher than for lower values. The best performing LDA threshold for the combination of prior probabilities and predictors were A_{2d} and A_{5d} at a prior of triggering events at 0.05; these data are plotted in Figure 6-14 b).

6.5.2 Step two of DA, scale 0.01

In the next step, the same methodology of exploratory multiple LDA where used, where best performance again was found using ERR. At this step the scale of prior intervals were set to 0.01 and the intervals of exploration were gathered around triggering prior of 0.95, the area from last stage showing the least error for the high prior probability of triggering conditions. The resulting distribution are shown in Figure 6-13 b), indicating a best fit of a rainfall threshold and corresponding discriminant function at ERR = 0.94, where the discriminant function are presented in Figure 6-14 c).

6.5.3 Final step of DA, scale 0.002

Further refinement of the exploratory analysis was conducted at same principals as the previous steps. This was considered the final step, as the scale was starting to get really small and as the results in Figure 6-13 c) and Figure 6-14 d) show, the change in the linear discriminant function was very low.

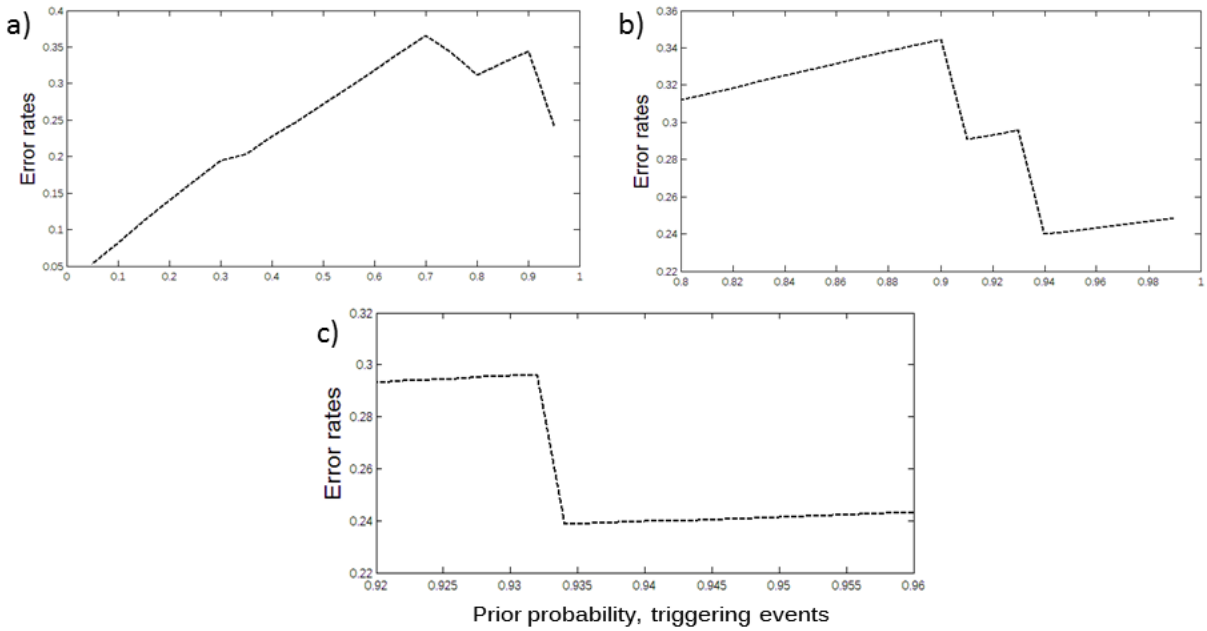


Figure 6-13. Distribution of ERR from the exploratory multiple LDA based on prior probability of triggering events. Figure shows the tree steps of analysis at different scale: a) scale 0.05, b) scale 0.01 and c) scale 0.002.

6.5.4 LDA –results and evaluation

The final proposed threshold in form of a linear discriminate function was

$$0 = 7.100 + (-0.133)x + (-0.161)y\,. \qquad \textbf{(12)}$$

The variables *x* and *y* are represented by the rainfall variables A_{6d} and A_{1d} respectively, and considering equation (5) to (8) this can then be written

$$A_1 = 44 - 0.826\,A_6\,. \qquad \textbf{(13)}$$

Using the established threshold from LDA, rainfall induced landslides are expected to occur if daily rainfall (A_{1d}) exceeds a value of 44 mm/day – 0.826 A_{6d} mm/day, where A_{6d} is the antecedent 6 days accumulated rainfall. Due to the high differences in rainfall distribution and intensities for the landslide triggering days a good performance of this threshold cannot be achieved, and application in an early warning system would not be recommended at this point. As visible in the figures below, there would be too many false positives and false negatives related to this threshold in this area.

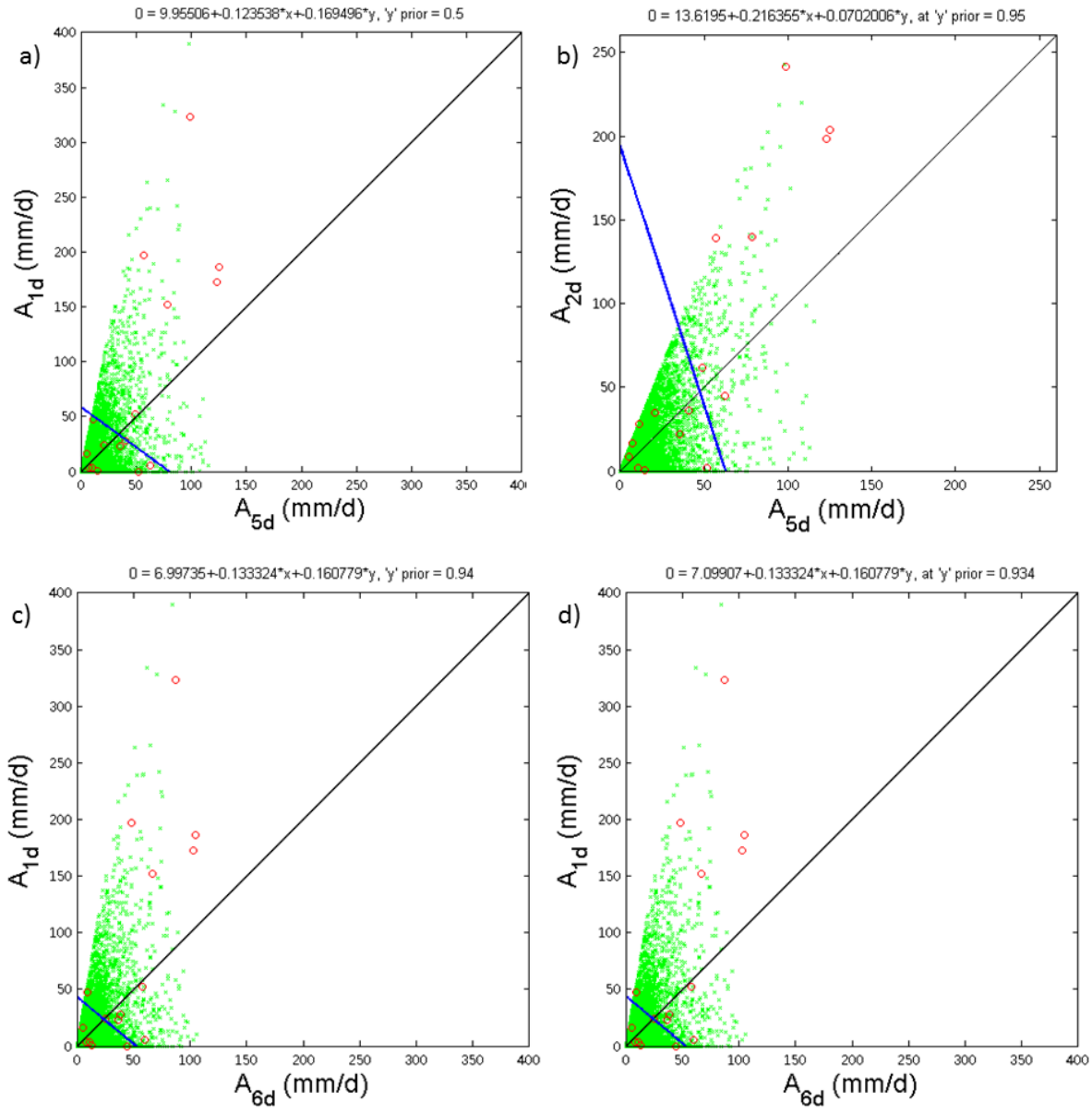


Figure 6-14. Succeeding results of LDA following the successive steps of exploratory multiple discriminant analyses: a) LDA using no prior probabilities (equals [0.50 0.50] prior), b) LDA at step one, c) LDA at step two and d) LDA at final step

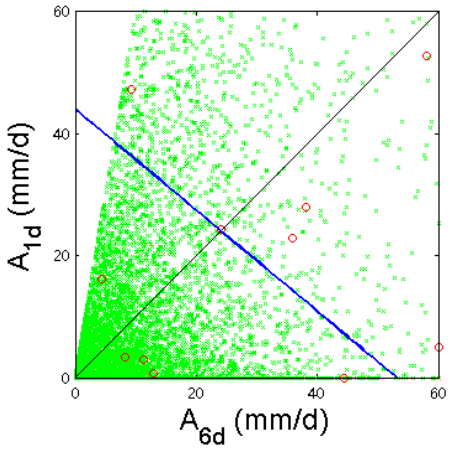


Figure 6-15. Threshold plot for last stage at prior scale 0.002, zoomed in at the area of estimated rainfall threshold

7 DISCUSSION

7.1 Comparison of rain gauge and TMPA-RT data

The limited amount of rain gauge and landslide data for Bangladesh posed a challenge for performing a good relationship study between rain gauge data and NASAs TRMM (Tropical Rainfall Measuring Mission)-based TMPA-RT (Real-Time - TRMM Multi-satellite Precipitation Analysis) data products; 3B41RT and 3B42RT. Still, the study conducted for the rain gauge of Chittagong did prove some important points that may be considered general. There were significant differences in TMPA-RT estimated rainfall values and rain gauge values; this was evidenced both by seasonal differences in rainfall amounts, annual mean rainfall values, rainfall intensity distributions, number of rainy days and by correlation plots. Correlation coefficients were very low. The most severe difference was found in the lack of rainfall amounts estimated by TMPA-RT data in the wettest period of year, June - August. Additionally, the rainfall estimates did not reflect the correct magnitude of the extreme rainfall event events. This problem had a direct consequence on the estimation of high intensity rainfall and was not promising for the threshold evaluation. An overall better performance for the 3B42RT product, based on merged infrared (IR) and passive microwave (PMW) rainfall estimates, compared to the IR based 3B41RT product, resulted that 3B42RT were selected for further studies. The IR based data provides better temporal resolution (1 hour) compared to the 3 hour resolution of the IR-PMW (IRmicro) data, but this was considered less important for this initial study as the available rain gauges only provided daily data.

There was also uncertainty related to the performance of TMPA-RT data at local level; one single grid box of TMPA-RT estimates were compared to one single rain gauge. This resulted in scale challenges, as one TMPA-grid is more than 0.25° x 0.25° (equal to approximately 25 x 25 km in Bangladesh). This partly explains the difference between the rain gauge and the TMPA-RT data.

Application of the SPE data to threshold studies in non-instrumented areas proved to represent a challenge; when too low landslide triggering rainfall values were observed, these could not be validated or be surely disregarded. As a result rainfall thresholds based on 3B42RT data may be established too low, or landslide triggering storms may be missed.

Note to data sets: It was found that the initial rainfall data for the rain gauge Chittagong was shifted with one day, so that the day of the fatal landslide events was registered with 88 mm instead of 425 mm; this was confirmed by e.g. (Ekram et al. 2007, IFRC 2007, Gov. Bangladesh 2010). This affected the TMPA-RT and rain gauge correlation, as this was conducted only on this rain gauge. The correlation plots and coefficients must thus be disregarded. The correlation of triggering event is correct, as this was conducted on a later stage, after the shift in the data was discovered.

7.2 Establishing rainfall thresholds for Bangladesh

The results of classification tree analyses (CTAs) and linear discriminant analyses (LDAs) based on 3B42RT data revealed a highly mixed data set, with a large variety of triggering conditions. The analysis was based on variables of antecedent rainfall, each variable of different rainfall accumulation duration. The rainfall variables were denoted A_i, where i equals 1 day, 2 days, 3 days,

etc. and A_i is the average rainfall intensity of duration i in mm/day. Considering one value of a variable A_i on a certain day as a rainfall condition or data sample, the following was found. The samples were sorted in triggering and non-triggering rainfall conditions (two classes of data), in a highly mixed group with no obvious border between the data classes (see e.g. **Feil! Fant ikke referansekilden.**). Some few data samples of triggering conditions were considered as extreme values. Some triggering conditions were really low. These may be underestimates by the 3B42RT product, such representing some minimum value outliers; this could not be determined due to lack of comparable rain gauge values. Some of the odd values resulting from some of the CTAs were probably caused by these values. As shown in Table A 6 and Figure 6-8, the extreme rainfall events clearly stood out from the rest of the registered events, these are further discussed in section 7.2.2.

7.2.1 Regional thresholds

One of the aims of this study was the assessment of the feasibility for establishing empirical rainfall thresholds encompassing the whole region of Bangladesh using simple rainfall variables. The high diversity of triggering conditions found in this study proves that this is not possible without a high level of uncertainty. Regional thresholds come with some level of uncertainty, regardless of rainfall data source, because of different levels of landslide susceptibility caused by differences in surface geology, vegetation, climate conditions, etc. Implementing such variables in a further assessment of landslide triggering conditions in Bangladesh would be one alternative for continuation. This may be a comprehensive process and requiring higher level of detail for the data. A simple first assessment of these conditions could be applied by rough classifications of physiographic and climate classifications or mean annual precipitation (MAP) as presented in chapter 0 in this study.

7.2.2 Local conditions

From the TMPA-RT data of triggering conditions extracted from the full 3B42RT inventory, as presented in Table A 6 and Figure 6-8, five rainfall series of extreme intensity were found to be clearly distinguishable from the rest of the data. As these were corresponding to the two most fatal landslide events in Bangladesh, these should be noticed. The I-D plots of these TMPA-RT data were clearly similar to the ones of rain gauges from the same area (Figure 6-9). Performance of the TMPA-RT data was generally found to be inadequate, and correlation of extreme daily rainfall did not seem very good based on the data correlation for Chittagong. Although the rainfall intensities were somewhat different, extreme intensities proved to be picked up by the TMPA-RT data. This discovery indicated that satellite based rainfall estimates may be used to successfully predict the most fatal storms, which is also suggested by Habib et al. (2009). A further step in assessing the applicability of TMPA-RT data for landslide prediction in non-instrumented areas is the further study of these storms events, and on non-triggering high intensity rainfall events.

It is suggested to establish thresholds for the hill tracts of Bangladesh alone, as this represents a relatively homogeneous area with similar physiographic and rainfall conditions. This is also where the most fatal events have occurred.

7.2.3 Multivariate analysis; CTA and LDA

Methodology for applying classification tree analysis (CTA) and linear discriminant analysis (LDA) where developed for the programming language Matlab and using functions available in the Matlab Statistical Toolbox. It was found that the methods could be developed and applied relatively easy using this software. The methodology was based on the goal of establishing thresholds with best possible performance for the full region of Bangladesh.

7.2.3.1 CTA

CTA proved to establish a relatively good threshold resulting in a false positive rate of just below 4 %. No susceptibility classes or variables other than antecedent rainfall were added in the analysis; hence is this considered a good result. The threshold resulted in low true positive rate (TPR) of only 0.625, but this included only 6 landslide events of which corresponding rainfall intensities is believed to be underestimated.

A peculiar feature of the CTA results is that the final performance of the established thresholds was virtually the same, independent of which variables were used. As a result the ROC curve was not able to distinguish which threshold performed the best. Using ROC to find which levels of the classification tree to include, and thus indirectly the number of variables to include in the threshold, worked well. As the method is not currently automatized, there are surely room for improvement of the method using the same principals.

The CTA method produces a threshold of a rather untraditional form, as described by Guzzetti et al. (2007) most thresholds are related to intensity-duration (I-D) plots or event duration or intensity. The only similar published rainfall threshold seems to be an event based thresholds using the variables A_{15d} and R (equal to A_{1d} in this thesis) at 3 different intensity levels (Lumb 1975). The method does successfully separate triggering and non-triggering rainfall conditions, to some degree, and further application of the method is suggested.

7.2.3.2 LDA

Numerical assessment of the thresholds established using linear discriminant analysis, using e.g. ROC, was not performed. Misclassification error rates (ERR) were used to identify which combination of rainfall variables resulted in the least misclassified samples, and proved to do this successfully. Applying high prior probability to the landslide triggering events resulted in a threshold fitting seemingly good to the lower bounds of the landslide triggering rainfall intensities. The number of misclassified samples of the final established threshold seems very high (se **Feil! Fant ikke referansekilden.**); a better performance may be found applying quadratic discriminant analysis (QDA).

Despite the low performance of the threshold suggested by this method, the prior probability-based exploratory LDA analysis seems to be a good method for selection of variables. Additionally, changing the prior may be a good tool for applying different levels of susceptibility. This support the use of discriminant analysis for parameter selection and refinement of other threshold types like the I-D

threshold as proposed by Cepeda et al. (2009). Further use of LDA would require application in I-D thresholds. Application of QDA may also prove to be useful.

7.2.3.3 Extreme events: adjustments of methods

A focus on these high intensity events would result in potential refinement of the thresholds, and limit the study area to the eastern hilly areas of Bangladesh. A threshold aiming for the high-intensity storms may simply be achieved selecting less threshold variables. The same application of ROC curves may then not be possible, and it is suggested to re-evaluate what decision criteria or method to be used. Excluding rainfall data below a certain intensity level could be a different approach. The applicability of CTA for this purpose is indicated by results in the current study; by removing the third variable of the already established threshold, five extreme rainfall intensity events remains above the threshold (see Table 6-3 and equation (11). 19 samples of false positives (FP) remain. These are promising numbers, and a further assessment of high-intensity storms would be advisable.

For LDA it is found that the linear discriminant function is drawn towards the extreme rainfall intensities for low prior probability of triggering events (**Feil! Fant ikke referansekilden.**). This suggests that a LDA should be conducted with focus on low triggering event prior, to see how this will affect the linear discriminant function. The same method as used for this study may be applied.

8 CONCLUSION

8.1 Conclusive remarks on the current study

Rainfall induced landslides represent an important part of the threat of natural hazards to human lives, as stressed by recent fatal landslide events in Bangladesh. Especially the densely populated areas in the southeast region of Bangladesh, Chittagong, are frequently hit by fatal events. High intensity rainfall, often of several days, was the main triggering factor.

It was found that satellite precipitation estimates (SPEs), also called satellite rainfall estimates, provide readily available rainfall estimates in relatively high temporal and spatial resolution. There is high uncertainty related to the use of rainfall estimates even when using daily accumulated rainfall values. Comparison of NASAs merged infrared and passive microwave (IR-PMW) TMPA-RT product 3B42RT and a rain gauge station in Chittagong showed no particular correlation.

Strongly heterogeneous triggering data and few registered landslides resulted in difficulties in defining reliable rainfall thresholds for the region of Bangladesh. For application of a regional rainfall threshold, there is a need of establishing different susceptibility levels for different areas; normalized rainfall threshold values could also be tried. Regarding application of TMPA-RT data in rainfall thresholds, it is recommended to focus on high-intensity rainstorms as caused by two recent fatal rainfall-triggered rainfall events in Bangladesh. It was found that 3B42RT data corresponding to these storms showed similar patterns to rain gauge data, only of slightly different rainfall intensity magnitude. Assessment of these storms and similar non-triggering rainstorms are proposed, as these high-intensity storms have caused most of the recent landslide fatalities. It is suggested to try establishing local threshold values for high intensity storms in the areas of highest risk, like Cox's Basar, Bandarban and Chittagong.

Classification tree analysis proved to be a good method for classifying highly mixed triggering and non-triggering rainfall conditions, given the highly heterogeneous 3B42RT data the established threshold performed well. Evaluation of CTA thresholds using Receiver Operating Characteristics (ROC) was successful to some degree. Linear discriminant analysis (LDA), weighting triggering and non-triggering conditions using prior probability and assessing the misclassification error rates, proved to be a feasible method for applying susceptibility levels to rainfall thresholds, but with less success on threshold performance.

Application of the established thresholds in a warning system is not advisable as the results may be considered preliminary. Still, they may provide some valuable initial information on levels of rainfall intensities related to the triggering of landslides in Bangladesh, and provide a basis for further studies on landslide triggering conditions in Bangladesh and further assessment on the application of SPE data for local or small-scale regional rainfall thresholds.

8.2 Suggested continuation of threshold studies in Bangladesh

As the extreme intensity landslide triggering events represent a relatively homogeneous set of rainfall data, and these are causing the most fatal events, it is suggested to focus further studies on assessing these events. It is important to understand which properties of these storms (intensity

peaks, total storm duration, etc.), which variables of landslide susceptibility (vegetation, human influence, soil conditions preceding the event, etc.) are more relevant. It is suggested to apply CTA for analyses of the high intensity events, as the method seems promising. Simultaneously, it is suggested to apply other threshold types like I-D thresholds and event-intensity (E-I) thresholds, to ensure that the optimum threshold criteria is established.

For these storm events it would be interesting to apply the TMPA-RT IR product 3B41RT, or other SPE products of high temporal resolution (e.g. CMORPH), and investigate if similar patterns are found for I-D plots of these data. If these data prove to successfully estimate hourly intensity peaks, they may have potential of predicting temporal occurrence of landslides during a rainstorm.

During autumn of 2010 two automatic rain gauges, providing rainfall measures every 15 minutes were established in Dhaka and Chittagong In collaboration between the Geological Survey of Bangladesh (GSB) and the Norwegian Geotechnical Institute (NGI). Data from these stations should be used in analysis of future landslide in these regions, and may also provide a good opportunity for future correlation between rain gauges and satellite-based rainfall estimates.

