

# **INVESTIGATING HURRICANE DRIVEN LANDSLIDES: FROM PHYSICALLY BASED TO STATISTICALLY BASED AND FROM SPACE TO SPACE-TIME**

DAYALLINI SENTHIL NATHAN

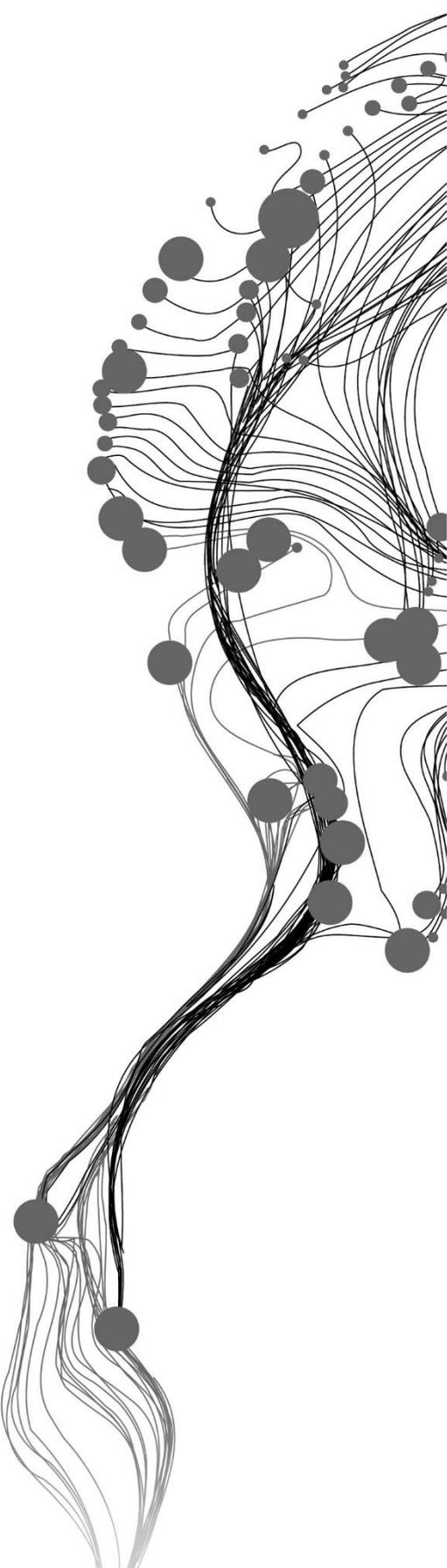
July 2020

SUPERVISORS:

Dr. L. Lombardo

Dr. C.J. van Westen





# **INVESTIGATING HURRICANE DRIVEN LANDSLIDES: FROM PHYSICALLY BASED TO STATISTICALLY BASED AND FROM SPACE TO SPACE-TIME**

**DAYALLINI SENTHIL NATHAN**

Enschede, The Netherlands.

July 2020

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.

Specialization: Natural Hazards and Disaster Risk Reduction

## **SUPERVISORS:**

Dr. L. Lombardo (Department of Earth Systems Analysis (ITC ESA))

Dr. C.J. van Westen (Department of Earth Systems Analysis (ITC ESA))

## **THESIS ASSESSMENT BOARD:**

Dr. V.G. Jetten (Department of Earth Systems Analysis (ITC ESA))