9 REFERENCES

Ang, A.H.-S. and Tang, W.H. 2007. *Probability concepts in engineering : emphasis on applications in civil & environmental engineering*. 2nd ed, New York: Wiley. xiii, 406 p. pp.

Baeza, C. and Corominas, J. 2001. Assessment of shallow landslide susceptibility by means of multivariate statistical techniques. *Earth Surface Processes and Landforms* 26, 1251-1263.

Baeza, C., Lantada, N. and Moya, J. 2010. Validation and evaluation of two multivariate statistical models for predictive shallow landslide susceptibility mapping of the Eastern Pyrenees (Spain). *Environmental Earth Sciences* 61.

Banglapedia. 2011. *Earthquake* 2011a [Accsessed: 10 June 2011]. Available at http://www.banglapedia.org/httpdocs/HT/E_0002.HTM.

Banglapedia. 2011. *Landslide* 2011b [Accsessed: 12 June 2011]. Available at http://www.banglapedia.org/httpdocs/HT/L_0057.HTM.

Beguería, S. 2006. Validation and Evaluation of Predictive Models in Hazard Assessment and Risk Management. *Natural Hazards* 37, 315-329.

Behrangi, A., Khakbaz, B., Jaw, T.C., AghaKouchak, A., Hsu, K.L. and Sorooshian, S. 2011. Hydrologic evaluation of satellite precipitation products over a mid-size basin. *Journal of Hydrology* 397, 225-237.

Bellerby, T., Todd, M., Kniveton, D. and Kidd, C. 2000. Rainfall estimation from a combination of TRMM precipitation radar and GOES multispectral satellite imagery through the use of an artificial neural network. *Journal of Applied Meteorology* 39, 2115-2128.

Breien, H., De Blasio, F.V., Elverhoi, A. and Hoeg, K. 2008. Erosion and morphology of a debris flow caused by a glacial lake outburst flood, Western Norway. *Landslides* 5, 271-280.

Caine, N. 1980. The Rainfall Intensity: Duration Control of Shallow Landslides and Debris Flows. *Geografiska Annaler Series A, Physical Geography* 62, 23-27.

Cardinali, M., Galli, M., Guzzetti, F., Ardizzone, F., Reichenbach, P. and Bartoccini, P. 2006. Rainfall induced landslides in December 2004 in south-western Umbria, central Italy: types, extent, damage and risk assessment. *Natural Hazards and Earth System Sciences* 6, 237-260.

Cepeda, J. 2009. *Characterisation and risk management of rainfall-induced landslides*, Faculty of Mathematics and Natural Science, University of Oslo, Oslo.

Cepeda, J., Díaz, M.R., Nadim, F., Høeg, K. and Elverhøi, A. 2009. An empirical threshold model for rainfall-induced landslides: application to the Metropolitan Area of San Salvador, El Salvador. *Characterisation and risk management of rainfall-induced landslides*: University of Oslo, Paper 3 of PhD, 48.

Cepeda, J., Hoeg, K. and Nadim, F. 2010. Landslide-triggering rainfall thresholds: a conceptual framework. *Quarterly Journal of Engineering Geology and Hydrogeology* 43, 69-84.

Chokngamwong, R. and Chiu, L.S. 2008. Thailand daily rainfall and comparison with TRMM products. *Journal of Hydrometeorology* 9, 256-266.

CIA. 2011. *South Asia: Bangladesh* 2011 [Accsessed: 10 June 2011]. Available at https://www.cia.gov/library/publications/the-world-factbook/geos/bg.html.

Crozier, M.J. 1999. Prediction of rainfall-triggered landslides: A test of the antecedent water status model. *Earth Surface Processes and Landforms* 24, 825-833.

Cruden, D.M. and Varnes, D.J. 1996. Landslide types and processes. *In* Turner, A.K. and Schuster, R.L. (eds). *Landslides - Investigation and mitigation*. Transportation Research Board: National Research Council - Special Report 247. Washington, D.C.: National Academy Press, 36-75.

Damiano, E. and Olivares, L. 2010. The role of infiltration processes in steep slope stability of pyroclastic granular soils: laboratory and numerical investigation. *Natural Hazards* 52, 329-350.

Ekram, A.R.M. and Khan, M.M.H. 2008. *Report on the landslides in July 2008 in and around Teknaf upazila, Cox's bazar, Bangladesh.* Dhaka. Geological Survey of Bangladesh.

Ekram, A.R.M., Khan, M.M.H. and Uddin, M.Z. 2007. *Report on the deadliest landslides of 11 June 2007 in and around Chittagong city corporation area, Chittagong, Bangladesh.* Dhaka. Geological Survey of Bangladesh.

Fawcett, T. 2006. An introduction to ROC analysis. *Pattern recognition letters* 27, 861-874.

Fiorillo, Guadagno, Aquino and De, B. 2001. The December 1999 Cervinara landslides: further debris flows in the pyroclastic deposits of Campania (southern Italy). *Bulletin of Engineering Geology and the Environment* 60, 171-184.

Giannecchini, R. 2005. Rainfall triggering soil slips in the southern Apuan Alps (Tuscany, Italy). *Advances in Geosciences* 2, 21-24.

Glade, T., Crozier, M. and Smith, P. 2000. Applying probability determination to refine landslide-triggering rainfall thresholds using an empirical "Antecedent Daily Rainfall Model". *Pure and Applied Geophysics* 157, 1059-1079.

Gov. Bangladesh. 2010. *National Plan for Disaster Management 2010-2015*. 104 pp.

Govi, M. and Sorzana, P.F. 1980. Landslide susceptibility as function of critical rainfall amount in Piedmont basin (Northwestern Italy). *Studia Geomorphologica Carpatho-Balcanica* 14, 43-60.

Guadagno, Martino and Scarascia, M. 2003. Influence of man-made cuts on the stability of pyroclastic covers (Campania, southern Italy): a numerical modelling approach. *Environmental Geology* 43, 371-384.

Guzzetti, F., Peruccacci, S., Rossi, M. and Stark, C.P. 2007. Rainfall thresholds for the initiation of landslides in central and southern Europe. *Meteorology and Atmospheric Physics* 98, 239-267.

Guzzetti, F., Peruccacci, S., Rossi, M. and Stark, C.P. 2008. The rainfall intensity-duration control of shallow landslides and debris flows: an update. *Landslides* 5, 3-17.

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M. and Galli, M. 2006. Estimating the quality of landslide susceptibility models. *Geomorphology* 81, 166-184.

Habib, E., Henschke, A. and Adler, R.F. 2009. Evaluation of TMPA satellite-based research and real-time rainfall estimates during six tropical-related heavy rainfall events over Louisiana, USA. *Atmospheric Research* 94, 373-388.

Highland, L.M. and Bobrowsky, P. 2008. *The landslide handbook—A guide to understanding landslides*. 1.0 ed, Reston, Virginia: U.S. Geological Survey. 129 pp.

Hong, Y., Adler, R. and Huffman, G. 2006. Evaluation of the potential of NASA multi-satellite precipitation analysis in global landslide hazard assessment. *Geophysical Research Letters* 33.

Huffman, G.J., Adler, R.F., Bolvin, D.T., Gu, G.J., Nelkin, E.J., Bowman, K.P., Hong, Y., Stocker, E.F. and Wolff, D.B. 2007. The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology* 8, 38-55.

Huffman, G.J., Adler, R.F., Bolvin, D.T., Nelkin, E.J., Hossain, F. and Gebremichael, M. 2009. The TRMM Multi-satellite Precipitation Analysis (TMPA). *Satellite Rainfall Applications for Surface Hydrology*, 3–22.

Huffman, G.J. and Bolvin, D.T. 2010. Real-Time TRMM Multi-Satellite Precipitation Analysis Data Set Documentation. 40.

Hutchinson, J.N. 1988. General report: Morphological and geotechnical parameters of landslides in relation to geology and hydrology. *Proc, Fifth International Symposium on Landslides*, A.A. Balkema. Rotterdam, Netherlands, 3-35.

IFRC. 2011. *Bangladesh: Landslides and floods.* Information Bulletin No.01/ 2007 [Accsessed: 10 June 2011]. Available at http://reliefweb.int/sites/reliefweb.int/files/resources/2F7AAB0CC77BAF7FC12572FA0044B74D-Full_Report.pdf.

IFRC. 2011. *Bangladesh: Landslides* DREF operation 2010a [Accsessed: 10 June 2011]. Available at http://www.ifrc.org/docs/appeals/10/DREF_operation_MDRBD007.pdf.

IFRC. 2010b. *World Disaster Report 2010* 215 pp.

IRIN. 2011. *BANGLADESH: 70,000 people vulnerable to landslides* 2008 [Accsessed: 10 June 2011]. Available at http://www.irinnews.org/Report.aspx?ReportId=79406.

Islam, M.N. and Uyeda, H. 2008. Vertical variations of rain intensity in different rainy periods in and around Bangladesh derived from TRMM observations. *International Journal of Climatology* 28, 273-279.

Iverson, R.M. 2000. Landslide triggering by rain infiltration. *Water Resources Research* 36, 1897-1910.

Jakob, M. and Weatherly, H. 2003. A hydroclimatic threshold for landslide initiation on the North Shore Mountains of Vancouver, British Columbia. *Geomorphology* 54, 137-156.

Kamal, A.S.M.M. 2008. Seismic hazard assessment for chittagong city corporation area, Bangladesh. *International Geological Congress*, August 6-14th, Oslo.

Kidd, C., Kniveton, D.R., Todd, M.C. and Bellerby, T.J. 2003. Satellite rainfall estimation using combined passive microwave and infrared algorithms. *Journal of Hydrometeorology* 4, 1088-1104.

Kirschbaum, D., Adler, R., Hong, Y., Hill, S. and Lerner-Lam, A. 2010. A global landslide catalog for hazard applications: method, results, and limitations. *Natural Hazards* 52, 561-575.

Lacasse, S. and Nadim, F. 2009. *Landslide Risk Assessment and Mitigation Strategy*. Edited by Sassa, K. and Canuti, P. Landslides - Disaster Risk Reduction, Berlin: Springer-Verlag Berlin. 31-61 pp.

Lacasse, S., Nadim, F. and Kalsnes, B. 2010. Living with Landslide Risk. *Geotechnical Engineering Journal of the SEAGS & AGSSEA* 41.

Lachenbruch, P.A. 1968. On Expected Probabilities of Misclassification in Discriminant Analysis, Necessary Sample Size, and a Relation with the Multiple Correlation Coefficient. *Biometrics* 24, 823-834.

Lachenbruch, P.A. and Goldstein, M. 1979. Discriminant Analysis. *Biometrics* 35, 69-85.

Landsberg, E.H., ed. 1981. *World Survey of Climatology*. Ed. Takahashi, K. and Arakawa, H. Vol. 9, *Climates of Southern and Western Asia*. Amsterdam: Elseviar.

Lee, S., Ryu, J.-H., Lee, M.-J. and Won, J.-S. 2003. Use of an artificial neural network for analysis of the susceptibility to landslides at Boun, Korea. *Environmental Geology* 44, 820-833.

Lockwood, J.G. 1974. *World climatology*, Norwich: Edward Arnold.

Lumb, P. 1975. Slope failures in Hong Kong. *Quarterly Journal of Engineering Geology and Hydrogeology* 8, 31-65.

Mahmood, A.B. and Khan, M.H. 2008. Landslide Vulnerability of Bangladesh Hills and Sustainable Management Options: A Case Study of 2007 Landslide in Chittagong City. *Paper prepared for the International Seminar on "Management and Mitigation of Water Induced Disaster"*, 21-22 April 2008, Kathmundu, Nepal.

Malet, J., Remaître, A., Maquaire, O. and Locat, J. 2003. Dynamics of distal debris-flows induced in clayey earthflows. Implications for hazard assessment, 341-348.

NASA. 2011. *Landslide database for 2003, 2007, 2008 and 2009* 2011 [Accsessed: 12 June 2011]. Available at http://trmm.gsfc.nasa.gov/publications_dir/landslide_catalog_2003_2007_2008_2009.xls.

NIST/SEMATECH. 2011. *e-Handbook of Statistical Methods*. *http://wwwitlnistgov/div898/handbook/ June 2011*.

NOAA CPC. 2011. *NOAA CPC Morphing Technique ("CMORPH")* 2011 [Accsessed: 21 March 2011]. Available at http://www.cpc.ncep.noaa.gov/products/janowiak/cmorph_description.html.

NOAA STAR. 2011. *STAR Satellite Rainfall Estimates Validation Algorithms* STAR 2011 [Accsessed: 21.03 2011]. Available at http://www.star.nesdis.noaa.gov/smcd/emb/ff/validationAlgorithms.php.

Olivares, L. and Damian, E. 2007. Postfailure mechanics of landslides: Laboratory investigation of flowslides in pyroclastic soils. *Journal of Geotechnical and Geoenvironmental Engineering* 133, 51-62.

Pasuto, A. and Silvano, S. 1998. Rainfall as a trigger of shallow mass movements. A case study in the Dolomites, Italy. *Environmental Geology* 35, 184-189.

Peel, M.C., Finlayson, B.L. and McMahon, T.A. 2007. Updated world map of the Koppen-Geiger climate classification. *Hydrology and Earth System Sciences* 11, 1633-1644.

Petley, D., Hearn, G., Hart, A., Rosser, N., Dunning, S., Oven, K. and Mitchell, W. 2007. Trends in landslide occurrence in Nepal. *Natural Hazards* 43, 23-44.

Rossi, M., Guzzetti, F., Reichenbach, P., Mondini, A.C. and Peruccacci, S. 2010. Optimal landslide susceptibility zonation based on multiple forecasts. *Geomorphology* 114, 129-142.

SAARC. 2007. *South Asia Disaster Report 2007*. SAARC Disaster Management Sentre (SDMS) - New Dehli. 213 pp.

Santacana, N., Baeza, B., Corominas, J., De Paz, A. and Marturia, J. 2003. A GIS-based multivariate statistical analysis for shallow landslide susceptibility mapping in La Pobla de Lillet area (Eastern Pyrenees, Spain). *Natural Hazards* 30, 281-295.

Sapiano, M.R.P. and Arkin, P.A. 2009. An Intercomparison and Validation of High-Resolution Satellite Precipitation Estimates with 3-Hourly Gauge Data. *Journal of Hydrometeorology* 10, 149-166.

Sarker, J., Ansary, M., Rahman, M. and Safiullah, A. 2010. Seismic hazard assessment for Mymensingh, Bangladesh. *Environmental Earth Sciences* 60, 643-653.

Scofield, R.A. and Kuligowski, R.J. 2003. Status and outlook of operational satellite precipitation algorithms for extreme-precipitation events. *Weather and Forecasting* 18, 1037-1051.

Shen, Y., Xiong, A.Y., Wang, Y. and Xie, P.P. 2010. Performance of high-resolution satellite precipitation products over China. *Journal of Geophysical Research-Atmospheres* 115.

Sohn, B.J., Han, H.J. and Seo, E.K. 2010. Validation of Satellite-Based High-Resolution Rainfall Products over the Korean Peninsula Using Data from a Dense Rain Gauge Network. *Journal of Applied Meteorology and Climatology* 49, 701-714.

STAR. 2011. *STAR Satellite Rainfall Estimates, Hydro-Estimator - Digital Global Data* 2011 [Accsessed: 28 June 2011]. Available at http://www.star.nesdis.noaa.gov/smcd/emb/ff/digGlobalData.php.

StatSoft Inc. 2011. *Electronic Statistics Textbook*. Tulsa, OK: Statsoft.

Su, F., Hong, Y. and Lettenmaier, D.P. 2008. Evaluation of TRMM Multisatellite Precipitation Analysis (TMPA) and its utility in hydrologic prediction in the La Plata Basin. *Journal of Hydrometeorology* 9, 622-640.

Tang, L., Hossain, F. and Huffman, G.J. 2010. Transfer of Satellite Rainfall Uncertainty from Gauged to Ungauged Regions at Regional and Seasonal Time Scales. *Journal of Hydrometeorology* 11, 1263-1274.

The Mathworks Inc. 2011. *Product Documentation - MATLAB* 2011 [Accsessed: 21 June 2011]. Available at http://www.mathworks.com/help/toolbox/stats/classify.html.

Tiranti, D. and Rabuffetti, D. 2010. Estimation of rainfall thresholds triggering shallow landslides for an operational warning system implementation. *Landslides*, 1-11.

Tsaparas, I., Rahardjo, H., Toll, D.G. and Leong, E.C. 2002. Controlling parameters for rainfall-induced landslides. *Computers and Geotechnics* 29, 1-27.

Turner, A.K. and Schuster, R.L., eds 1996. *Landslides - Investigation and mitigation*, *Transportation Research Board: National Research Council - Special Report 247*. Washington, D.C.: National Academy Press.

USGA. 2011. *Earthquake Hazard Program - Earthquakes*, 25 August, 2010 2011 [Accsessed: 23 June 2011]. Available at http://earthquake.usgs.gov/earthquakes/.

van Westen, C.J. and Daag, A.S. 2005. Analysing the relation between rainfall characteristics and lahar activity at Mount Pinatubo, Philippines. *Earth Surface Processes and Landforms* 30, 1663-1674.

Varnes, D.J. 1978. *Slope Movement Types and Processes*. *I* Schuster, R.L. and Krizek, R.J. (red). *Special Report 176: Landslides: Analysis and Control* Washington, DC: Transportation and Road Research Board, National Academy of Science.

Vicente, G.A., Scofield, R.A. and Menzel, W.P. 1998. The operational GOES infrared rainfall estimation technique. *Bulletin of the American Meteorological Society* 79, 1883-1898.

Wan, S., Lei, T.C. and Chou, T.Y. 2010. A novel data mining technique of analysis and classification for landslide problems. *Natural Hazards* 52, 211-230.

Wan, S.A. and Lei, T.C. 2009. A knowledge-based decision support system to analyze the debris-flow problems at Chen-Yu-Lan River, Taiwan. *Knowledge-Based Systems* 22, 580-588.

Wieczorek, G.F. 1987. Effect of rainfall intensity and duration on debris flows in central Santa Cruz Mountains, California. *In* Costa, J.E. and Wieczorek, G.F. (eds). *Debris flows/avalanches; process, recognition, and mitigation*. Boulder, CO, United States.: Geological Society of America (GSA)

Wieczorek, G.F. and Glade, T. 2005. Climatic factors influencing occurrence of debris flows. *In* Jakob, M. and Hungr, O. (eds). *Debris-flow Hazards and Related Phenomena*. Springer Praxis Books: Springer Berlin Heidelberg, 325-362.

Wieczorek, G.F. and Sarmiento, J. 1988. Rainfall, piezometric levels, and debris flows near La Honda, California, in storms between 1975 and 1983. *In* Ellen, S.D. and Wieczorek, G.F. (eds). *Landslides, floods, and marine effects of the storm of January 3-5, 1982, in San Francisco Bay region, California*. Reston, VA, United States: U. S. Geological Survey

Wilson, R.C. 1989. Rainstorms, pore pressures, and debris flows; a theoretical framework. *In* Sadler, P.M. and Morton, D.M. (eds). *Landslides in a semi-arid environment with emphasis on the inland valleys of Southern California*. Redland, Ca, United States.: The Inland Geological Society 2, 101-117.

Wilson, R.C. 1997. *Normalizing rainfall/debris-flow thresholds along the US Pacific coast for long-term variations in precipitation climate*. Edited by Chen, C.L. Debris-Flow Hazards Mitigation: Mechanics, Prediction & Assessment, New York: Amer Soc Civil Engineers. 32-43 pp.

Yeon, Y.K., Han, J.G. and Ryu, K.H. 2010. Landslide susceptibility mapping in Injae, Korea, using a decision tree. *Engineering Geology* 116, 274-283.

Yong, B., Ren, L.L., Hong, Y., Wang, J.H., Gourley, J.J., Jiang, S.H., Chen, X. and Wang, W. 2010. Hydrologic evaluation of Multisatellite Precipitation Analysis standard precipitation products in basins beyond its inclined latitude band: A case study in Laohahe basin, China. *Water Resources Research* 46.

Appendix A. Landslide- and rainfall data

Initial landslide and rainfall data

These initial data were provided by Mr. Reshad Ekram (Director Geological Survey of Bangladesh) and ADPC (Asian Disaster Preparedness Centre). Note that these data are extracts of the available data. Requests for access of these data may be addressed to Mr. Reshad Ekram (reshadekram@gmail.com).

Table A 1. Landslide events and rain gauge data provided by the Asian Disaster and Preparedness Centre (ADCP)

Location	Date of Occurrence	6 days before	5 days before	4 days before	3 days before	2 days before	1 day before	**The day of disaster**	1day after
1. Chittagong	11 June 2007	3	23	22	4	42	3	**88**	425
2. Teknaf	3 July 2008	0	0	0	0	61	33	**367**	53
3. Teknaf	6 July 2008	0	61	33	367	53	73	**209**	144
4. Srimongal	18 May 2009	0	0	90	0	0	13	**328**	51
5. Teknaf Sadar and surrounding areas	15 June 2010	0	41	0	15	75	78	**132**	186
6.Cox's Bazar Sadar and Ukhia Upazila	15 June 2010	0	41	0	15	75	78	**132**	186

Table A 2. Daily rainfall data (mm/day) from rain gauge station Chittagong situated at longitude 91.82 and latitude 22.27.

Year	Month	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	Total
2002	7	46	37	4	17	129	59	7	3	2	13	22	0	0	14	3	4	3	2	0	23	156	119	89	71	0	19	0	1	0	77	0	920
2002	8	0	0	39	20	35	24	13	4	11	54	6	5	1	1	25	14	85	15	0	6	0	3	74	2	0	1	2	1	15	0	0	456
2002	9	0	0	5	5	0	36	0	3	17	0	1	0	0	0	2	20	2	5	0	0	18	22	5	0	0	0	1	1	2	0	145	
2002	10	0	0	0	0	0	0	2	1	0	5	0	0	0	1	47	22	0	0	0	49	2	0	0	0	0	0	0	0	0	0	0	129
2002	11	0	0	0	0	0	0	0	0	0	0	0	23	69	21	0	0	0	0	0	0	0	0	0	0	0	2	3	0	2	8	128	
2002	12	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10
2003	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2003	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
2003	3	0	0	0	0	0	0	0	0	0	0	0	0	8	1	1	16	7	20	0	0	0	0	0	0	0	0	0	0	0	0	0	53
2003	4	97	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	11	26	0	0	12	0	0	0	0	0	0	17	0	167	
2003	5	0	0	8	55	0	1	0	0	0	0	0	0	0	1	0	31	25	0	0	0	26	8	0	8	25	0	0	10	0	0	0	198
2003	6	0	0	0	1	14	53	8	133	5	63	116	34	0	6	0	5	8	2	0	2	95	175	63	1	2	51	206	43	20	103	1209	
2003	7	10	54	0	30	0	1	4	6	0	2	4	41	17	0	0	0	0	0	12	8	36	5	0	10	28	0	0	4	31	64	5	372
2003	8	7	0	0	0	0	12	0	0	0	53	27	80	13	2	4	1	12	0	1	28	8	2	1	0	0	0	0	0	3	1	5	260
2003	9	2	40	23	0	0	4	8	4	2	0	60	16	5	0	0	0	2	1	2	0	4	0	2	10	0	0	5	11	9	0	210	
2003	10	0	0	8	0	0	0	11	4	43	33	71	0	0	0	0	3	0	0	0	0	0	0	4	42	4	0	0	0	11	0	0	234
2003	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2003	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	37	25	0	0	0	0	0	0	0	0	0	0	66
2004	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2004	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
2004	3	0	0	0	0	0	0	0	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	6
2004	4	0	0	0	0	5	1	7	0	4	26	21	0	0	0	0	0	0	0	41	0	0	26	0	34	0	0	0	0	0	0	165	
2004	5	0	0	0	0	0	0	0	0	0	0	0	12	122	0	0	24	0	5	0	4	0	56	1	47	0	0	0	0	4	0	0	275
2004	6	3	1	0	62	0	0	9	7	1	1	37	18	2	8	2	0	0	52	119	94	91	8	18	0	50	18	37	0	0	0	638	
2004	7	0	5	17	0	5	1	87	75	264	50	72	52	0	0	15	4	15	44	6	108	59	0	5	0	0	15	0	4	2	0	0	905
2004	8	0	0	0	1	13	6	3	1	2	0	0	3	41	26	4	0	2	1	1	5	0	0	0	0	1	2	11	0	2	1	16	142
2004	9	34	0	5	16	13	0	2	1	1	10	15	129	117	154	32	1	7	1	0	0	1	4	15	0	27	4	0	1	0	0	590	
2004	10	0	20	0	3	8	38	18	95	3	0	0	0	0	0	9	6	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	203
2004	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2004	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2005	1	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5
2005	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
2005	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	7	29	14	1	0	0	0	0	0	57
2005	4	0	0	0	0	0	0	0	0	0	0	43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	1	58	4	1	121	
2005	5	0	0	0	0	0	11	0	0	6	24	3	0	7	4	0	0	0	16	0	4	0	26	15	13	19	0	0	0	0	0	0	148
2005	6	0	0	0	0	24	0	0	0	1	0	21	1	0	0	0	0	3	0	0	0	0	0	3	20	19	4	11	45	50	31	233	
2005	7	44	90	58	18	11	5	6	0	0	5	1	11	0	83	103	37	2	1	1	12	94	0	0	0	0	1	7	1	0	5	0	596
2005	8	0	22	3	12	33	101	41	53	13	1	14	4	0	5	26	21	0	0	1	36	60	16	42	47	96	0	0	5	1	1	0	654
2005	9	0	23	6	18	2	0	0	0	2	2	20	2	3	19	0	14	0	0	8	2	64	1	0	15	164	0	6	13	0	2	386	
2005	10	31	11	0	0	0	0	0	0	0	0	6	0	0	0	0	1	0	2	0	0	8	11	3	25	1	0	0	0	0	0	1	100
2005	11	1	18	0	0	0	0	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	
2005	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	3	0	0	0	0	0	0	0	8

Year	Month	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	Total
2006	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2006	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
2006	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2006	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	1	0	0	0	0	0	0	0	0	0	40	0	50	
2006	5	0	0	0	0	0	1	20	7	0	0	15	23	28	0	4	0	0	0	0	19	47	0	0	14	0	16	210	116	69	233	37	859
2006	6	13	5	0	2	13	10	42	10	1	54	35	0	0	4	4	2	0	0	0	5	30	15	0	0	0	20	17	5	0	1	288	
2006	7	0	0	0	30	30	7	19	84	61	38	24	21	2	0	1	4	0	0	19	13	0	2	8	29	0	0	42	50	1	10	0	495
2006	8	9	0	0	0	3	0	8	3	8	4	10	11	0	0	0	0	0	0	8	0	23	2	9	9	11	10	0	6	1	3	0	138
2006	9	2	8	11	0	0	0	0	26	5	0	40	14	0	0	2	11	0	0	13	8	66	69	95	10	2	0	0	0	50	6	438	
2006	10	1	0	0	0	6	9	0	0	1	28	0	25	2	9	0	0	0	0	0	1	2	0	0	0	0	0	0	7	0	0	0	91
2006	11	0	0	0	0	0	0	0	0	15	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	
2006	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2007	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2007	2	0	0	0	0	16	0	3	0	0	0	0	0	0	7	21	2	0	0	0	0	0	0	0	0	0	0	0	0	49			
2007	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	4
2007	4	0	0	0	0	0	0	0	0	0	13	17	53	27	0	0	0	0	0	0	0	3	14	66	4	19	0	0	7	0	0	223	
2007	5	0	0	0	0	0	56	0	8	0	0	0	0	0	9	61	2	0	0	0	37	43	44	0	3	0	0	0	0	0	0	0	263
2007	6	0	0	0	0	3	23	22	4	42	3	88	425	48	0	71	88	42	8	29	0	2	1	0	4	0	0	5	6	10	29	953	
2007	7	29	55	8	6	34	41	70	0	18	0	0	0	2	0	11	18	50	81	51	206	44	64	29	6	1	37	53	11	4	1	0	930
2007	8	4	0	5	0	3	4	0	0	0	19	31	6	62	21	58	64	100	0	0	1	0	1	13	4	73	73	10	1	4	30	2	589
2007	9	5	1	0	0	7	35	84	160	40	50	76	0	0	0	0	0	0	0	0	0	1	0	12	13	2	0	0	0	1	25	512	
2007	10	10	0	0	20	19	20	40	62	104	0	2	1	3	0	17	49	284	2	0	0	2	0	0	0	0	0	0	0	0	0	0	635
2007	11	0	0	20	7	0	0	0	0	0	36	1	0	0	4	19	82	13	0	0	0	0	0	0	0	0	0	0	0	0	0	182	
2007	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2008	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	55	1	0	0	0	56
2008	2	0	0	0	0	0	0	0	0	0	9	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13		
2008	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13	0	0	1	0	0	0	0	0	0	0	14
2008	4	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
2008	5	0	0	0	13	5	85	0	0	0	0	0	0	0	11	0	0	2	10	57	26	0	0	0	0	0	2	0	4	0	23	34	272
2008	6	0	0	1	0	14	39	0	0	5	6	0	4	1	0	4	0	195	204	23	17	1	42	0	0	0	0	12	0	26	24	618	
2008	7	82	84	55	135	28	38	69	20	10	14	14	3	24	228	47	13	25	8	3	18	0	0	0	6	7	1	9	16	0	1	4	962
2008	8	0	0	0	0	2	0	0	2	15	5	11	106	0	8	31	2	38	104	129	11	2	2	4	2	0	5	0	55	197	76	2	809
2008	9	0	0	15	6	32	23	0	0	0	0	9	23	2	0	1	14	0	5	0	5	13	4	31	15	7	0	1	40	0	20	266	
2008	10	0	21	15	0	0	0	59	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	34	7	24	55	0	0	0	216
2008	11	0	0	0	0	8	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	43	
2008	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

[Note : ****=missing data,0=rain nil]

Table A 3. Landslide inventory from combining all sources of landslide data for Bangladesh (explained in chapter 5.1.1 Landslide inventory)

New ID	Org. ID	Location	District	Region	Closest Rain Gauge	Latitude	Longitude	Sample type	Source	Type of landslide*	Casual-ties	Confidence radius (km)	Year	Month	Day
1	1	Fakirpura, Teknaf	Cox's Bazar	Chittagong	Cox's Bazar RG	20,874	92,291	Single LS	GSB	-		0,1	2 008	7	3
2	2	Dhumperang, Teknaf	Cox's Bazar	Chittagong	Cox's Bazar RG	20,872	92,290	Single LS	GSB	-		0,1	2 008	7	3
3	3	Puran Pollanpara, Teknaf	Cox's Bazar	Chittagong	Cox's Bazar RG	20,875	92,292	Single LS	GSB	-		0,1	2 008	7	6
4	1	Baizid Bostami	Chittagong	Chittagong	Chittagong RG	22,388	91,818	Single LS	GSB	-		0,1	2 007	6	11
5	2	Kushumbagh	Chittagong	Chittagong	Chittagong RG	22,355	91,822	Single LS	GSB	-		0,1	2 007	6	11
6	3	Lalkhanbazar	Chittagong	Chittagong	Chittagong RG	22,347	91,816	Single LS	GSB	-		0,1	2 007	6	11
7	4	Pahartoli	Chittagong	Chittagong	Chittagong RG	22,350	91,800	Single LS	GSB	Wall collapse		0,1	2 007	6	11
8	5	Chittagong University Campus	Chittagong	Chittagong	Chittagong RG	22,469	91,789	Single LS	GSB	-		0,1	2 007	6	11
9	6	Sikandarpara	Chittagong	Chittagong	Chittagong RG	22,435	91,798	Single LS	GSB	-		0,1	2 007	6	11
10	7	Sikandarpara	Chittagong	Chittagong	Chittagong RG	22,435	91,798	Single LS	GSB	-		0,1	2 007	6	11
11	8	Sikandarpara	Chittagong	Chittagong	Chittagong RG	22,434	91,798	Single LS	GSB	Debris flow		0,1	2 007	6	11
12	9	Lebubagan	Chittagong	Chittagong	Chittagong RG	22,417	91,810	Single LS	GSB	-		0,1	2 007	6	11
13	10	Lebubagan	Chittagong	Chittagong	Chittagong RG	22,416	91,810	Single LS	GSB	-		0,1	2 007	6	11
14	11	Lebubagan	Chittagong	Chittagong	Chittagong RG	22,416	91,794	Single LS	GSB	-		0,1	2 007	6	11
15	12	Lebubagan	Chittagong	Chittagong	Chittagong RG	22,415	91,810	Single LS	GSB	-		0,1	2 007	6	11
16	13	Lebubagan	Chittagong	Chittagong	Chittagong RG	22,415	91,811	Single LS	GSB	-		0,1	2 007	6	11
17	14	Lebubagan	Chittagong	Chittagong	Chittagong RG	22,415	91,811	Single LS	GSB	-		0,1	2 007	6	11
18	15	Kechuarghona	Chittagong	Chittagong	Chittagong RG	22,422	91,808	Single LS	GSB	-		0,1	2 007	6	11
19	16	Kechuarghona	Chittagong	Chittagong	Chittagong RG	22,424	91,806	Single LS	GSB	-		0,1	2 007	6	11
20	17	Kechuarghona	Chittagong	Chittagong	Chittagong RG	22,424	91,806	Single LS	GSB	-		0,1	2 007	6	11
21	18	Kechuarghona	Chittagong	Chittagong	Chittagong RG	22,424	91,804	Single LS	GSB	-		0,1	2 007	6	11
22	19	Workshopghona	Chittagong	Chittagong	Chittagong RG	22,421	91,807	Single LS	GSB	Debris flow		0,1	2 007	6	11
23		Hathazari	Chittagong	Chittagong	Chittagong RG	22,396	91,774	Single LS	news	-		5	2 007	6	11
24	1	Chittagong and outer northern areas	Chittagong	Chittagong	Chittagong RG	22,386	91,805	LS event	ADPC	Debris flow (DF)		5	2 007	6	11
25	2	Teknaf	Cox's Bazar	Chittagong	Chittagong RG	20,873	91,290	LS event	ADPC	DF and rock fall		5	2 008	7	3
26	3	Teknaf	Cox's Bazar	Chittagong	Chittagong RG	20,875	92,292	LS event	ADPC	DF and rock fall		5	2 008	7	6
27	4	Srimangal city	Maulvibazar	Sylhet	Sylhet RG	24,306	91,729	LS event	ADPC	Debris flow		5	2 009	5	18
28	5	Teknaf Sadar and surroundings	Cox's Bazar	Chittagong	Cox's Bazar RG	20,871	92,293	LS event	ADPC	DF and rock fall		5	2 010	6	15
29	6	Ukhia Upazila	Cox's Bazar	Chittagong	Cox's Bazar RG	21,229	92,169	LS event	ADPC	DF and rock fall		5	2 010	6	15
30	6	Cox's Bazar, 'Himchori drive	Cox's Bazar	Chittagong	Cox's Bazar RG	21,397	92,008	LS event	ADPC	DF and rock fall		5	2 010	6	15
36	90	Chittagong	Chittagong	Chittagong	Chittagong RG	22,416	91,833	LS event	NASA	-	128	5	2 007	6	11
37	91	Rangamati	Rangamati	Chittagong	Chittagong RG	22,637	92,145	LS event	NASA	Mudslide	3	10	2 007	6	11
38	256	Nabinagar in Chittagong	Brahmanbaria	Chittagong	Daka/Comilla RG	23,892	90,973	Single LS	NASA	Earth fall	2	4	2 007	9	10
39	311	Betbunia	Khulna	Khulna	Kuna/Satkira RG	22,533	89,400	Single LS	NASA	Mudslide	3	5	2 007	10	15
40	320	Madamdevihat and Dolhajra	Feni	Chittagong	Majidicort RG	22,894	91,533	Single LS	NASA	-		150	2 007	10	19
41	224	Cox's Bazar, Moheshkhali,Teknaf	Cox's Bazar	Chittagong	Cox's Bazar RG	21,440	92,000	LS event	NASA	-	20	4	2 008	7	3
42	281	Meghalaya, Kalaphahar kills	Sylhet	Sylhet	Sylhet RG	24,896	91,902	LS event	NASA	-		45	2 008	8	1
43	325	Lalkhan Bazar	Cox's Bazar	Chittagong	Chittagong RG	22,340	91,824	LS event	NASA	-	11	10	2 008	8	18
44	339	Motijharna of Lalkhan Bazar	Cox's Bazar	Chittagong	Chittagong RG	22,344	91,819	Single LS	NASA	-		10	2 008	8	23
45	71	Sreemangal upazila, Moulvibazer.	Maulvibazar	Sylhet	Srimangal RG	24,308	91,733	Single LS	NASA	Mudslide	6	10	2 009	5	18
46	108	Lama village, Bandarban district	Bandarban	Chittagong	Chittagong RG	22,225	92,190	LS event	NASA	-	10	4	2 009	7	31

Table A 4. Example from TMPA-RT rainfall inventory, here presenting hourly 3B41RT data

	Pos1	Pos2	Pos3	Pos4	Pos5	Pos6	Pos7	Pos8	Pos9	Pos10	Pos11	Pos12
'TMPA_lat'	22,625	23,875	20,875	24,375	22,875	24,875	22,375	22,625	22,125	21,375	21,125	20,875
'TMPA_long'	89,375	90,875	91,375	91,625	91,625	91,875	91,875	92,125	92,125	92,125	92,125	92,375
'2007010100'	0	0	0	0	0	0	0	0	0	0	0	0
'2007010101'	0	0	0	0	0	0	0	0	0	0	0	0
'2007010102'	0	0	0	0	0	0	0	0	0	0	0	0
'2007010103'	0	0	0	0	0	0	0	0	0	0	0	0
'2007010104'	0	0	0	0	0	0	0	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
'2008100107'	0	0	0	0	0	0	0	0	0	0	0	0
'2008100108'	0	0	0	0	0	0	0	0	0	0	0	2,71
'2008100109'	0	0	0	0	0	0	0	0	0	0,12	3,75	3,93
'2008100110'	0	0	0,4	0	0	0	0	0	0	2,6	4,57	7,43
'2008100111'	0	0	2,59	0	0	0	0	0	0,59	5,51	6,47	2,81
'2008100112'	0	0	3,18	0	0	0	0,88	9,27	1,54	4,06	3,07	1,42
'2008100113'	0	0	1,15	0	6,74	0	3,8	8,88	3,83	3,89	1,98	0,27
'2008100114'	0	0	0,8	0	10,61	0	3,8	7,92	4,17	4,32	2,6	0
'2008100115'	0	0	3,75	0	7,69	0	5,65	3,56	3,56	3,89	2,6	0
'2008100116'	0	0	10,05	0	3,04	0	1,73	1,29	2,09	3,89	3,69	0,12
'2008100117'	0	0	5,23	0,33	0,83	1,12	0,86	0,42	0,96	3,2	3,06	0,2
'2008100118'	0	0,04	4,3	0	0,2	2,27	0,18	0	0,29	1,87	1,87	0
'2008100119'	0,05	0	3,8	0	0	0,55	0	0	0	0,21	0,5	0
'2008100120'	0	0	7,49	0	0	0,05	0	0	0	0,3	0,5	0
'2008100121'	0	0	3,05	0	0	0	0	0	0	0,24	0,3	0
'2008100122'	0	0	2,22	0	0	0	0,04	0	0,21	0,8	0,52	0
'2008100123'	0	0	1,01	0	0	0	0,33	0	0,64	0,46	0,34	0
'2008100200'	0	0	0,45	0	0	0	0	0	0,38	0,31	0,12	0
'2008100201'	0	0	0,48	0	0	0	0	0	0,23	0	0,03	0
'2008100202'	0	0	0,23	0	0	0	0	0	0,2	0	0	0
'2008100203'	0	0	0	0	0	0	0	0	0	0	0	0
'2008100204'	0,21	0	0	0	0	0	0	0	0	0	0	0
'2008100205'	0,32	0	0	0	0	0	0	0	0	0	0	0
'2008100206'	0,11	0	0	0	0	0	0	0	0	0	0	0
'2008100207'	0	0,02	0	0	0	0	0	0	0	0	0	0
'2008100208'	0	0,15	0	0	0	0	0	0	0	0	0	0
'2008100209'	0	0,04	0	0	0	0	0	0	0	0	0	0
'2008100210'	0	0	0	0,8	0	0	0	0	0	0	0	0
'2008100211'	0	0	0	1,52	0	0,48	0	0	0	0	0	0
'2008100212'	0	0	0	1,2	0	0,63	0	0	0	0	0	0
'2008100213'	0	0	0	0,64	0	0,27	0	0	0	0	0	0
'2008100214'	0	0	0	0,17	0	0,12	0	0	0	0	0	0
'2008100215'	0	0	0	0	0	0	0	0	0	0	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
'2010123118'	0	0	0	0	0	0	0	0	0	0	0	0
'2010123119'	0	0	0	0	0	0	0	0	0	0	0	0
'2010123120'	0	0	0	0	0	0	0	0	0	0	0	0
'2010123121'	0	0	0	0	0	0	0	0	0	0	0	0
'2010123122'	0	0	0	0	0	0	0	0	0	0	0	0
'2010123123'	0	0	0	0	0	0	0	0	0	0	0	0

Explanation to table: The table shows rainfall data extracted from 3B41RT data files for all 12 different TMPA grid box positions where triggered landslide(s) are located. Grid coordinates are listed in the top most lines ('TMPA_lat = latitude, 'TMPA_long' = longitude). The left column list the time range related to each rainfall measurement at the following format; yyyymmddHH, where y = year, m = month, d = day and H = hour.