Dr. Rens van Beek, University of Utrecht

#### DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

## ABSTRACT

Hazards are generally defined by three components: Where they occurred? When have they occurred? And How destructive were they? This study focuses on these components for the event of landslides. Initially, where the landslides have occurred is analysed, this concept is commonly termed as landslide susceptibility. Over the years, there had been several techniques via which the susceptibility is estimated. This study specifically researches the quantitative methods namely: statistical and physically-based models. While the physically-based is process driven, the statistical model is data driven. Statistical framework aids in a physical model for the interpolation and parameterisation of the physical parameters but an incorporation of these physical parameters into a statistical model is hardly ever done. Thus, this study aims to visualise the difference in the spatial patterns obtained from statistical analysis done on the different parametric datasets: i) the traditional parameters used in statistical framework for landslide susceptibility; ii) physical parameters which explains the slope instabilities and iii) combination of both in the above-mentioned. In this study the physically-based model is not carried out rather the inputs and the outputs of a previously executed model is utilised. Thereby with this regard, this initial study is done on the region of Grand Bay at Dominica. For this a Generalised Linear Model (GLM) with binomial probability distribution integrated with the Least Absolute Shrinkage and Selection Operator (LASSO) as variable selector is implemented. The traditional parameters were more adept in capturing the spatial characteristics of the landslide susceptibility, this was because of the increased spatial variability of the conditioning factors. The next phase of the study focuses on “When” the landslides have occurred. An attempt to examine whether a statistical framework is capable of recognising a spatial pattern of the temporal dependency on landslide occurrences, is undertaken. In order to scrutinise this, a generalised additive model (GAM) with its temporal counterpart and considering non-linear parameters is executed. Specifically, an autoregressive model acts upon this GAM in order to speculate on capturing a temporal latency effect on the landslide susceptibility. While the previous study was done at pixel level, this was carried out at slope unit level for the whole island of Dominica and the available five landslide inventories for the region was utilised. For this tropical region, no significant temporal latency on susceptibility was observed. This might be due to the fact that there is a spatial variability of the triggering events over the period of years, thus the model is dominated by the spatial trends rather than temporal ones. The final phase of this study was on “How”, and, explicitly approached to model a specific characteristic of the landslide, the percentage of landslide area per mapping unit. A GLM with Gaussian probability distribution was implemented for the whole island for the five time periods. These models were also executed at the slope unit level and though they reflected on the increase/ decrease of the percentages adequately, they were unable to efficiently capture the variance. This was due to the data inadequacy in terms of sample size (a small dimension of input dataset) and spatial invariability (a uniformity in the characteristics of the covariates used). While this study focuses on the components individually, a more established future research would be on ways to integrate all the three components and facilitate the learning ability of the statistical framework which in turn would increase the performance of the susceptibility model.

**Keywords:** Landslide Susceptibility, Generalised Linear Model, Generalised Additive Model, Binary Logistic Regression, Physically-based model, Geotechnical and Hydrological parameters, Landslide Area Percentage, Dominica

## ACKNOWLEDGEMENTS

I would first like to express my gratefulness to my supervisors, Dr. Luigi Lombardo and Dr. Cees van Westen. Whenever I ran into a trouble spot or had doubts regarding my research or writing, their invaluable insight, exemplary guidance, encouragement, and patience steered me in the right direction.

Further, I am thankful to Bastian who shared his insights and am indebted for his valuable comments. I am grateful to all the professors and technical staffs who helped me in gaining the knowledge which allowed me to complete the study.

My sincere gratitude towards ITC for giving me this amazing opportunity to study in the Netherlands and thank the God for this blessing. The university exposed me to the international culture, and I am thankful for the support and help I had received from my friends during my study.

Finally, I must express my very profound gratitude to my family for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Dayallini Senthil Nathan  
Enschede, July 2020

# TABLE OF CONTENTS

---

1.	Introduction.....	8
1.1.	Rationale of this Study.....	9
1.2.	Research Questions and Objectives.....	10
1.3.	Structure of this thesis .....	11
2.	Study Area.....	12
2.1.	Climatological Context of the Region.....	12
2.2.	Geological and Pedological Context of the Region .....	14
2.3.	Landslide inventory of the Study Area .....	15
3.	Methodology.....	17
3.1.	Landslide Conditioning Factors for the various Models of this Study .....	21
4.	Results and Discussions.....	28
4.1.	Importance of the covariates among morphometric, thematic, and physical parameters for the landslide susceptibility modelling .....	28
4.2.	Attempting to capture temporal effects in the landslide susceptibility. ....	35
4.3.	Spatial patterns of the percentage of landslide area per mapping unit over the years .....	41
5.	Conclusions and Recommendations .....	46