Table A 5. Small selection of data from the 3B42RT rainfall inventory prepared for threshold analysis

Date	Predictor 1	Predictor 2	Predictor 3	Predictor 4	Predictor 5	…	Predictor 11	Predictor 12	Predictor 13	Outcome
01.01.2007	0,000	NaN	NaN	NaN	NaN	…	NaN	NaN	NaN	n
02.01.2007	0,000	0,000	NaN	NaN	NaN	…	NaN	NaN	NaN	n
03.01.2007	0,000	0,000	0,000	NaN	NaN	…	NaN	NaN	NaN	n
04.01.2007	0,000	0,000	0,000	0,000	NaN	…	NaN	NaN	NaN	n
05.01.2007	0,000	0,000	0,000	0,000	0,000	…	NaN	NaN	NaN	n
06.01.2007	1,300	0,650	0,433	0,325	0,260	…	NaN	NaN	NaN	n
07.01.2007	0,000	0,650	0,433	0,325	0,260	…	NaN	NaN	NaN	n
08.01.2007	0,000	0,000	0,433	0,325	0,260	…	NaN	NaN	NaN	n
09.01.2007	0,000	0,000	0,000	0,325	0,260	…	NaN	NaN	NaN	n
10.01.2007	0,000	0,000	0,000	0,000	0,260	…	NaN	NaN	NaN	n
11.01.2007	0,000	0,000	0,000	0,000	0,000	…	NaN	NaN	NaN	n
12.01.2007	0,000	0,000	0,000	0,000	0,000	…	NaN	NaN	NaN	n
13.01.2007	0,580	0,290	0,193	0,145	0,116	…	NaN	NaN	NaN	n
14.01.2007	0,160	0,370	0,247	0,185	0,148	…	NaN	NaN	NaN	n
15.01.2007	0,000	0,080	0,247	0,185	0,148	…	0,136	NaN	NaN	n
16.01.2007	3,430	1,715	1,197	1,043	0,834	…	0,365	NaN	NaN	n
17.01.2007	0,140	1,785	1,190	0,933	0,862	…	0,374	NaN	NaN	n
18.01.2007	0,000	0,070	1,190	0,893	0,746	…	0,374	NaN	NaN	n
19.01.2007	0,000	0,000	0,047	0,893	0,714	…	0,374	NaN	NaN	n
20.01.2007	1,080	0,540	0,360	0,305	0,930	…	0,446	NaN	NaN	n
21.01.2007	0,000	0,540	0,360	0,270	0,244	…	0,359	NaN	NaN	n
22.01.2007	0,000	0,000	0,360	0,270	0,216	…	0,359	NaN	NaN	n
23.01.2007	0,000	0,000	0,000	0,270	0,216	…	0,359	NaN	NaN	n
24.01.2007	0,000	0,000	0,000	0,000	0,216	…	0,359	NaN	NaN	n
25.01.2007	0,000	0,000	0,000	0,000	0,000	…	0,359	NaN	NaN	n
26.01.2007	0,000	0,000	0,000	0,000	0,000	…	0,359	NaN	NaN	n
27.01.2007	0,000	0,000	0,000	0,000	0,000	…	0,359	NaN	NaN	n
28.01.2007	0,000	0,000	0,000	0,000	0,000	…	0,321	NaN	NaN	n
29.01.2007	6,115	3,058	2,038	1,529	1,223	…	0,718	NaN	NaN	n
30.01.2007	3,045	4,580	3,053	2,290	1,832	…	0,921	0,528	NaN	n
31.01.2007	1,230	2,138	3,463	2,598	2,078	…	0,774	0,569	NaN	n
01.02.2007	0,000	0,615	1,425	2,598	2,078	…	0,765	0,569	NaN	n
02.02.2007	0,000	0,000	0,410	1,069	2,078	…	0,765	0,569	NaN	n
03.02.2007	1,000	0,500	0,333	0,558	1,055	…	0,831	0,603	NaN	n
04.02.2007	0,310	0,655	0,437	0,328	0,508	…	0,780	0,613	NaN	n
05.02.2007	0,000	0,155	0,437	0,328	0,262	…	0,780	0,570	NaN	n
06.02.2007	0,000	0,000	0,103	0,328	0,262	…	0,780	0,570	NaN	n
07.02.2007	28,835	14,418	9,612	7,286	6,029	…	2,702	1,531	NaN	n
08.02.2007	4,275	16,555	11,037	8,278	6,684	…	2,987	1,673	NaN	n
09.02.2007	0,000	2,138	11,037	8,278	6,622	…	2,987	1,673	NaN	n
10.02.2007	0,000	0,000	1,425	8,278	6,622	…	2,987	1,673	NaN	n
11.02.2007	0,000	0,000	0,000	1,069	6,622	…	2,987	1,673	NaN	n
12.02.2007	0,265	0,133	0,088	0,066	0,908	…	3,005	1,663	NaN	n
13.02.2007	31,685	15,975	10,650	7,988	6,390	…	4,710	2,714	NaN	n
14.02.2007	0,000	15,843	10,650	7,988	6,390	…	4,507	2,714	NaN	n
15.02.2007	0,000	0,000	10,562	7,988	6,390	…	4,425	2,599	NaN	n
16.02.2007	0,000	0,000	0,000	7,921	6,390	…	4,425	2,595	NaN	n
17.02.2007	0,000	0,000	0,000	0,000	6,337	…	4,425	2,595	NaN	n
18.02.2007	0,000	0,000	0,000	0,000	0,000	…	4,358	2,595	NaN	n
19.02.2007	0,000	0,000	0,000	0,000	0,000	…	4,337	2,559	NaN	n
20.02.2007	0,000	0,000	0,000	0,000	0,000	…	4,337	2,559	NaN	n
21.02.2007	0,000	0,000	0,000	0,000	0,000	…	4,337	2,559	NaN	n
22.02.2007	0,000	0,000	0,000	0,000	0,000	…	2,415	2,559	NaN	n
23.02.2007	0,000	0,000	0,000	0,000	0,000	…	2,130	2,559	NaN	n
24.02.2007	0,000	0,000	0,000	0,000	0,000	…	2,130	2,559	NaN	n
25.02.2007	0,000	0,000	0,000	0,000	0,000	…	2,130	2,559	NaN	n
26.02.2007	0,000	0,000	0,000	0,000	0,000	…	2,130	2,559	NaN	n
27.02.2007	0,000	0,000	0,000	0,000	0,000	…	2,112	2,559	NaN	n
28.02.2007	0,000	0,000	0,000	0,000	0,000	…	0,000	2,355	NaN	n

Explanation to table: Each predictor variable represents a rainfall series of antecedent rainfall, where the duration of rainfall accumulation is different for all predictors (see Table 5-4). Outcome variable: 'y' if landslide(s) occurred.

Table A 6. Rainfall series of landslide triggering data extracted from TMPA-RT rainfall product 3B42RT. Note the most intense landslide triggering rainfalls in bold.

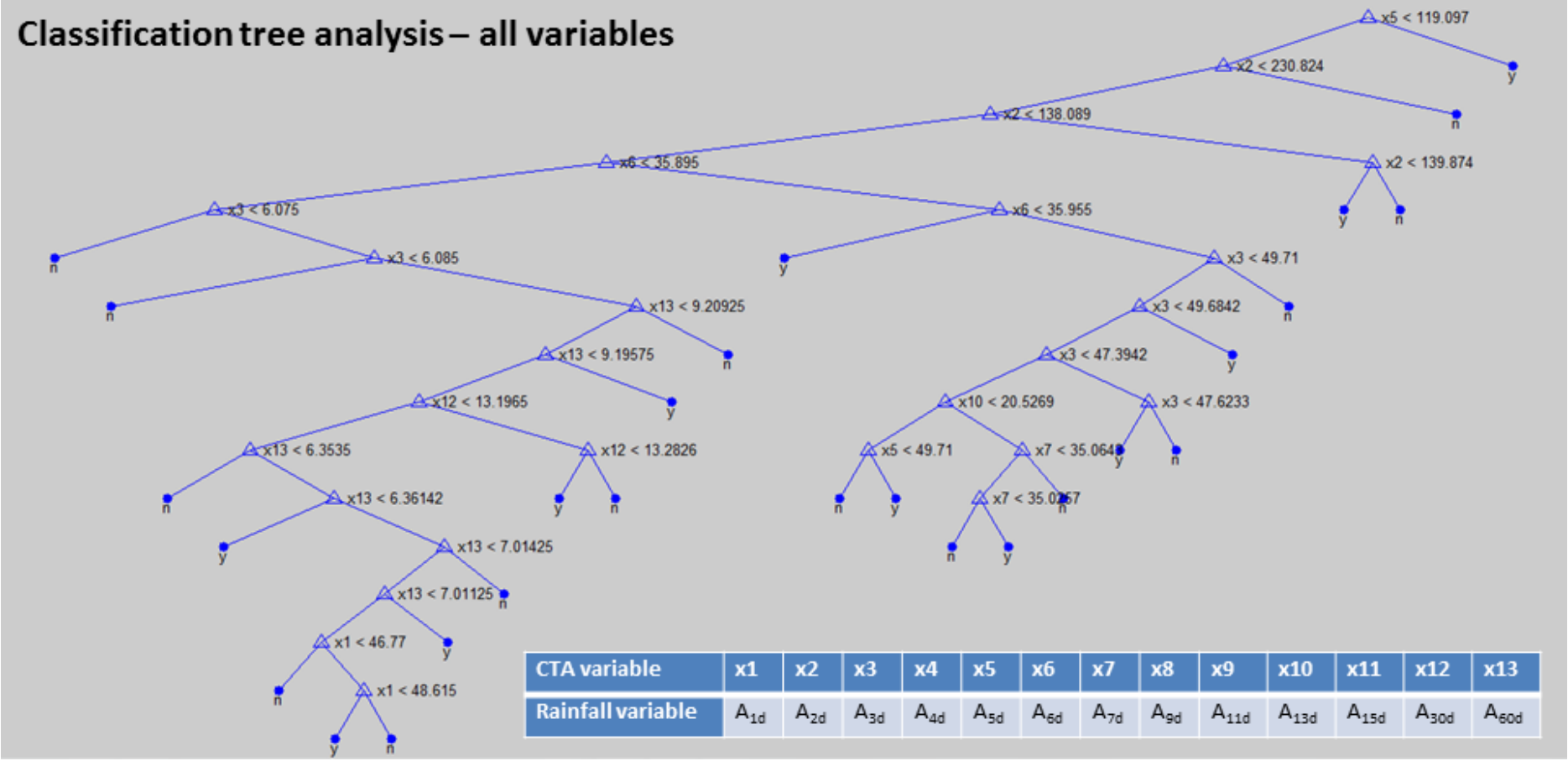
	Latitude	Longitude	Date	1	2	3	4	5	6	7	9	11	13	15	30	60
Chittagong	**22,375**	**91,875**	**11.06.2007**	**323,5**	**241,6**	**163,6**	**123,5**	**98,8**	**87,0**	**74,8**	**61,2**	**50,1**	**42,4**	**36,7**	**23,3**	**14,2**
Chittagong	**22,625**	**92,125**	**11.06.2007**	**196,8**	**138,8**	**94,6**	**71,2**	**57,0**	**47,5**	**41,1**	**33,5**	**27,4**	**23,2**	**20,1**	**14,6**	**11,1**
Plains	23,875	90,875	**10.09.2007**	0,7	0,4	6,1	15,7	14,4	12,9	11,4	9,1	7,6	6,6	5,8	5,3	5,7
Plains	22,625	89,375	**15.10.2007**	47,3	27,8	18,5	13,9	11,1	9,3	7,9	18,9	16,3	13,8	12,0	8,1	6,6
Cox' Bazar	20,875	91,375	**03.07.2008**	5,0	44,6	47,5	53,1	62,2	60,0	53,5	45,0	36,9	31,6	32,2	23,6	14,2
Cox' Bazar	21,375	92,125	**03.07.2008**	22,9	21,9	29,5	30,0	35,5	35,9	30,8	25,4	20,8	17,6	15,6	10,7	11,3
Cox' Bazar	20,875	92,375	**03.07.2008**	52,7	61,7	49,7	41,5	48,6	58,1	50,4	41,2	33,7	29,0	26,6	20,6	15,6
Cox' Bazar	20,875	92,375	**06.07.2008**	27,9	36,1	26,4	33,0	40,5	38,1	35,0	47,5	40,9	34,6	30,3	22,7	16,9
Sylhet	24,875	91,875	**01.08.2008**	3,5	16,3	10,8	8,2	6,8	8,2	7,4	8,8	13,5	13,9	12,5	11,0	9,2
Chittagong	22,375	91,875	**18.08.2008**	16,1	8,4	8,1	6,1	4,8	4,4	3,9	5,1	8,3	7,2	7,1	5,5	6,4
Chittagong	22,375	91,875	**23.08.2008**	3,0	1,8	16,6	12,5	10,2	11,2	9,7	8,4	7,0	6,7	8,3	6,9	7,0
Sylhet	24,375	91,625	**18.05.2009**	24,3	34,7	34,6	25,9	20,8	24,1	20,7	21,2	17,5	15,5	16,4	13,2	8,9
Chittagong	22,125	92,125	**31.07.2009**	0,0	1,7	25,9	56,8	51,7	44,4	38,1	29,6	24,2	20,5	21,9	12,6	19,6
Cox' Bazar	**21,375**	**92,125**	**15.06.2010**	**152,6**	**139,7**	**117,6**	**95,0**	**78,6**	**66,2**	**56,8**	**44,2**	**38,0**	**34,8**	**30,1**	**36,0**	**19,2**
Cox' Bazar	**21,125**	**92,125**	**15.06.2010**	**173,0**	**198,6**	**179,5**	**143,0**	**122,7**	**102,6**	**88,5**	**68,9**	**57,9**	**51,2**	**44,4**	**46,5**	**24,3**
Cox' Bazar	**20,875**	**92,375**	**15.06.2010**	**186,3**	**203,6**	**170,2**	**135,8**	**125,1**	**104,5**	**91,5**	**71,6**	**58,8**	**57,1**	**49,5**	**50,5**	**26,1**

Explanation to table: The table show accumulated rainfall intensities for all TMPA-RT locations (each representing a 0.25° x 0.25° grid box) where landslides have been registered. Note that the highest rainfall intensities are in bold letters and that all rainfall events register in the ADPC data in Table A 1 are shaded light grey.

Table A 7. Comparison of TMPA-RT product 3B42RT and rain gauges for main fatal storm events

Located area	Position			Rainfall intensity (mm/d), time series of different duration							
	Latitude	**Longitude**	**Date**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Source**
Chittagong	**22,375**	**91,875**	**11.06.2007**	**323,5**	**241,6**	**163,6**	**123,5**	**98,8**	**87,0**	**74,8**	TMPA-RT
Chittagong	**22,625**	**92,125**	**11.06.2007**	**196,8**	**138,8**	**94,6**	**71,2**	**57,0**	**47,5**	**41,1**	TMPA-RT
Chittagong	city + surroundings			425,0	256,5	172,0	139,5	112,4	97,3	86,7	ADPC
Cox' Bazar	**21,375**	**92,125**	**15.06.2010**	**152,6**	**139,7**	**117,6**	**95,0**	**78,6**	**66,2**	**56,8**	TMPA-RT
Cox' Bazar	**21,125**	**92,125**	**15.06.2010**	**173,0**	**198,6**	**179,5**	**143,0**	**122,7**	**102,6**	**88,5**	TMPA-RT
Cox' Bazar	**20,875**	**92,375**	**15.06.2010**	**186,3**	**203,6**	**170,2**	**135,8**	**125,1**	**104,5**	**91,5**	TMPA-RT
Cox' Bazar/Teknaf	+ surroundings			132,0	105,0	95,0	75,0	60,0	56,8	48,7	ADPC

Appendix B. Classification tree analysis



CTA variable	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10	x11	x12	x13
Rainfall variable	A_{1d}	A_{2d}	A_{3d}	A_{4d}	A_{5d}	A_{6d}	A_{7d}	A_{9d}	A_{11d}	A_{13d}	A_{15d}	A_{30d}	A_{60d}

Figure B 1. Resulting classification tree form CTA without excluding predictors due to high correlation. Rainfall variable corresponding to the predictors included are listed in figure table.

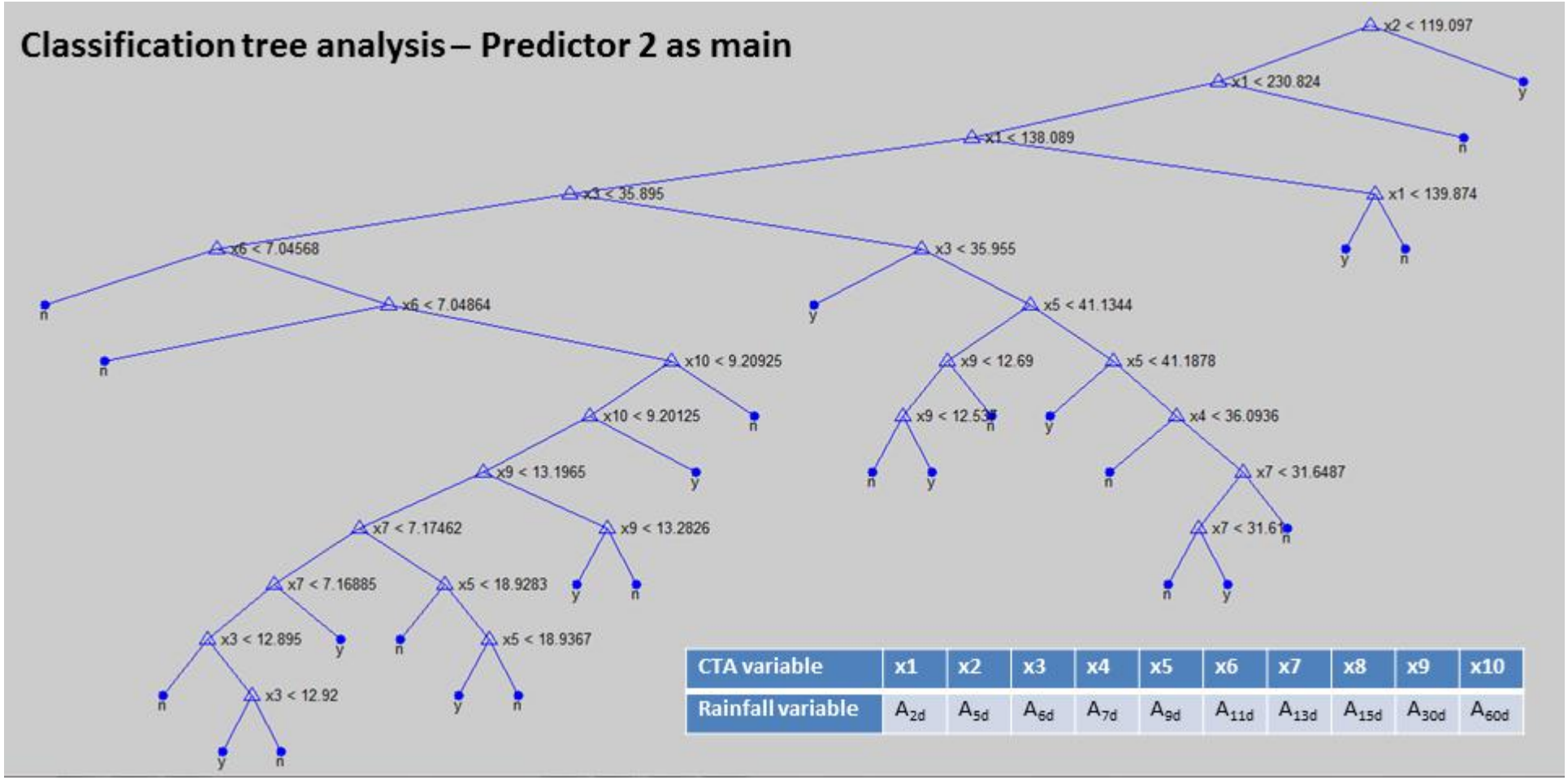


Figure B 2. Result of CTA analysis not including predictors of high correlation (ρ>0.8) to predictor 2; rainfall variable corresponding to the predictors included are listed in figure table.

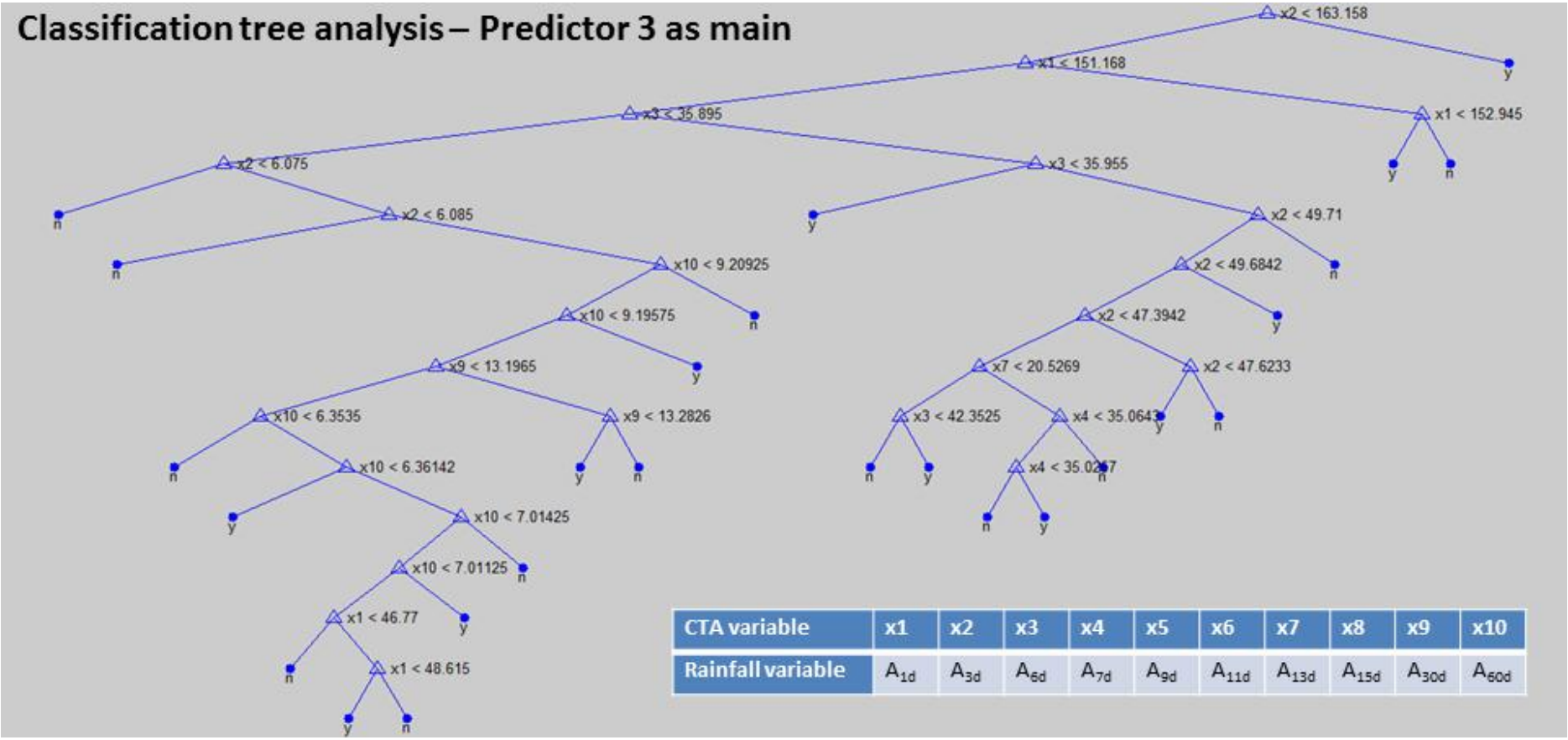


Figure B 3. Result of CTA analysis not including predictors of high correlation (ρ>0.8) to predictor 3; rainfall variable corresponding to the predictors included are listed in figure table.

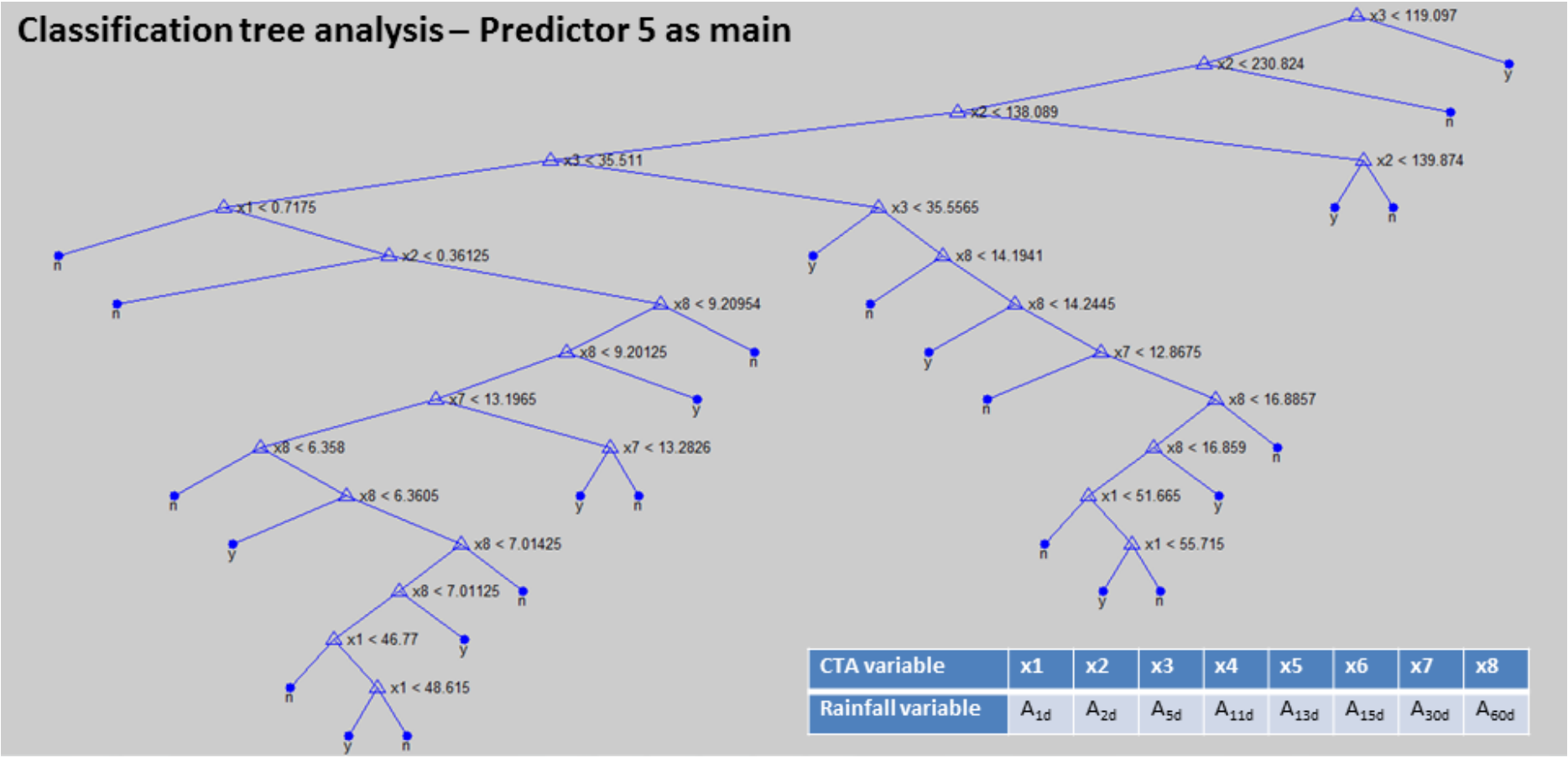


Figure B 4. Result of CTA analysis not including predictors of high correlation (ρ>0.8) to predictor 5; rainfall variable corresponding to the predictors included are listed in figure table.

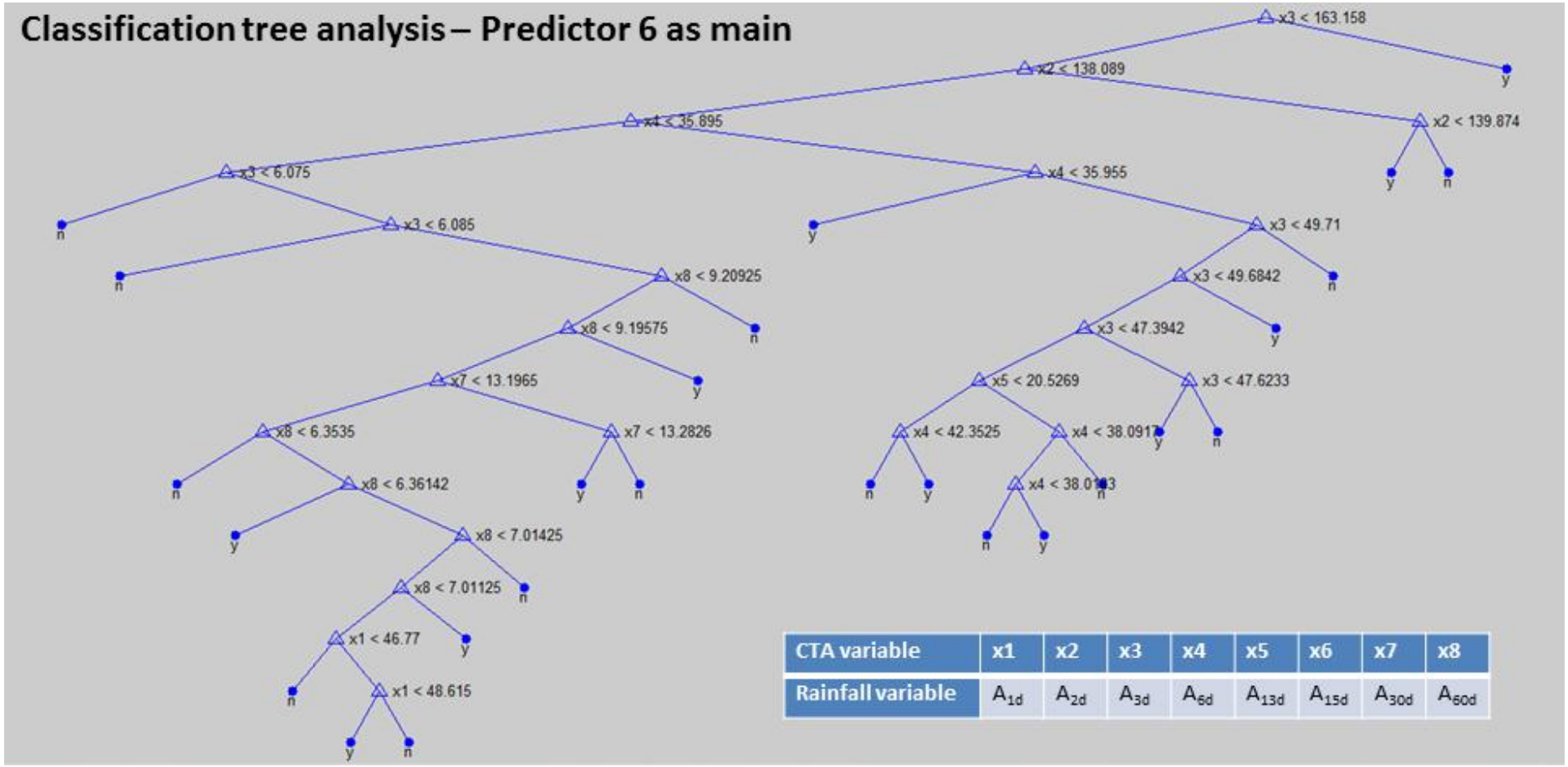


Figure B 5. Result of CTA analysis not including predictors of high correlation (ρ>0.8) to predictor 6; rainfall variable corresponding to the predictors included are listed in figure table.

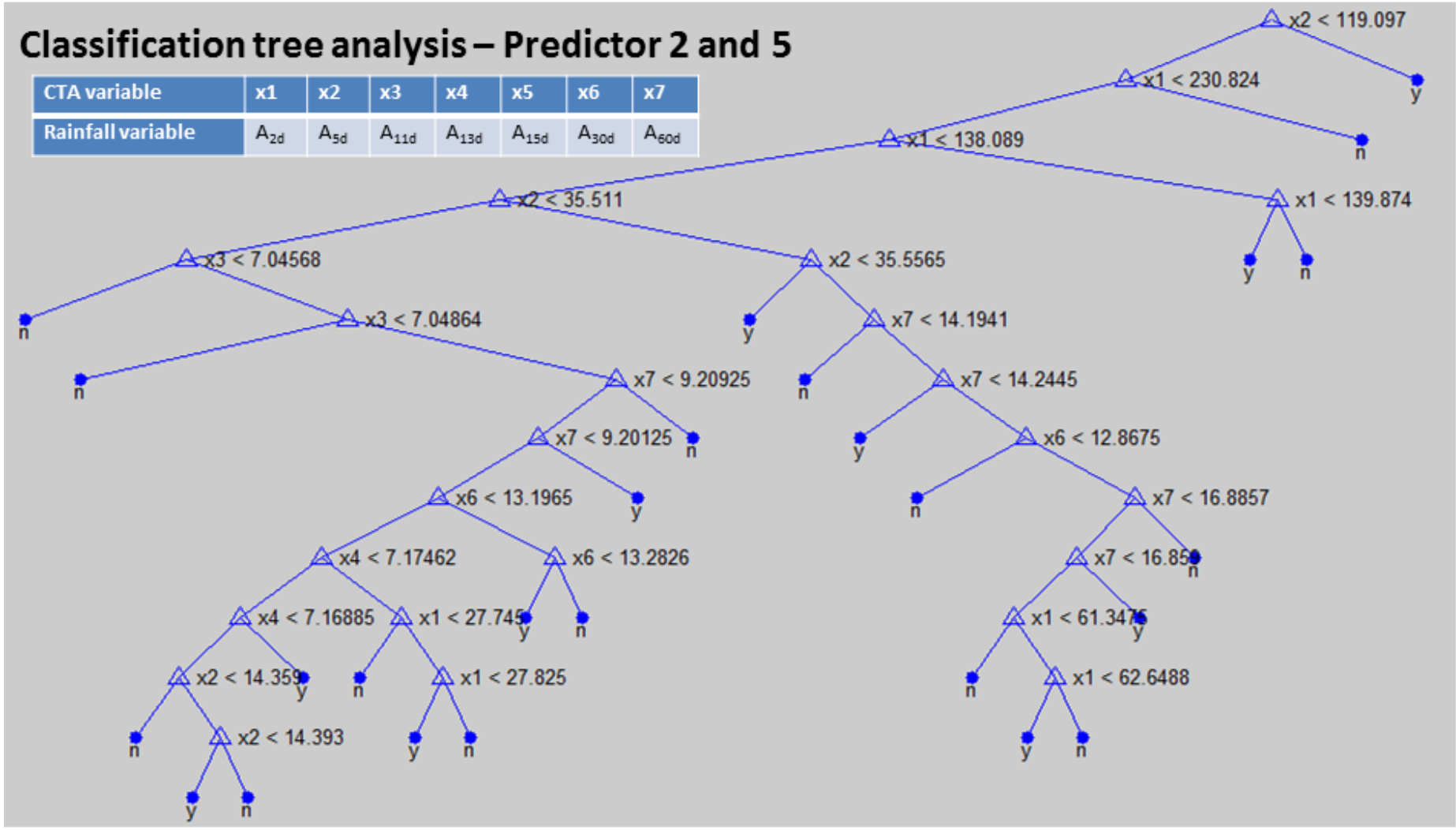


Figure B 6. Result of CTA analysis not including predictors of high correlation (ρ>0.8) to predictor 2 and 5; rainfall variable corresponding to the predictors included are listed in figure table.

Appendix C. Matlab-scripts

Matlab was the main tool used for this study. The program, including its Statistical Toolbox, was use both for acquiring data, data handling and processing, statistical analyses and visualisation of plots and tables. As a result, several Matlab-scripts have been developed for different tasks. All scripts representing key elements of the methodology, including data acquisition, analyses and results are presented in this appendix.

The files are tried sorted in 3 section representing different parts of the thesis work;

1. Data acquisition: Acquiring TMPA-RT files, extracting rainfall data and preparing inventories.
2. Validation part: Analysis and comparison of rain gauge and TMPA-RT products.
3. Threshold part: Preparing inventory of predictor and executing the multivariate analyses.

In all parts most scripts are linked such that the execution of one requires the execution of another first. As result of this, the scripts are listed in ascending order according to the processes and required order of execution to be able to run al scripts.

It should be noted that the structure of TMPA-RT files and TRMM FTP-server has been changes since this study was conducted.

Table C 1. Overview of Matlab scripts used for data acquisition part.

Script nr.	Script name	Script purpose	Starting page
1	GetAllFiles.m	Downloads files from ftp-server and create file archive on local hard disk drive.	96
2	TRMMscript_Several.m	Handles several scripts (nr. 2-9) used to extract rainfall data from archived files of both TRMM-based products.	98
3	InsertData.m	Insert filename and other parameters needed to extract rainfall data.	98
4	Load_LSinventory.m	Extract data from landslide inventory.	99
5	CheckTime.m	Check if time data from landslide inventory is okay.	100
6	FindFilenamesFullSet.m	Find name of all files needed to extract rainfall data corresponding to data input and landslide inventory.	101
7	FindTMPApositions.m	Find the position in the global rainfall grid corresponding to the landslide positions in landslide inventory.	102
8	FindOneTMPAposition.m	As script 7 for only one landslide position input.	103
9	ExtractRainfallData.m	Extract rainfall data for all landslide inventory positions for all time-series needed for threshold analysis.	104
10	SaveData.m	Saves all position and rainfall data in *.mat-files.	105
11	GetDailyRainfallNew.m	Calculates the daily rainfall from the original 1hr and 3hr rainfall estimates of the TRMM-based rainfall products (result of script nr. 2, saved by script nr. 9).	107

Table C 2. Overview of Matlab scripts used for validation part.

Script nr.	Script name	Script purpose	Line position
12	ValidationStatistics.m	Handle several scripts (nr. 12– 20) with purpose to validate TMPA data.	108
13	LoadData.m	Load data needed for TMPA validation	109
14	PrepareSeries.m	Connect TMPA and rain gauge series of same date	109
15	VisInspection.m	Plot full rainfall series of daily data	110
16	LandslideData.m	Prepare landslide data for use in validation	110
17	DistributionCurve.m	Prepare data for creation of distribution curve	111
18	MonthlyRainfall.m	Prepare data for creation of monthly data used for visual inspection	112
19	CorrelationPlots	Prepare and plot correlation graphs	114
20	CheckNaN.m	Find and analyse missing rainfall data (NaNs)	115

Table C 3. Overview of Matlab scripts used for threshold part.

Script nr.	Script name	Script purpose	Line position
21	ThresholdAnalysis.m	Handle the complete TMPA threshold analysis part and all related scripts (22-25).	116
22	PrepareRFdata.m	Change rainfall series to meet criteria for analyses	117
23	ExtractYdays.m	Extract all rainfall data related to landslide triggering events	118
24	ClassTreeAnalysis.m	Perform all stages of CTA	118
25	DiscriminantAnalysis.m	Perform all stages of discriminant analysis	119
26	DA_subplots.m	Run multiple DAs in a matrix and find best prior	121
27	DA_final.m	Run final DA with best prior and output figure	123

1. GetAllFiles.m

```
%%%%%%%%%% Download all NASA TMPA-RT files %%%%%%%%%%

%This script downloads files form the ftp-server and store them on the
local
%hard drive in a file archive.

%The script is used for download of both TRMM rainfall products, but ftp-
%folder structure in script must be changed between products, as it is
%slightly different between the two.


%% Calibrated IR
RainfallProduct = '3B41RT';     %'3B41RT' or '3B42RT'
Archive = ['C:\Users\Fagerlandet 2.0\Skole\_File Archive\' ...
    RainfallProduct '\'];
cd(FileArchive)
ftp-productname = 'calibratedIR';   %name of product folder in ftp-dir
    %'calibratedIR' or 'mergedIRmicro'
fileduration = 1;   %1 hr for 3B41RT and 3 hrs for 3B42RT

TRMM = ftp('trmmopen.gsfc.nasa.gov');

for n=2005  %2005 starts 1 march
    ftpFolder = ['/pub/merged/' ftp-productname '/V5/' num2str(n) '/'];
    cd(TRMM, ftpFolder);
    starttime = [num2str(n) '030100'];
    endtime = [num2str(n) '123123'];
    startnum = datenum(starttime, 'yyyymmddHH');
    endnum = datenum(endtime, 'yyyymmddHH');
    timediff = endnum - startnum;
    filecount = timediff * 24/fileduration +1;  %total nr of files to dwnld
    for i=1:filecount
        filename = [RainfallProduct '.' ...
            datestr(startnum+(i-1)/24, 'yyyymmddHH') '.bin.gz'];
        ArchiveFolder = [FileArchive 'V5\' num2str(n) '\' filename(12:13)];
        if isdir(ArchiveFolder) == 0;   %create folder if non-existent
            mkdir(ArchiveFolder)
        end
        cd(ArchiveFolder)   %select folder for file download
        try
            mget(TRMM, filename);   %download file
        catch
            ErrorMsg = ['File ' filename ' was not found on server.'];
            disp(ErrorMsg)
        end
    end
end

for n=2006:2008
    ftpFolder = ['/pub/merged/' ftp-productname '/V5/' num2str(n) '/'];
    cd(TRMM, ftpFolder);
    starttime = [num2str(n) '010100'];
    endtime = [num2str(n) '123123'];
    startnum = datenum(starttime, 'yyyymmddHH');
    endnum = datenum(endtime, 'yyyymmddHH');
    timediff = endnum - startnum;
    filecount = timediff * 24/fileduration +1;
    for i=1:filecount
```

```
        filename = [RainfallProduct '.' ...
            datestr(startnum+(i-1)/24, 'yyyymmddHH') '.bin.gz'];
        ArchiveFolder = [FileArchive 'V5\' num2str(n) '\' filename(12:13)];
        if isdir(ArchiveFolder) == 0;   %create folder if non-existent
            mkdir(ArchiveFolder)
        end
        cd(ArchiveFolder)   %select folder for file download
        try
            mget(TRMM, filename);   %download file
        catch
            ErrorMsg = ['File ' filename ' was not found on server.'];
            disp(ErrorMsg)
        end
    end
end

for n=2008
    ftpFolder = ['/pub/merged/' ftp-productname '/' num2str(n) '/'];
    cd(TRMM, ftpFolder);
    starttime = [num2str(n) '011000'];
    endtime = [num2str(n) '123123'];
    startnum = datenum(starttime, 'yyyymmddHH');
    endnum = datenum(endtime, 'yyyymmddHH');
    timediff = endnum - startnum;
    filecount = timediff * 24/fileduration +1;    %one file each 3 hours
    for i=1:filecount
        filename = [RainfallProduct '.' ...
            datestr(startnum+(i-1)/24, 'yyyymmddHH') ...
            '.6.bin.gz'];
        ArchiveFolder = [FileArchive num2str(n) '\' filename(12:13)];
        if isdir(ArchiveFolder) == 0;   %create folder if non-existent
            mkdir(ArchiveFolder)
        end
        cd(ArchiveFolder)   %select folder for file download
        try
            mget(TRMM, filename);
        catch
            ErrorMsg = ['File ' filename ' was not found on server.'];
            disp(ErrorMsg)
        end
    end
end

for n=2009:2010
    ftpFolder = ['/pub/merged/' ftp-productname '/' num2str(n) '/'];
    cd(TRMM, ftpFolder);
    starttime = [num2str(n) '102000'];
    endtime = [num2str(n) '123123'];
    startnum = datenum(starttime, 'yyyymmddHH');
    endnum = datenum(endtime, 'yyyymmddHH');
    timediff = endnum - startnum;
    filecount = timediff * 24/fileduration +1;    %one file each 3 hours
    for i=1:filecount
        filename = [RainfallProduct '.' ...
            datestr(startnum+(i-1)/24, 'yyyymmddHH') ...
            '.6.bin.gz'];
        ArchiveFolder = [FileArchive num2str(n) '\' filename(12:13)];
        if isdir(ArchiveFolder) == 0;   %create folder if non-existent
            mkdir(ArchiveFolder)
        end
        cd(ArchiveFolder)   %select folder for file download
```

```
        try
            mget(TRMM, filename);
        catch
            ErrorMsg = ['File ' filename ' was not found on server.'];
            disp(ErrorMsg)
        end
    end
end

cd(TRMM, '/pub/merged/' ftp-productname '/');
starttime = '2010010100';
endtime = '2011040118';
startnum = datenum(starttime, 'yyyymmddHH');
endnum = datenum(endtime, 'yyyymmddHH');
timediff = endnum - startnum;
filecount = timediff * 24/fileduration +1;
    for i=1:filecount
        filename = [RainfallProduct '.' ...
            datestr(startnum+(i-1)/24, 'yyyymmddHH') ...
            '.6.bin.gz'];
        ArchiveFolder = [FileArchive filename(8:11) '\' filename(12:13)];
        if isdir(ArchiveFolder) == 0;   %create folder if non-existent
            mkdir(ArchiveFolder)
        end
        cd(ArchiveFolder)   %select folder for file download
        try
            mget(TRMM, filename);
        catch
            ErrorMsg = ['File ' filename ' was not found on server.'];
            disp(ErrorMsg)
        end
    end

close(TRMM)
```