## LIST OF FIGURES

---

Figure 2.1 The relief map of the island (Central Intelligence Agency, 1990); the region of Grand Bay has been highlighted which would be the area of focus for first phase of the study.....	12
Figure 2.2 Geological map of Dominica (Roobol and Smith, 2004).....	14
Figure 2.3 Landslide occurrences of various years.....	16
Figure 3.1 The Methodology Flowchart for First Phase of the Study.....	19
Figure 3.2 The Methodology Flowchart for the Second Phase to investigate the Temporal Dependency on Landslide Susceptibility.....	20
Figure 3.3 The Methodology Flowchart for the Final Phase of the Study where an Initial analysis is done on the Spatial Patterns of the Landslide Run-out Areas.....	21
Figure 4.1 The results of the variable selector in the models with Dataset A, Dataset B and Dataset C.....	28
Figure 4.2 The Regression Coefficients of the continuous covariates of Dataset A and Dataset C.....	30
Figure 4.3 The Regression Coefficients of the covariates of Dataset B and Dataset C.....	31
Figure 4.4 Landslide Susceptibility obtained from the three Models.....	33
Figure 4.5 ROC curves of the Models.....	33
Figure 4.6 Scatterplot between the Predicted Values of the Model with dataset C and those of the Model with dataset A and B.....	34
Figure 4.7 The Regression Coefficients of the Significant Covariates in the Temporal Dependency Model.....	35
Figure 4.8 The Regression Coefficient of the Classes of the Slope Steepness.....	36
Figure 4.9 The Regression Coefficient of the Classes of the Aspect.....	36
Figure 4.10 Landslide Susceptibility and its Uncertainty Maps of the year 1987.....	37
Figure 4.11 Landslide Susceptibility and its Uncertainty Maps of the year 1990.....	38
Figure 4.12 Landslide Susceptibility and its Uncertainty Maps of the year 2014.....	38
Figure 4.13 Landslide Susceptibility and its Uncertainty Maps of the year 2015.....	39
Figure 4.14 Landslide Susceptibility and its Uncertainty Maps of the year 2017.....	39
Figure 4.15 Mean and Maximum of Landslide Susceptibility over the years.....	40
Figure 4.16 The Temporal Dependencies generated for the Various Slope Units.....	40
Figure 4.17 The Regression Coefficients of the Covariates (excl. geology and soil type) in Models for the Landslide Area Percentages of the years.....	41
Figure 4.18 The Actual and the Predicted Percentages of the Landslide Areas of 1987.....	42
Figure 4.19 The Actual and the Predicted Percentages of the Landslide Areas of 1990.....	43
Figure 4.20 The Actual and the Predicted Percentages of the Landslide Areas of 2014.....	43
Figure 4.21 The Actual and the Predicted Percentages of the Landslide Areas of 2015.....	44
Figure 4.22 The Actual and the Predicted Percentages of the Landslide Areas of 2017.....	44
Figure 4.23 The Scatterplots between the Actual and the Predicted Percentage of Landslide Area per Slope Unit.....	45

## LIST OF TABLES

---

Table 2.1 Characteristics of the hurricanes and tropical storms that had triggered landslides in the island (The date denotes the day it hit the island).....	13
Table 2.2 Summary of the available landslide inventories for Dominica.....	15
Table 3.1 Characteristics of the Conditioning Factors (Median for continuous data with unique values and Mean for the continuous data is tabulated).....	25
Table 4.1 Regression Coefficients of the conditioning factors of Dataset A and Dataset C.....	29
Table 4.2 Regression Coefficients of the conditioning factors of Dataset B and Dataset C.....	29