2. TRMMscript_Several.m

```
%%%%%%%%%  RUN ALL TRMM SCRIPTS FOR EXTRACTION OF RAINFALL DATA %%%%%%%%%%

%Script is used for both TRMM-based rainfall products

InsertData;             %Input of variables needed to run script-series
Load_LSinventory;       %Load data from landslide inventory (*.xls(x)-file)
CheckTime;              %Checks for incorrect time input
FindFilenamesFullSet;          %Find TMPA filenames needed
FindTMPApositions;      %Finds TMPA grid positions needed
ExtractRainfallData;    %Extracts rainfall data from TMPA_files
SaveData;               %Saves all position- and rainfall-data in result
                            %files.
```

3. InsertData.m

```
%%%%%%%%%%% INSERT VARIABLES AND DATA FOR ANALYSIS %%%%%%%%%%

%This script run as a part of the script TRMMscript_Several.
RainfallProduct = '3B41RT';     %select type of rainfall product
                                %(3B41RT or 3B42RT).

UTCzone = 6;    %Insert UTC zone of the area (e.g. El Salvadore is UTC -6,
```

```
                %Bangladesh is UTC +6). This to use local time.
RGstart = 7;    %Insert the time of day when rain gauge data are collected
                %in the area. The number is equivalent to the time of day
                %in hours (0-23, equals 00:00 - 23:00). If time is not
                %known, it is recommended to use 7 (7 AM / 07:00).
PreceedingDays = 60;    %Insert number of rainfall day to analyse preceding
                        %the day of landslide triggering.
FollowingDays  = 1; %Insert number of following days to include in analysis
            %The day of the landslide occurrence is automatically included.

xlsName = 'Full inventory_Bangladesh.xlsx'; %Insert name of landslide
                                            %inventory file.
sheetName = 'Landslides';   %Insert sheet name.
Range = 'A1:U52';   %Insert range of landslide data in sheet.

startline = 2;      %Select range for input in analysis (skip header line).
endline   = 42;

%Insert column position in spread sheet of the different data required:
colNr_LSnr = 1;     %Give each line of registered data a number, to
                %discriminate between landslides with same name/area.
colNr_LSname = 3;   %Name of landslides (e.g. often nearest town/village)
colNr_LSlat = 7;    %Latitude position of landslide
colNr_LSlong = 8;   %Longitude position of landslide
colNr_Year = 14;    %Date of landslide in;  year
colNr_Month = 15;   %                       month
colNr_Day = 16;     %                       day
                    %Hours are left out as only date of triggering are
                    %known for most of the landslides in the inventory.
```

4. Load_LSinventory.m

```
%%%%%%%%% LOAD DATA FROM LANDSLIDE INVENTORY %%%%%%%%%%

%This script run as a part of the script TRMMscript_Several.

%Extracts data from landslide inventory. Data- and parameter-input from
%the script InsertData.m are used here.

[X, Y, Z] = xlsread(xlsName, sheetName, Range);
    %X = numerical data, Y = string data as cell, Z = all data as cell
    disp('Xls-file read.');
    clear('X', 'Y');

for i = startline:endline;      %Get all landslide/rain gauge coordinates
    temp=[num2str(cell2mat(Z(i,colNr_LSnr))) '_' char(Z(i,colNr_LSname))];
    PosName(i-1,1:9) = temp(1:9);       %Enters name of surface position
            %(usually a landslide or rain gauge)
    Latitude(i-1,1)  = cell2mat(Z(i,colNr_LSlat));  %Position on earth
surface
    Longitude(i-1,1) = cell2mat(Z(i,colNr_LSlong)); %latitude and longitude
end
    disp('All lat/lon positions stored.');

for i = startline:endline;      %Load dates into vectors
    Date(i-1,3) = cell2mat(Z(i,colNr_Day));
    Date(i-1,2) = cell2mat(Z(i,colNr_Month));
    Date(i-1,1) = cell2mat(Z(i,colNr_Year));
end
```

```
236
237 for i = 1:length(Date(:,1));      %Create dates for filenames
238     if Date(i,3) <10
239         dd(i,1:2) = ['0' num2str(Date(i,3))];
240     else dd(i,1:2) = num2str(Date(i,3));
241     end
242     if Date(i,2) < 10
243         mm(i,1:2) = ['0' num2str(Date(i,2))];
244     else mm(i,1:2) = num2str(Date(i,2));
245     end
246     yyyy(i,1:4) = num2str(Date(i,1));
247 end
248
249 hour_num = 1/24;    %numerical date number for one hour
250 for i = 1:length(Date(:,1));
251     initialTime(i,:) = [dd(i,:) '.' mm(i,:) '.' yyyy(i,:)]; %LS-day
252     endnum(i,:) = datenum(initialTime(i,:), 'dd.mm.yyyy') + ...
253         FollowingDays + (24 * hour_num);
254             %end of day at 24:00 = 00:00 next day => +24 hours
255     endingtime(i,:) = datestr(endnum(i,:), 'dd.mm.yyyy HH:MM');
256     startnum(i,:) = datenum(initialTime(i,:), 'dd.mm.yyyy') - ...
257         (PreceedingDays) + 1*hour_num; %first measurement at 01:00
258     startingtime(i,:) = datestr(startnum(i,:), 'dd.mm.yyyy HH:MM');
259 end
260
```

5. CheckTime.m

```
261 %%%%%%% CHECK TIME INPUT FROM LANDSLIDE INVENTORY %%%%%%
262
263 %This script run as a part of the script TRMMscript_Several.
264 %Checks if entered time data is ok
265
266 for n=1:length(startingtime(:,1))
267     if str2num(startingtime(n,4:5)) > 12
268         error('The starting month(mm) is incorrect(>12). Please correct
269 this.')
270     end
271     if str2num(endingtime(n,4:5)) > 12
272         error('The ending month(mm) is incorrect(>12). Please correct
273 this.')
274     end
275 end
276
277 startnum = datenum(startingtime, 'dd.mm.yyyy HH:MM');
278 endnum = datenum(endingtime, 'dd.mm.yyyy HH:MM');
279 hour_num = 1/24;
280 if startnum > endnum
281     tempText = ['Your end time is less than your start time. ' ...
282         'Please change time input and try again.'];
283     error(tempText)
284 end
285
286 disp('Time input seems correct.')
287
288
289
```

6. FindFilenamesFullSet.m

```
%%%%%%%%% FIND FILENAMES %%%%%%%%%%%%

%This script run as a part of the script TRMMscript_Several.

%Find filename of all files needed for data extraction. All filenames are
%saved temporally in both zipped und un-zipped format in the file
%filenames.mat.

if RainfallProduct == '3B42RT'
    filefrequency = 3;  %3 hrs rainfall estimates for 3B42RT
end
if RainfallProduct == '3B41RT'
    filefrequency = 1;  %1 hr rainfall estimates for 3B41RT
end
hour_num = 1/24;        %find the numerical number for 1 hour, 1 day = 1

startnum = datenum('01.01.2007', 'dd.mm.yyyy');%hours removed to download..
endnum = datenum('31.12.2010', 'dd.mm.yyyy'); %..full days (at 3 hr
interval)

NrOfDays = endnum - startnum +1; %days to download per landslide
NrOfDownloads = NrOfDays*24/filefrequency;

for n=1:length(startnum)
    for i=1:NrOfDownloads
        fileTime_num(((n-1)*NrOfDownloads+i),:) = ...
            startnum(n,:) + (i-1)*(filefrequency/24);
    end         %creates numerical time for each file to download
end
%sort filetimes (all files) and remove duplicate file time-inputs
fileTime_sorted = sort(fileTime_num);
for i=1:length(fileTime_sorted)-1
    fileTime_diff(i,:) = fileTime_sorted(i+1) - fileTime_sorted(i);
end
find_diff = find(fileTime_diff == 0);
fileTime_new = fileTime_sorted;
fileTime_new(find_diff+1) = [];
fileTimestr = datestr(fileTime_new, 'yyyymmddHH');

%% Store filenames needed for rainfall data processing
for n=1:length(fileTimestr)
    if fileTime_new(n,:) < datenum('01.01.2009', 'dd.mm.yyyy')
         filename2(n,:) = [RainfallProduct '.' fileTimestr(n,:) '.bin  '];
    else filename2(n,:) = [RainfallProduct '.' fileTimestr(n,:) '.6.bin'];
    end
end
filename2_cell = cellstr(filename2);    %filenames of (g)unzipped files

for n=1:length(filename2)
    if length(char(filename2_cell(n,:))) == 23
        filename1(n,:) = [char(filename2_cell(n,:)) '.gz'];
    else filename1(n,:) = [char(filename2_cell(n,:)) '.gz  '];
    end
end
filename1_cell = cellstr(filename1);    %filenames of g-zipped files

save('filenames.mat', 'filename1_cell', 'filename2_cell')
    tempText = [num2str(length(fileTimestr)) ' filenames found and ' ...
```

```
        'stored in filename-array.'];
    disp(tempText)
    clear('tempText')
```

7. FindTMPApositions.m

```
%%%%%% FIND GRID POSITION %%%%%%%%

%This script run as a part of the script TRMMscript_Several.

%Search for grid positions (TMPA) closest to positions of landslides (or
%rain gauges, etc.). The grid positions are stored a variable that tells
%where in the rainfall product files to find rainfall estimates
%corresponding to rainfall of landslide positions in the landslide
%inventory.

%To get these data from one single position, the script
%FindOneTMPAposition.m (almost completely similar to the current script)
%is used.

%% Create latitude and longitude grid position variables
for i=1:1440/2
    lonGridPos(i,:) = 0.125 +(i-1)*0.25;   %longitudes (0-180 E)
end
for i=1:1440/2      %longitude (180 to 360 E = -180 to 0), adjustment to...
    lonGridPos(i+(1440/2),:) = -179.875 +(i-1)*0.25;%...NOAA coord settings
end
for j=1:480
    latGridPos(j,:) = 59.875 -(j-1)*0.25;   %N(pos.latitudes)
                                            %S(neg.latitudes)
end

%% Find grid positions closest to landslides
for n=1:length(Latitude)
    latDiffGrid = abs(Latitude(n,:) - latGridPos);
    lonDiffGrid = abs(Longitude(n,:) - lonGridPos);
    latDiff(n,:) = min(latDiffGrid(:,1));
    find_lat(n,:) = max(find(latDiffGrid == latDiff(n,:)));
    lonDiff(n,:) = min(lonDiffGrid);
    find_lon(n,:) = max(find(lonDiffGrid == lonDiff(n,:)));
end
find_GridPos = [find_lat find_lon Latitude Longitude];

%% Sort by latitude and remove duplicate positions (equal positions)
sortedGrid = sortrows(find_GridPos, 1);
for i=1:length(sortedGrid(:,1))-1
    find_equal(i,1) = sortedGrid(i+1,1) - sortedGrid(i,1);
    find_equal(i,2) = sortedGrid(i+1,2) - sortedGrid(i,2);
end
is_equal = find(abs(find_equal(:,1)) + abs(find_equal(:,2)) == 0);
    %finds positions where absolute value of lat and lon difference is zero
sortedGrid_temp = sortedGrid;
sortedGrid_temp(is_equal+1,:) = [];

%% Sort by longitude and remove duplicates
sortedGrid2 = sortrows(sortedGrid_temp, 2);
for i=1:length(sortedGrid2(:,2))-1
    find_equal2(i,1) = sortedGrid2(i+1,1) - sortedGrid2(i,1);
    find_equal2(i,2) = sortedGrid2(i+1,2) - sortedGrid2(i,2);
end
is_equal2 = find(abs(find_equal2(:,1)) + abs(find_equal2(:,2)) == 0);
```

```
    %finds positions where absolute value of lat and lon difference is zero
filePos_new = sortedGrid2;
filePos_new(is_equal2+1,:) = [];   %all duplicate positions removed
filePos = filePos_new(:,1:2);
LatLonLS = filePos_new(:,3:4);

%% Get grid position coordinates
GridPos_new = [latGridPos(filePos(:,1)) lonGridPos(filePos(:,2))];

%% Get surface distance
deg2rad = (pi/180);
LatDist(:,1) = GridPos_new(:,1)-LatLonLS(:,1); %surface dist lat (Easting)
LonDist(:,1) = GridPos_new(:,2)-LatLonLS(:,2); %surface dist lon (Northing)

%Ellipsoidal Earth projected to a plane
K1 = 111.13209 - 0.56605*cos(2*LatDist) + 0.00120*cos(4*LatDist);
K2 = 111.41513*cos(LatDist) - 0.09455*cos(3*LatDist) + ...
    0.00012*cos(5*LatDist);
for i=1:length(K1)
    SurfDistKM(i,:) = sqrt((K1(i,:)*LatDist(i,:))^2 + ...
        (K2(i,:)*LonDist(i,:))^2);
end

%% Save variables
save('PositionData.mat', 'latDiff', 'find_lat', 'lonDiff', 'find_lon' ...
    , 'filePos', 'GridPos_new', 'SurfDistKM');
    tempText = [num2str(length(GridPos_new(:,1))) ...
        ' unique TMPA positions found.'];
    disp(tempText)
    clear('tempText')
```

8. FindOneTMPAposition.m

```
%%%%%% FIND GRID POSITION %%%%%%%%

%Search for grid positions (TMPA) closest to positions of landslides (or
%rain gauges, etc.). The grid position is stored in variable "sortGrid_new"
%and tell where in rainfall files to find the rainfall equivalent to the
%landslide positions.

%% Create latitude and longitude grid position variables
for i=1:1440/2
    lonGridPos(i,:) = 0.125 +(i-1)*0.25;   %longitudes (0-180 E)
end
for i=1:1440/2      %longitude (180 to 360 E = -180 to 0), adjustment to...
    lonGridPos(i+(1440/2),:) = -179.875 +(i-1)*0.25;%...NOAA coord settings
end
for j=1:480
    latGridPos(j,:) = 59.875 -(j-1)*0.25;   %N(pos.latitudes)
                                            %S(neg.latitudes)
end

%% Find grid positions closest to landslides
for n=1:length(Latitude)
    latDiffGrid = abs(Latitude(n,:) - latGridPos);
    lonDiffGrid = abs(Longitude(n,:) - lonGridPos);
    latDiff(n,:) = min(latDiffGrid(:,1));
    find_lat(n,:) = max(find(latDiffGrid == latDiff(n,:)));
```

```
    lonDiff(n,:) = min(lonDiffGrid);
    find_lon(n,:) = max(find(lonDiffGrid == lonDiff(n,:)));
end
find_GridPos = [find_lat find_lon Latitude Longitude];

%% Sort by latitude and remove duplicate positions (equal positions)
sortedGrid = sortrows(find_GridPos, 1);
sortedGrid_temp = sortedGrid;

%% Sort by longitude and remove duplicates
sortedGrid2 = sortrows(sortedGrid_temp, 2);
filePos_new = sortedGrid2;
filePos = filePos_new(:,1:2);
LatLonLS = filePos_new(:,3:4);

%% Get grid position coordinates
GridPos_new = [latGridPos(filePos(:,1)) lonGridPos(filePos(:,2))];

%% Get surface distance
deg2rad = (pi/180);
LatDist(:,1) = GridPos_new(:,1)-LatLonLS(:,1); %surface dist lat (Easting)
LonDist(:,1) = GridPos_new(:,2)-LatLonLS(:,2); %surface dist lon (Northing)

%Ellipsoidal Earth projected to a plane
K1 = 111.13209 - 0.56605*cos(2*LatDist) + 0.00120*cos(4*LatDist);
K2 = 111.41513*cos(LatDist) - 0.09455*cos(3*LatDist) + ...
    0.00012*cos(5*LatDist);
for i=1:length(K1)
    SurfDistKM(i,:) = sqrt((K1(i,:)*LatDist(i,:))^2 + ...
        (K2(i,:)*LonDist(i,:))^2);
end

%% Save variables
save('PositionData.mat', 'latDiff', 'find_lat', 'lonDiff', 'find_lon' ...
    , 'filePos', 'GridPos_new', 'SurfDistKM');
    tempText = [num2str(length(GridPos_new(:,1))) ...
        ' unique TMPA positions found.'];
    disp(tempText)
    clear('tempText')
```

9. ExtractRainfallData.m

```
%%%%%%%%%%%%%%%%% READ RAINFALL FILES %%%%%%%%%%%%

%This script runs as a part of the script TRMMscript_Several.m

%Rainfall data (TRMM-based estimates) are extracted from all needed files
%and positions, and saved in a rainfall data variable including rainfall
%estimates from all positions and a time string corresponding to each file.


%% Extract rainfall data
ScriptFolder = pwd;     %save current directory as string
Archive = ['C:\Users\Fagerlandet 2.0\Skole\_File Archive\' ...
    RainfallProduct '\'];
cd(Archive)

for n=1:length(fileTime_new)
    %create correct filenames
```

```
    if fileTime_new(n,:) < datenum('01.01.2009', 'dd.mm.yyyy')
         filename2 = [RainfallProduct '.' fileTimestr(n,:) '.bin'];
         filename1 = [RainfallProduct '.' fileTimestr(n,:) '.bin.gz'];
         Folder=[Archive 'V5\' fileTimestr(n,1:4) '\' fileTimestr(n,5:6)];
         cd(Folder)     %fileTimestr: 1-4=year, 5-6=month
            %opens the correct folder from the data archive
    else filename2 = [RainfallProduct '.' fileTimestr(n,:) '.6.bin'];
         filename1 = [RainfallProduct '.' fileTimestr(n,:) '.6.bin.gz'];
         Folder = [Archive fileTimestr(n,1:4) '\' fileTimestr(n,5:6)];
         cd(Folder)
    end
    try     %prosess file if present in archive folder
      gunzip(filename1);
      fileID = fopen(filename2, 'r', 'b', 'ISO-8859-1');
        %opens binary file, read only, using machine format "big-endian"
        %and textformate ISO-8859-1
      rainfallGrid = fread(fileID, [1440 480], 'int16', 0, 'b');
        %Read file in the given range, as 2 byte integer (int16), skipping
        %no bits(0), using machineformat "big-endian".
      errorGrid = fread(fileID, [1440 480], 'int16', 0, 'b'); %2 byte int
      sourceGrid = fread(fileID, [1440 480], 'int8', 0, 'b'); %1 byte int
      %uncalRFgrid = fread(fileID, [1440 480], 'int16', 0, 'b'); %2 byte
        %integer (uncalRFgrid is only used on 3BXXRT version 6 files)
      fclose(fileID);
      delete(filename2);

      for i=1:length(filePos(:,1))  %Rainfall values in file is *100 mm/hr
        Rainfall(n,i) = rainfallGrid(filePos(i,2),filePos(i,1))/100;
        Error(n,i) = errorGrid(filePos(i,2),filePos(i,1))/100;
        Source(n,i) = sourceGrid(filePos(i,2),filePos(i,1));
      end
            tempText = ['File number ' num2str(n) ' (' filename2 ...
                ') processed.'];
            disp(tempText)
            clear('tempText')
    catch
        %if file is not present in arcive folder, values = -9
        tempText = ['File ' filename1 ' not found in archive.' ...
            ' Rainfall = -9.'];
        disp(tempText)
        clear('tempText')
        for i=1:length(filePos(:,1))
            Rainfall(n,i) = -9;
            Error(n,i) = -9;
            Source(n,i) = -9;
        end
    end

end

ScriptFolder = 'C:\Users\Fagerlandet 2.0\Skole\Masteroppgave\Matlab\TRMM';
cd(ScriptFolder)
    tempText = [num2str(length(fileTime_new)) ' files of rainfall data' ...
        ' processed and saved in rainfall-array.'];
    disp(tempText)
    clear('tempText')
```

10. SaveData.m

```
%%%%%%%%%%%%%% SAVE DATA INTO FINAL RESULT FILE %%%%%%%%%%
```

```
583
584 %This script runs as a part of the script TRMMscript_Several.m.
585
586 %Saves all rainfall and position data in *.mat-files in the folder Results.
587 %Name of saved files corresponds to the name of the landslide inventory
588 %file (given in InsertData.m).
589
590 %Data needed        - Variables:
591     %Latitude           - LatLonLS
592     %Longitude          - LatLonLS
593     %Latitude_TMPA      - GridPos_new
594     %Longitude_TMPA     - GridPos_new
595     %SurfaceDistKM      - SurfDistKM
596     %
597     %RainfallData       - Rainfall
598     %ErrorData          - Error
599
600 if isdir('Results') == 0
601     mkdir('Results')
602 end
603 cd Results
604
605 %% Save position data
606 PositionData(1:2,:) = LatLonLS';
607 PositionData(3:4,:) = GridPos_new';
608 PositionData(5,:) = SurfDistKM';
609
610 Header = {'LS_latitude'; 'LS_longitude'; 'TMPA_latitude'; ...
611     'TMPA_longitude'; 'SurfaceDistance(KM)';};
612
613 PositionData_cell = [Header num2cell(PositionData)]';
614
615 tempName = [xlsName(1:max(strfind(xlsName, '.'))-1)];
616 PosFileName = [tempName '_position data.mat'];
617 save(PosFileName, 'PositionData', 'Header', 'PositionData_cell');
618
619 %% Save rainfall data
620 TMPA_Rainfall(1:2,:) = PositionData(3:4,:);
621 TMPA_Rainfall(3:length(Rainfall(:,1))+2,:) = Rainfall;
622 TMPA_Rainfall_header(3:length(fileTimestr)+2,:) = fileTimestr;
623 TMPA_Rainfall_header(1:2,:) = ['TMPA_lat  '; 'TMPA_long '];
624 TMPA_Rainfall_cell=[cellstr(TMPA_Rainfall_header) num2cell(TMPA_Rainfall)];
625
626 TMPA_Error(1:2,:) = PositionData(3:4,:);
627 TMPA_Error(3:length(Error(:,1))+2,:) = Error;
628 TMPA_Error_header(3:length(fileTimestr)+2,:) = fileTimestr;
629 TMPA_Error_header(1:2,:) = ['TMPA_lat  '; 'TMPA_long '];
630 TMPA_Error_cell=[cellstr(TMPA_Error_header) num2cell(TMPA_Error)];
631
632 RainfallFilename = [tempName '_rainfall data.mat'];
633 save(RainfallFilename, 'TMPA_Rainfall', 'TMPA_Rainfall_cell', ...
634     'TMPA_Error_cell', 'TMPA_Error')
635
636 cd ..
637
638 disp('Rainfall data saved in \Results.')
```

11. GetDailyRainfallNew.m

```
%%%%%%%%%%%%% ESTIMATE DAILY RAINFALL FROM NASAS 3B41RT/3B42RT %%%%%%%%%%

%Files resulting from running the script TRMMscript_Severeal.m is needed to
%run this script, and must be present in folder 'Results' in current Matlab
%directory. Filenames and rainfall product type must be inserted before
%running the script.

%The rainfall data loaded into this file contain the coordinates to the
%rainfall data and rainfall data each hour. The rainfall is given in mm/hr.

%The mean time of each file (time for data collection) is at 00, 01, 02,
%etc. Thus data collection related to a file is actually at +/- 0.5 or 1.5
%hours depending on the type of rainfall product, and that the rainfall
%files of 00-hours and 01-hours, for 3B42RT and 3B41RT respectively, must
%be divided between two days.

%Bangladesh is located in UTC zone +6, and the local rainfall data is
%probably collected at 07:00 (equal to 01:00 UTZ 0). The 01-hours files
%rainfall data thus has to be divided between two days: 00:30 to 01:00
%(0.5 hours) are related to the previous day, and 01:00 to 01:30
%(0.5 hours)to the coming day. The rainfall intensity given in the file in
%(mm/hr) is spread evenly between the two days => 0.5hr*rainfall(RF).

FilenameBeginning = 'Full time series';    %enter filename of result-file
    %leave out the ending, like '_position data' or '_rainfall data'.
RainfallProduct = '3B41RT';     %select type of rainfall product
                                %(3B41RT or 3B42RT).
if RainfallProduct == '3B41RT'
    SplitFileTime = 1;
    PartPrevDay = 0.5;
    FileDuration = 1;
end
if RainfallProduct == '3B42RT'
    SplitFileTime = 0;
    PartPrevDay = 2.5;
    FileDuration = 3;
end
    PartCurrDay = 0.5;

cd Results
Loadfilename1 = [FilenameBeginning '_rainfall data.mat'];
Loadfilename2 = [FilenameBeginning '_position data.mat'];
load(Loadfilename1)
load(Loadfilename2)
FileName = [FilenameBeginning '_daily rainfall data.mat'];
LatLonTMPA = [TMPA_Rainfall_cell(1,2) TMPA_Rainfall_cell(2,2)];
AllTimes = datenum(char( ...
    TMPA_Rainfall_cell(3:length(TMPA_Rainfall_cell),1)), 'yyyymmdd');
Startingday = AllTimes(1);
Endingday = AllTimes(length(AllTimes));
NrOfDays = Endingday - Startingday +1;
CountRFseries = length(TMPA_Rainfall(1,:));

    for n=1:CountRFseries
        temp_Rainfall = TMPA_Rainfall(:,n);
for i=1:NrOfDays
    NaNcount(i,:) = 0;
    DailyRainfall(i,:) = 0;
```

```
    CurrDay(i,:) = Startingday + (i-1);  %numerical day (for each day in
loop)
    findRF = find(AllTimes == CurrDay(i,:));
        %selects rainfall-data positions of current day (8 for each day)
    for j=min(findRF):max(findRF)   %for each day...
        Datestr = char(TMPA_Rainfall_cell(j+2,1));
        if (temp_Rainfall(j+2) == -9) || ...    %count hours of NaN
                (temp_Rainfall(j+2) == -319.99);  %2 header lines in TMPA_
            if str2num(Datestr(9:10)) == SplitFileTime
                if (i-1)>0
                    NaNcount(i-1,:) = NaNcount(i-1,:) + PartPrevDay;
                        %adds 2.5 or 0.5 hrs of NaN to previsious day
                end
                NaNcount(i,:) = NaNcount(i,:) + PartCurrDay; %0.5 hrs NaN
            else NaNcount(i,:) = NaNcount(i,:) + FileDuration; %3 hrs NaN
            end
        else  %accumulate rainfall
            if str2num(Datestr(9:10)) == SplitFileTime
                if (i-1)>0
                  DailyRainfall(i-1,:) = DailyRainfall(i-1,:) + ...
                    temp_Rainfall(j+2,:) * PartPrevDay; %2.5 or 0.5 hrs RF
                end
                DailyRainfall(i,:) = DailyRainfall(i,:) + ...
                    temp_Rainfall(j+2,:) * PartCurrDay; %0.5 hours of RF
            else DailyRainfall(i,:) = DailyRainfall(i,:) + ...
                    temp_Rainfall(j+2,:) * FileDuration; %1 or 3 hrs of RF
            end
        end
    end
end
DailyRainfall_cell(3:length(DailyRainfall)+2,n+1) =
num2cell(DailyRainfall);
NaNcount_cell(3:length(DailyRainfall)+2,n+1) = num2cell(NaNcount);
    end

DailyRainfall_cell(1:2,1:length(PositionData_cell(:,1))) = ...
    [PositionData_cell(1:length(PositionData_cell(:,1)),3:4)'];
DailyRainfall_cell(3:length(DailyRainfall)+2,1) = ...
    cellstr(datestr(CurrDay, 'yyyymmdd'));

NaNcount_cell = DailyRainfall_cell;
NaNcount_cell(3:length(DailyRainfall)+2,1) = ...
    cellstr(datestr(CurrDay, 'yyyymmdd'));

save(FileName, 'DailyRainfall', 'DailyRainfall_cell', 'CurrDay', ...
    'NaNcount', 'NaNcount_cell')

cd ..
```

12. ValidationStatistics.m

```
%%%%%%% VALIDATION STATISTICS OF RAINFALL ESTIMATES %%%%%%%%

%This script is used to compare TRMM-based rainfall estimates
%with daily rain gauge data from Chittagong. The rainfall estimate products
%used are the NASA TRMM 3B41RT using
%infrared imagery (IR) and the NASA TRMM 3B42RT using merged IR and
%microwave imagery (IRMW). The rainfall series selected is the longest
%possible providing complete years of rainfall from all sources
```

```
LoadData;   %load all data needed to create data series
PrepareSeries;  %prepare rainfall for equal time series
LandslideData;  %prepare landslide data for use in validation

%% Visual inspection
VisInspection;      %prepare rainfall series for visual inspection
DistributionCurve;  %prepare data for making distrubution plot
MontlyRainfall;     %prepare data for monthly rainfall figures
FigureSetup;        %create bar-plot and landslide event data line-plot

%% Correlation
CorrelationPlots;       %prepare graphs for correlation

%% Check NaN
CheckNaN;       %analyse severity of missing rainfall estimates
```

13. LoadData.m

```
%% Load IR, IRMW and HE data for use in ValidationStatistics.m
CurrentFolder = pwd;
IRfolder = ['C:\Users\Fagerlandet 2.0\Skole\Masteroppgave\Matlab' ...
    '\TRMM\3B41RT\Results'];   %folder of 3B41RT (IR estimates)
IRMWfolder = ['C:\Users\Fagerlandet 2.0\Skole\Masteroppgave\Matlab' ...
    '\TRMM\3B42RT\Results'];   %folder of 3B42RT (combined IR/MW estimates)
DailyRFfile = 'Raingauge_Daily_Chittagong_daily rainfall data.mat';

cd(IRfolder)
load(DailyRFfile)
rainfallIR = DailyRainfall;     %Get IR data (3B41RT)
dateIR = CurrDay;
nanIR = NaNcount;

cd(IRMWfolder)
load(DailyRFfile)
rainfallIRMW = DailyRainfall;   %Get IR/MW data (3B42RT)
dateIRMW = CurrDay;
nanIRMW = NaNcount;

cd(CurrentFolder)
load('Validation file_Chittagong.mat')
ValStat(1,:) = [];  %remove header line
rainfallHE = cell2mat(ValStat(:,5));
dateHE = datenum(cell2mat(ValStat(:,1:3)));
nanHE = cell2mat(ValStat(:,6));

load('ChittagongDailyRainGaugeData.mat')
rainfallRG = cell2mat(DailyRainfall_RG(:,4));  %Get rain gauge data
dateRG = cell2mat(DailyRainfall_RG(:,5));
```

14. PrepareSeries.m

```
%% Create data series of similar dates for IR, IR/MW and RG data, for use
%% in ValidationStatistics.m

startdate = '01-Mar-2005';  %No rainfall in any data sets before march 05
```

```
startdatenum = datenum(startdate);
enddate = '31-Dec-2008';    %Rainfall registered until end of December 08
enddatenum = datenum(enddate);

%Find start and end possitions for rainfall data corresponding to dates
startIR = find(dateIR == startdatenum);
endIR = find(dateIR == enddatenum);
startIRMW = find(dateIRMW == startdatenum);
endIRMW = find(dateIRMW == enddatenum);
startRG = find(dateRG == startdatenum);
endRG = find(dateRG == enddatenum);

%Create final rainfall data series
rainfalldates_new = dateRG(startRG:endRG);
rainfallIR_new = rainfallIR(startIR:endIR);
rainfallIRMW_new = rainfallIRMW(startIRMW:endIRMW);
rainfallRG_new = rainfallRG(startRG:endRG);
rainfallCompare_new = [rainfalldates_new rainfallIR_new ...
    rainfallIRMW_new rainfallRG_new];
```

15. VisInspection.m

```
%% Initial visual inspection of the data

x = [1:1:length(rainfallRG_new)];
y = rainfallIR_new;
y(:,2) = rainfallIRMW_new;
y(:,3) = rainfallRG_new;
plot(x, y)
legend('IR', 'IRMW', 'RG');
title('Daily rainfall correlation');
```

16. LandslideData.m

```
%This script loads landslide data from spread sheet into Matlab workspace.
%Date and time vectors for landslide events in landslide inventory are
%created and events of too high position uncertainty are removed.

xlsName = 'ChittagongDistrict.xlsx';
sheetName = 'Landslides';
Range = 'A1:U15';   %Insert range of landslide data in sheet.
startline = 2;      %Select range for input in analysis (skip header line).
endline   = 15;

%Insert column position in spread sheet of the different data required:
colNr_LSnr = 1;
colNr_LSname = 3;
colNr_LSlat = 7;
colNr_LSlong = 8;
colNr_Year = 14;
colNr_Month = 15;
colNr_Day = 16;
colNr_Region = 5;
colNr_District = 4;
colNr_ConfRadius = 13;
colNr_Casualties = 12;
```

```
%Read inventory file
[X, Y, Z] = xlsread(xlsName, sheetName, Range);
    %X = numerical data, Y = string data as cell, Z = all data as cell
    disp('Xls-file read.');

%Remove data with very high spatial uncertainty
ConfRadius = Z(2:length(Z(:,1)),colNr_ConfRadius);
Unacurate = find(cell2mat(ConfRadius(:,1)) > 50);
Z(Unacurate+1,:) = [];

%Get time data
LStime_cell = Z(:,colNr_Year:colNr_Day);
LStime_cell(1,:) = [];   %remove header line
LStime_num = datenum(cell2mat(LStime_cell));
timenum_sorted = sort(LStime_num);
for i=1:length(LStime_num)-1
    timediff(i,:) = timenum_sorted(i+1,:) - timenum_sorted(i,:);
end
finddiff = find(timediff == 0);
LStime_num_new = timenum_sorted;
LStime_num_new(finddiff) = [];

LStime = datestr(LStime_num_new, 'dd.mm.yyyy');

%Prepare sorting of LS event by month
mm = str2num(LStime(:,5:6))     %used in visual inspection figure (se
                                %MontlyRainfall2.m).
```

17. DistributionCurve.m

```
%% Compare distribution of different data

%This script create a bar-plot representing a rainfall intensity
%distribution curve for each rainfall products, by plotting intensity
%intervals. Figure was adjusted using Matlab figure GUI.

clear('distrIR_new', 'distrIRMW_new', 'distrRG_new')
Resolution = 20; %resolution of distribution columns
for j=1:20
    distrIR_new(j,:) = length(find(rainfallIR_new > (j-1)*Resolution & ...
        rainfallIR_new <= (j)*Resolution));
    distrIRMW_new(j,:) = length(find(rainfallIRMW_new > (j-1)* ...
        Resolution & rainfallIRMW_new <= (j)*Resolution));
    distrRG_new(j,:) = length(find(rainfallRG_new > (j-1)*Resolution & ...
        rainfallRG_new <= (j)*Resolution));
end

distrAll = [distrIR_new distrIRMW_new distrRG_new]/4;

%% Create bar diagram of yearly average rainfall intensity distribution

%Add data
y = distrAll    %y1 = IR, y2 = IRMW, y3 = RG
x = .5:1:19.5;

%Create figure
```

```
F = figure('Color','w','Position',[100 100 900 450]);   %figure properties:
background color,
b = bar(x,y,1,'grouped');   %plot data in figure with parameters x-axis,
                            %y-axis and plot format
colormap(   [0.8157 0.2549 0.2549;...
             0.4784 0.7098 0.2627; ...
             0.3333 0.4745 0.6157;] ...
         )
%Add title and labels
title('Rainfall distribution, yearly average', ...
    'FontWeight','bold','FontSize',18)
ylabel('Rainfall (mm)','FontSize',18)

%Change figur properties
box  on
set(gca, 'FontSize',10,'FontName','Kalinga')
set(gca, 'YGrid','On')
set(gca, 'YTick', 10:10:120, 'XTick', 0:20)
set(gca, 'XTickLabel', (x-0.5)*Resolution)
```

18. MonthlyRainfall.m

```
%% Creates monthly average rainfall values for plotting monthly rainfall
values. Plots where made using GUI

%The variable 'rainfalldates_new' give the numerical date for the whole 4
      %year data series.
startdatenum = min(rainfalldates_new);
enddatenum = max(rainfalldates_new);
NrOfDays = enddatenum - startdatenum +1;

%Create numerical days, months and years for data handling/processing
for i=1:NrOfDays
    iAllDays(i,:) = str2num(datestr(startdatenum+(i-1), 'dd'));
    iAllMonths(i,:) = str2num(datestr(startdatenum+(i-1), 'mm'));
    iAllYears(i,:) = str2num(datestr(startdatenum+(i-1), 'yyyy'));
end

%% Create histogram variables
montly_LS = [0 0 0 0 0 2 2 2 1 0 0 0];  %number of landslide events (most
    %of multiple landslides) in the Chittagong district from the landslide
    %inventory, sorted on month of occurrence. Data found in variable mm
    %from LandslideData.m.
montly_LSall = [0 0 0 0 1 5 5 3 1 1 0 0 ];
montly_LS = montly_LS/4;
montly_LSall = montly_LSall/4;

for i=1:12  %for all 12 months (Jan-Dec)
    iOneMonth = find(iAllMonths == i);   %get position of month(i) rainfall
    NrOfYears = round(NrOfDays/365);

  %Montly average rainfall
    iAvgRF_IR(i,:) = sum(rainfallIR_new(iOneMonth))/NrOfYears;
    iAvgRF_IRMW(i,:) = sum(rainfallIRMW_new(iOneMonth))/NrOfYears;
    iAvgRF_RG(i,:) = sum(rainfallRG_new(iOneMonth))/NrOfYears;
    SumRainfall = [sum(iAvgRF_IR) sum(iAvgRF_IRMW) sum(iAvgRF_RG)]
```

```
984       %FIGURES:   Histograms of average monthly rainfall created using GUI
985       %           and the variables iAvgRF_IR, iAvgRF_IRMW and iAvgRF_RG
986       %           in separated columns. A second y-axis was created to
987       %           include landslide data (variable montly_LS).
988
989
990     %Monthly average rainfall (using thresholds)
991       Threshold = [90 50 30];     %big landslide event had 88 mm on event day
992       for T = 1:length(Threshold)
993
994           %Above threshold
995           findT_IRa =find(rainfallIR_new > Threshold(1,T) & iAllMonths == i);
996           findT_IRMWa = ...
997               find(rainfallIRMW_new > Threshold(1,T) & iAllMonths == i);
998           findT_RGa =find(rainfallRG_new > Threshold(1,T) & iAllMonths == i);
999
1000          iAvgRF_IR_Ta(i,T) = sum(rainfallIR_new(findT_IRa))/NrOfYears;
1001          iAvgRF_IRMW_Ta(i,T) = sum(rainfallIRMW_new(findT_IRMWa))/NrOfYears;
1002          iAvgRF_RG_Ta(i,T) = sum(rainfallRG_new(findT_RGa))/NrOfYears;
1003
1004          %Below threshold
1005          findT_IRb =find(rainfallIR_new < Threshold(1,T) & iAllMonths == i);
1006          findT_IRMWb = ...
1007              find(rainfallIRMW_new < Threshold(1,T) & iAllMonths == i);
1008          findT_RGb =find(rainfallRG_new < Threshold(1,T) & iAllMonths == i);
1009
1010          iAvgRF_IR_Tb(i,T) = sum(rainfallIR_new(findT_IRb))/NrOfYears;
1011          iAvgRF_IRMW_Tb(i,T) = sum(rainfallIRMW_new(findT_IRMWb))/NrOfYears;
1012          iAvgRF_RG_Tb(i,T) = sum(rainfallRG_new(findT_RGb))/NrOfYears;
1013      end
1014
1015      %FIGURES:   Same type of histograms as the one above was created for
1016      %           the variables where only rainfall above("variable"_Ta) or
1017      %           below("variable"_Tb) a treshold value (T) is included
1018      %           (variable_Tx(i,T). Landslides not included here.
1019
1020
1021    %Count number of rainfall days
1022      iRFdays = find(iAllMonths == i);
1023      countAvgRF_IR(i,:) = ...
1024          length(find(rainfallIR_new(iRFdays) > 0))/NrOfYears;
1025      countAvgRF_IRMW(i,:) = ...
1026          length(find(rainfallIRMW_new(iRFdays) > 0))/NrOfYears;
1027      countAvgRF_RG(i,:) = ...
1028          length(find(rainfallRG_new(iRFdays) > 0))/NrOfYears;
1029
1030      BHKDF = [sum(countAvgRF_IR) sum(countAvgRF_IRMW) sum(countAvgRF_RG)];
1031      %FIGURE:    Histogram presenting the average number of rainfall days
1032      %           each month for each rainfall product created using GUI and
1033      %           the variables countAvgRF_IR, countAvgRF_IRMW and
1034      %           countAvgRF_RG.
1035
1036  end
1037
1038
```

19. CorrelationPlots.m

```
%% This script is used for creating correlation plots for comparing
different rainfall sources. Figures was adjusted using GUI before presented
in thesis.

%create julian dates
z_temp_num = rainfalldates_new; %datenum corresponding to RG- and IR-data
yyyy = datestr(z_temp_num, 'yyyy');
for i=1:length(yyyy(:,1))
    removeYear(i,:) = ['01-01-' yyyy(i,:)];
end
removeYear_num = datenum(removeYear);
julianDay = z_temp_num - removeYear_num + 1;

%Prepare data for plotting
x = [rainfallRG_new(2:length(rainfallRG_new),1); 0]  %raingauge data
y = rainfallIR_new; %TMPA IR-data
y2 = rainfallIRMW_new;
z = julianDay;      %data may be sorted by julian date in z-axis, creating
                    %different colors depending on season
Plot_1to1 = 1:500;    %get values for plotting 1 to 1-line

%Plot figure 1
figure('Color','w','Position',[100 100 1000 500]);
subplot(1,2,1)
Corr = scatter(x, y, 30, z, 'filled')
    %(x-axis, y-axis, marker size, plot color)
        %plot color may be one color ([x x x]) og specter defined on a
        %z-variable.
hold on
perfectCorr = plot(Plot_1to1,Plot_1to1);

%Add title and labels
title('Correlation of daily rainfall', ...
    'FontWeight','bold','FontSize',24)
ylabel('TMPA IR (mm)','FontSize',20)
xlabel('Rain gauge (mm)','FontSize',20)

%Change figur properties
set(gca, 'XScale','log', 'YScale','log')
axis square
box  on
%colormap('hot')
set(Corr,'Marker','o')
set(gca, 'FontSize',16,'FontName','Kalinga')

%set color map to plot
set(perfectCorr,'LineWidth',1, 'Color',[0 0 0])
    %Colour map is edited using the function colormap editor.


%% Plot figure 2
subplot(1,2,2)
%F = figure('Color','w','Position',[100 100 1000 500]);
Corr2 = scatter(x, y2, 30, z, 'filled')
    %(x-axis, y-axis, marker size, plot colour)
        %plot colour may be one colour ([x x x]) of spectre defined on a
        %z-variable.
hold on
perfectCorr = plot(Plot_1to1,Plot_1to1);
```

```
%Add title and labels
title('Correlation of daily rainfall', ...
    'FontWeight','bold','FontSize',24)
ylabel('TMPA IRmicro (mm)','FontSize',20)
xlabel('Rain gauge (mm)','FontSize',20)

%Change figur properties
set(gca, 'XScale','log', 'YScale','log')
axis square
box  on
%colormap('hot')
set(Corr2,'Marker','o')
set(gca, 'FontSize',16,'FontName','Kalinga')

%set color map to plot
set(perfectCorr,'LineWidth',1, 'Color',[0 0 0])


%% Plot TMPA-correlation
F3 = figure('Color','w','Position',[100 100 700 500]);
Corr3 = scatter(y, y2, 30, [0 0 0], 'filled');
hold on
perfectCorr = plot(Plot_1to1,Plot_1to1);

%Add title and labels
title('Correlation of IR and IRmicro', ...
    'FontWeight','bold','FontSize',24)
ylabel('TMPA IRmicro (mm)','FontSize',20)
xlabel('TMPA IR (mm)','FontSize',20)

%Change figur properties
set(gca, 'XScale','log', 'YScale','log')
axis square
box  on
colormap('hot')
set(Corr3,'Marker','o')
set(gca, 'FontSize',16,'FontName','Kalinga')

%set color map to plot
set(perfectCorr,'LineWidth',1, 'Color',[0 0 0])
```

20. CheckNaN.m

```
%% investigate the NaNs in the rainfall data estimate series

%Create NaN-series corresponding to rainfall series (from PrepareSeries.m)

%ValDate = dateRG(startRG:endRG);
nanIR_new = nanIR(startIR:endIR);
nanIRMW_new = nanIRMW(startIRMW:endIRMW);

%Find positions where rainfall data are missing by number of NaN hours pr
%day. Data is divided in groups;
%   days with more than 18 hours of data missing (18 hours of NaN)
%   days with more than 12 hours of NaN
%   days with < 6 hours of NaN
```

```
%   days with < 0 hours of NaN
NrOfNan0 = 0;   NrOfNan6 = 6;   NrOfNan12 = 12;   NrOfNan18 = 18;

nanPosIR0 = find(nanIR_new > NrOfNan0);
nanPosIRMW0 = find(nanIRMW_new > NrOfNan0);
nanPosIR6 = find(nanIR_new > NrOfNan6);
nanPosIRMW6 = find(nanIRMW_new > NrOfNan6);
nanPosIR12 = find(nanIR_new > NrOfNan12);
nanPosIRMW12 = find(nanIRMW_new > NrOfNan12);
nanPosIR18 = find(nanIR_new > NrOfNan18);
nanPosIRMW18 = find(nanIRMW_new > NrOfNan18);

NrOfNanIR = [length(nanPosIR0) length(nanPosIR6) ...
    length(nanPosIR12) length(nanPosIR18)];
NrOfNanIRMW = [length(nanPosIRMW0) length(nanPosIRMW6) ...
    length(nanPosIRMW12) length(nanPosIRMW18)];

missingIR18 = rainfallIR_new(nanPosIR18);
missingIR12 = rainfallIR_new(nanPosIR12);
missingIR6  = rainfallIR_new(nanPosIR6);
missingIR0  = rainfallIR_new(nanPosIR0);
missingIRavg = [sum(missingIR0)/length(missingIR0) ...
    sum(missingIR6)/length(missingIR6) ...
    sum(missingIR12)/length(missingIR12) ...
    sum(missingIR18)/length(missingIR18)];
missingIRmax = [max(missingIR0) max(missingIR6) max(missingIR12) ...
    max(missingIR18)];

missingIRMW18 = rainfallIRMW_new(nanPosIR18);
missingIRMW12 = rainfallIRMW_new(nanPosIR12);
missingIRMW6  = rainfallIRMW_new(nanPosIR6);
missingIRMW0  = rainfallIRMW_new(nanPosIR0);
missingIRMWavg = [sum(missingIRMW0)/length(missingIRMW0) ...
    sum(missingIRMW6)/length(missingIRMW6) ...
    sum(missingIRMW12)/length(missingIRMW12) ...
    sum(missingIRMW18)/length(missingIRMW18)];
missingIRMWmax = [max(missingIRMW0) max(missingIRMW6) max(missingIRMW12)...
    max(missingIRMW18)];
```

21. ThresholdAnalysis.m

```
%%%%%% CREATE THRESHOLDS %%%%%%

%Load threshold inventory
TMPAproduct = '3B42RT';
cd Results
FileName = [TMPAproduct '_threshold inventory.mat'];
load(FileName)
cd ..
    disp('Files loaded.')

%Variables loaded from file:
%PredDate_num: List of continuous dates in numerical date values (T)
%A_RFdaily: 3-dimensional vector of all rainfall data corresponding to the
%           numerical dates. Rainfall is accumulated antecedent rainfall of
%           different durations(D) up to and including the date (T). This
%           way D=1 gives daily rainfall at date T. Rainfall in (mm/day).
```

```
%           Each line correspond to a day, each row to a TMPA-lat/lon
%           position (P), and each layer in 3rd dimension to a duration.
%           [T x P x D].
%A_RF_dailyAvg: Same vector format as A_RFdaily, but numerical values is
%               given in average daily rainfall (TA) for the duration D,
%               as TA = T(D)/D [mm/day].
%LS_outcome: Predictor of if landslides happened or not. The
%             variable contain only 'y' and 'n' in a T x P-matrix.
%NaNacumulated: Count of missing hours corresponding to the acumulated
%               rainfall series A_RFdaily.
%Lat: Latitude TMPA position for rainfall data in A_RFdaily
%Lon: Longitudal TMPA position

%% Prepare data
PrepRFdata;     %change structure of RF data and remove NaN inputs

%% Pre-analyses
Corr = corr(A_RF_d_noNaN);     %correlation of all predictors
    disp('Correlation analysis done.')

%Extract data related to landslide days
ExtractYdays;

%% Classification tree analysis (CTA)
ClassTreeAnalysis;

%% Discriminant analysis
```

22. PrepareRFdata.m

```
%% Prepare data

%Changes structure of data and remove NaN inputs

%Variables prepared here change [T x P x D]-matrixes to [T' x P]-matrixes,
%as explained above, where T' = T x D. This is to prepare the data for
%2D-ploting.

%%
%landslide triggering- and rainfall data
    count = 0;
    tic;
for n=1:length(LS_outcome(1,:))
    for i=1:length(LS_outcome(:,1))
        count = count + 1;  %used to output data in order; line 1,2,3, etc.
        Dates(count,:) = datestr(PredDate_num((i),:), 'dd.mm.yyyy');
            %output dates of landslides (for verification of dates)
        A_RF_dailyAvg_new(count,:) = A_RF_dailyAvg((i),n,:);
            %outputs different predictors for threshold analysis
        NaNacumulated_new(count,:) = NaNacumulated((i),n,:);
            %outputs the total amount of NaNs in the corresponding rainfall
            %data vector A_RF_dailtAvg_new.
        LS_outcome_new(count,:) = LS_outcome((i),n,:);
            %new outcome vector in format [T' x P]
    end
end
clear count
```

```
toc;

%remove lines containing NaN
A_RF_d_noNaN = A_RF_dailyAvg_new;
LS_outcome_noNaN = LS_outcome_new;
isNaN = isnan(A_RF_d_noNaN(:,13));
findNaN = find(isNaN == 1);   %locate all lines with no actual number (NaN)

Dates(findNaN,:) = [];
A_RF_d_noNaN(findNaN,:) = [];
NaNacumulated_new(findNaN,:) = [];
LS_outcome_noNaN(findNaN,:) = [];

%create logarithmig rainfall variables
A_RF_d_log = log10(A_RF_d_noNaN);
find_inf = find(A_RF_d_log == -Inf);    %remove infinite values; cannot be
A_RF_d_log(find_inf) = log10(0.001);    %included in discriminant analysis


%create vector of durations corresponding to the predictors
for i=1:length(A_RF_d_noNaN(:,1))
    D_new(i,:) = [1 2 3 4 5 6 7 9 11 13 15 30 60];
end
    disp('Data prepared.')
```

23. ExtractYdays.m

```
%Extract data related to landslide days
    count = 0;
for n=1:length(LS_outcome(1,:))
    findLS = find(LS_outcome(:,n) == 'y');    %find landslide days
    for i=1:length(findLS)
        count = count + 1;  %used to output data in order; line 1,2,3, etc.
        LS_yes(count,:) = A_RF_dailyAvg(findLS(i),n,:);
            %outputs different predictors for threshold analysis
        LSdate_y(count,:) = datestr(PredDate_num(findLS(i),:), ...
            'dd.mm.yyyy');
            %output dates of landslides (for verification of dates)
        LS_NaNs_y(count,:) = NaNacumulated(findLS(i),n,:);
        LStempPos_y(count,:) = n;
        LSpos_y(count,:) = [Lat(n) Lon(n)];
    end
end
clear count
```

24. ClassTreeAnalysis.m

```
%% Classification tree analysis (CTA)

t = classregtree(A_RF_dailyAvg_new,  LS_outcome_new);

view (t)

%-------------------------------------------------------------------------%
```

```
%CTA adjusted for predictor correlation%

%P2 - P1 and P3-4 are left out of analysis
A_RF_CTAtemp = [A_RF_dailyAvg_new(:,2) A_RF_dailyAvg_new(:,5:13)];
t2 = classregtree(A_RF_CTAtemp,  LS_outcome_new);
view (t2)

%P3 - P2 and P4-5 are left out of analysis
A_RF_CTAtemp = [A_RF_dailyAvg_new(:,1) A_RF_dailyAvg_new(:,3) ...
    A_RF_dailyAvg_new(:,6:13)];
t3 = classregtree(A_RF_CTAtemp,  LS_outcome_new);
view (t3)

%P5 - P3-4 and P6-8 are left out of analysis
A_RF_CTAtemp = [A_RF_dailyAvg_new(:,1:2) A_RF_dailyAvg_new(:,5) ...
    A_RF_dailyAvg_new(:,9:13)];
t5 = classregtree(A_RF_CTAtemp,  LS_outcome_new);
view (t5)

%P6 - P4-5 and P7-9 are left out of analysis
A_RF_CTAtemp = [A_RF_dailyAvg_new(:,1:3) A_RF_dailyAvg_new(:,6) ...
    A_RF_dailyAvg_new(:,10:13)];
t6 = classregtree(A_RF_CTAtemp,  LS_outcome_new);
view (t6)

%P2 and P5 - P1, P3-4 and P6-8 are left out of the analysis
A_RF_CTAtemp = [A_RF_dailyAvg_new(:,2) A_RF_dailyAvg_new(:,5) ...
    A_RF_dailyAvg_new(:,9:13)];
t2_5 = classregtree(A_RF_CTAtemp,  LS_outcome_new);
view (t2_5)

%custom P2 and X - P1, P3-8 are left out of the analysis
A_RF_CTAtemp = [A_RF_dailyAvg_new(:,2)  ...
    A_RF_dailyAvg_new(:,9:13)];
t2custom = classregtree(A_RF_CTAtemp,  LS_outcome_new);
view (t2custom)
```

25. DiscriminantAnalysis.m

```
%This script run discriminant analysis according to methodology developed
%for this thesis.Final output of discriminant function and scatter plot are
%produced.

%Input data for analysis
classType = 'linear';   %set classification type 'linear' or 'quadratic'
Range_min = 1;
Range_max = length(A_RF_d_noNaN(1,:));
DispFig = 'n';     %boolean 'y' or 'n'
resolu_Pri = 0.01;     %resolution of prior probability testing
priStart = 1.00;        %1.00 for resolution 0.05, 0.01, 0.005
                        %0.951 for resolution 0.001
NrOfIntervals = 20;     %19 for res 0.05, 20 for res 0.01, 21 for res 0.001

%give N according to what prior probability range to do discr. analysis
                            for N=1:NrOfIntervals               tic;
priY(1,N) = priStart - N*resolu_Pri;
```

```
priority = [1-priY(1,N) priY(1,N)];    %prior probability for linear class


%Run DA for all predictor combinations
if DispFig == 'y';
    figure('Color','w', 'Position',[10 10 780 680], 'Name', ...
        'Discriminant analysis');    %add figure for plott
end
classError = zeros(13,13);
classError(find(classError == 0)) = 999;
DA_subplots_linear;    %RUNS DA-ANAYSIS!


%Locate minimum ERR and corresponding variables for each prior prob (N)
find0 = find(classError == 0);    %remove all zeros in err-variable and
classError(find0) = 999;          %replace with 999
find_best_MinERR(N,:) = min(classError);     %find min of each col
best_MinERR(N,:) = min(find_best_MinERR(N,:));  %find smallest minERR value
find_bestRFvar = find(find_best_MinERR(N,:) == best_MinERR(N,:));
if length(find_bestRFvar) == 2
    best_RFvar(N,:) = find_bestRFvar;
        %selects RFvariable of lowest correlation
else if length(find_bestRFvar > 2)
    best_RFvar(N,1) = min(find_bestRFvar);
    for i=2:length(find_bestRFvar)
        tempCorr(:,i-1) = Corr(find_bestRFvar(1,1),find_bestRFvar(1,i));
            %locate RFvariable with lowest correlation to the selected
            %RFvariable of lowest duration
        best_RFvar(N,2) = find_bestRFvar(1,find(min(tempCorr))+1);
    end
    end
end
find_bestPri(N,:) = priority;

    clear('find0')
    clear('minERR')
    temptext = ['Discriminant analysis for ' num2str(priY(1,N)) ...
        ' prior probability of landslide triggering events.'];
    disp(temptext)
                                                                toc;

                            end

%% Save results
if isdir('Results') == 0
    mkdir('Results')
end
cd Results
Filename = ['Discriminant Analysis_' TMPAproduct '_' classType '_'...
    num2str(resolu_Pri*100) 'pst_pristart' ...
    num2str(priStart) 'new.mat'];
%save(Filename,'find_best_MinERR','best_MinERR','best_RFvar','find_bestPri'
)
cd ..


%% Find lowest miss class. error, without high predictor correlation

%Locate absolute minimum ERR and variables (to be used in final threshold)
```

```
final_MinERR = min(best_MinERR);
find_final = find(best_MinERR == final_MinERR);
final_RFvar = best_RFvar(find_final,:);
final_priority = find_bestPri(find_final,:);

DispFig = 'y';

DA_final_linear;
%%
%Plot distribution of ERR by prior probability
y = best_MinERR;
x = find_bestPri(:,2)

F = figure('Color','w','Position',[50 100 800 400]);
p = plot(x,y);

title('Distribution of misclassification error rates', ...
    'FontWeight','bold','FontSize',18)
ylabel('Error rates','FontSize',18)
xlabel('Prior probability of triggering events','FontSize',18)

box  on
set(p, 'LineStyle','--','LineWidth',2,'Color',[0 0 0])
set(gca, 'FontSize',10,'FontName','Kalinga')
```

26. DA_subplots.m

```
%%%%% DISCRIMINANT ANALYSIS %%%%%

    %give range from variables in ThresholdAnalysis.m
for I=Range_min:Range_max   %column in subplots
    for A=Range_min:Range_max   %line in subplots
        if Corr(I,A) >= 0.8   %A == I
            continue    %skipp DA for high correlation
        end
        if DispFig == 'y'
            plotpos(A,I) = I + (A-1)*13;    %gives position to subplots
            subplot(length(A_RF_d_noNaN(1,:)), ...
                length(A_RF_d_noNaN(1,:)), plotpos(A,I))
        end
        %prepare axis values
        temp_maxAxis = max([max(A_RF_d_noNaN(:,I)) ...
            max(A_RF_d_noNaN(:,A))]);          %find max axis value
        maxAxis = round((temp_maxAxis+10)/20)*20;   %round up to closest 20

        %discriminant classification Matlab-function
        [C,err,Posterior,logP,coeff] = classify( ...
            [A_RF_d_noNaN(:,I) A_RF_d_noNaN(:,A)], ...
            [A_RF_d_noNaN(:,I) A_RF_d_noNaN(:,A)], ...
            LS_outcome_noNaN(:,1), classType, priority);
                %err = missclassification error rate
                %coeff = coefficiants to discriminant function for each
                %        combination of variables

        %Plot rainfall data and discriminant function
    if DispFig == 'y'
            plotMarkers = 'x*';
            markerSize = [1 2.5];
```

```
        h = gscatter(A_RF_d_noNaN(:,A), A_RF_d_noNaN(:,I), ...
            LS_outcome_noNaN(:,1),'gr',plotMarkers,markerSize,'off',[],[]);
        box on
        hold on
        h1 = gscatter(X,Y,C,'kk','.',0.2,'off'); %add 1-to-1 line in figure
    end
        if strcmp(classType, 'linear') == 1
                %compare string with classType
            K = coeff(1,2).const;
            L = coeff(1,2).linear;
            f = sprintf('0 = %g+%g*x+%g*y',K,L); %linear threshold curve

            %check which combinations meets the criteria for a threshold
            condition = (L(1,1) < 0 && L(2,1) < 0 && K > 0)  ...
                || (L(1,1) > 0 && L(2,1) > 0 && K < 0);
            if condition
                %last part removed as no good threshold eq.'s are produced
                classK(A,I) = K;
                classL1(A,I) = L(1,1);
                classL2(A,I) = L(2,1);    %store coeff's for threshold-eq's
                ThresholdEq(A,I) = cellstr(f);
                classError(A,I) = err;            %classification error
            end

        end
        if strcmp(classType, 'quadratic') == 1
            K = coeff(1,2).const;
            L = coeff(1,2).linear;
            Q = coeff(1,2).quadratic;
            f = sprintf('0 = %g+%g*x+%g*y+%g*x^2+%g*x.*y+%g*y.^2',...
                K,L,Q(1,1),Q(1,2)+Q(2,1),Q(2,2));   %quadratic threshold

            if condition
                %last part removed as no good threshold eq.'s are produced
                classK(A,I) = K;
                classL1(A,I) = L(1,1);
                classL2(A,I) = L(2,1);    %store coeff's for threshold-eq's
                classQ11(A,I) = Q(1,1);
                classQ12(A,I) = Q(1,2);
                classQ21(A,I) = Q(2,1);
                classQ22(A,I) = Q(2,2);
                ThresholdEq(A,I) = cellstr(f);
                classError(A,I) = err;
            end
        end
        plotSize = [0 maxAxis 0 maxAxis];
    if DispFig == 'y'
        h2 = ezplot(f,plotSize);    %plots threshold line

    %Adjust figure
        axis(plotSize)  %set axis to closest 50 above max RF value
        axis square
        set(h2,'Color','blue','LineWidth',2)
        set(gca, 'xticklabel',[], 'yticklabel',[])

    %adjust figure for plots
            xlabel([])
            ylabel([])
            title([])
            %axis square
```

```
    end
            clear('X','Y', 'Fig')
        tempText = ['Discriminant analysis: Duration ' num2str(I) ...
            ' x ' num2str(A) ' done.'];
        disp(tempText)
    end
end
```

27. DA_final.m

```
%%%%% DISCRIMINANT ANALYSIS %%%%%
figure('color','w')
 I = final_RFvar(1,1);
 A = final_RFvar(1,2);
 priority = final_priority;

  %prepare axis values
temp_maxAxis = max([max(A_RF_d_noNaN(:,I)) ...
    max(A_RF_d_noNaN(:,A))]);          %find max axis value
maxAxis = round((temp_maxAxis+10)/20)*20;   %round up to closest 20

%Create X and Y sample values for classify-function
X = (1:length(A_RF_d_noNaN))';
Y = (1:length(A_RF_d_noNaN))';
X = X/length(A_RF_d_noNaN)*maxAxis;
Y = Y/length(A_RF_d_noNaN)*maxAxis;

%discriminant classification Matlab-function
[C,err,R,logP,coeff] = ...
            classify([A_RF_d_noNaN(:,I) A_RF_d_noNaN(:,A)], ...
            [A_RF_d_noNaN(:,I) A_RF_d_noNaN(:,A)], ...
            LS_outcome_noNaN(:,1), classType, priority);

%Plot rainfall data and discriminant function
h = gscatter(A_RF_d_noNaN(:,A), A_RF_d_noNaN(:,I), ...
    LS_outcome_noNaN(:,1),'gr','xo',[3 5],'off',[],[]);
box on
hold on
h1 = gscatter(X,Y,C,'kk','.',0.2,'off'); %add 1-to-1 line in figure

if strcmp(classType, 'linear') == 1
                %compare string with classType
   K2 = coeff(1,2).const;
   L2 = coeff(1,2).linear;
   f2 = sprintf('0 = %g+%g*x+%g*y',K2,L2);    %linear threshold curve
end
if strcmp(classType, 'quadratic') == 1
   K2 = coeff(1,2).const;
   L2 = coeff(1,2).linear;
   Q2 = coeff(1,2).quadratic;
   f2 = sprintf('0 = %g+%g*x+%g*y+%g*x^2+%g*x.*y+%g*y.^2',...
                K2,L2,Q2(1,1),Q2(1,2)+Q2(2,1),Q2(2,2));   %quadratic
threshold
end

plotSize = [0 maxAxis 0 maxAxis];
h2 = ezplot(f2,plotSize);    %plots threshold line
```

```
%Adjust figure
axis(plotSize)  %set axis to closest 50 above max RF value
axis square
set(h2,'Color','blue','LineWidth',2)
set(gca, 'xticklabel',[], 'yticklabel',[])

 %adjust figure for plots
 x_lab = ['A_' num2str(A) '_d (mm/d)'];
 xlabel(x_lab,'FontSize',18)
 y_lab = ['A_' num2str(I) '_d (mm/d)'];
 ylabel(y_lab,'FontSize',18)
 figTitle = [f2 ', ''y'' prior = ' num2str(final_priority(1,2))];
 title(figTitle,'FontSize',10) %'FontWeight','bold'
 hold off
 set(gca, 'XTick',0:20:maxAxis, 'YTick',0:20:maxAxis)

 set(gca, 'FontSize',10,'FontName','Kalinga')
 axis square

    clear('X','Y', 'Fig')
    tempText = ['Discriminant analysis: Duration ' num2str(I) ...
        ' x ' num2str(A) ' done.'];
    disp(tempText)

% Calculate new DA-equation with y on left hand side
Aj = L2(1,1)/L2(1,1);
Ai = L2(2,1)/-L2(1,1); %equal to x/y or Ai/Aj
Alpha = K2/-L2(1,1);
new_eq = ['A_' num2str(I) '_d = ' num2str(Alpha) ' + ' ...
    num2str(Ai) 'A_' num2str(A) '_d'];
```