# 1. INTRODUCTION

For at least three decades, predicting where landslides would occur has been a notion known as landslide susceptibility whereas the temporal and magnitude components of widespread landslide events has often been included in the landslide hazard definition (Brabb 1985; Hansen 1984; Varnes DJ 1984). Several approaches have been proposed through the years for mapping landslide susceptibility, starting from geomorphological mapping (Hansen et al., 1995), investigation of landslide inventories (Campbell, 1973; DeGraff, 1985; Galli et al., 2008), zonation of susceptibility by terrain analysis (Nilsen and Brabb, 1977; Abella and van Westen, 2008), statistically-based methods (Carrara, 1983; Chacón et al., 2006) and physically-based numerical models (Montgomery and Dietrich, 1994; Rigon et al., 2006; Simoni et al., 2008). These approaches all share some common requirement:

*i)* the partition of a study area in a specific mapping unit. A mapping unit corresponds to the geographic entity used to subdivide the whole area and to ultimately assign the outcome of the susceptibility/hazard models (Reichenbach et al., 2018). Among the available mapping units, one can commonly find the use of grid-cells (regular squared lattices), slope units (irregular moderate-scale polygons encompassing ridges to the closest streamline), catchments (irregular small-scale polygons from the ridges to the outlets), administrative units (irregular polygons corresponding to counties, cities or provinces).

*ii)* the information on the distribution of past landslides. For instance, geomorphological mapping (Reichenbach et al., 2005), which essentially corresponds to the ability of a geomorphologist to interpret the landscape and recognize landslide-prone slopes, relies on the understanding of a geomorphologist due to his/her past field experience. As for statistically based methods, these learn from the functional relations existing between past landslides and associated landscape properties (Guzzetti et al., 1999; van Westen et al., 2008). Physically-based models implement the hydro-mechanical laws which govern the instability process, and the optimal parameterization is calibrated to maximize the numerical- to the real – landslide scenario (Baum et al., 2008; Anagnostopoulos and Burlando, 2012; Bout et al., 2018).

Despite these commonalities, substantial differences also exist in the way the various approaches are implemented and the data they require (Glade and Crozier, 2005). For instance, physically-based models require the actual geotechnical and hydrological properties featured in the governing equations (Terlien et al., 1995; Borga et al., 1998). Conversely, statistically based models usually feature proxies (Guzzetti et al., 2005; Goetz et al., 2015) of the geotechnical and hydrological parameters mentioned above. This context implies a similar predictive outcome when it comes to interpretation. But it also implies that significant differences exist in the respective predictive capacity. This study initially focuses on the different level of information carried by the parameters mentioned above, these being tested in the context of statistically based susceptibility models. Specifically, checking the differences in performance and interpretability among: *i)* a model featuring terrain as well as lithological and pedological properties; *ii)* a model featuring geotechnical and hydrological parameters only; *iii)* a model featuring the combination of both.

In addition to this, the susceptibility literature has seen the effect of pre-existing slope instabilities to be relevant if not dominant at times, with respect to subsequent landslide events. This effect has been defined as landslide path dependency (Samia et al., 2017) and it has been demonstrated that its inclusion improved the prediction in statistically based models. In this work, it has also been tested whether the landslide path-dependency can be captured via statistical models in tropical areas.

And, referring to the differences between landslide susceptibility and hazard, one of the differences is due to the inclusion of the landslide event-magnitude. This concept describes the magnitude (a proxy for the destructiveness) of a population of landslides. This is traditionally included in landslide hazard models in terms of landslide size (area) characteristics (Malamud et al., 2004). However, when using a slope-unit partition (Carrara et al., 1995), one could express the magnitude of a population of landslide as the proportion of failed slope-units. This metric can indicate how much of a given slope unit has failed in response to single or multiple failures occurred inside the same slope unit polygonal extent. Here, a trial has been undergone to complement the susceptibility information together with the hazard information expressed as the percentage of the failed slopes, testing the framework mentioned above.

### **1.1. Rationale of this Study**

In this study, the major focus is on the quantitative methods, as described above these could be grouped as either physically-based approach or statistical approach. While physically-based models are process-driven the statistical models are data-driven (Canli et al., 2015). The physically-based model simulates the physics behind the processes thus making it directly interpretable because they feature physical quantities that can be generally measured in laboratories or via field tests (Wu and Sidle, 1995; Anagnostopoulos et al., 2015; Alvioli and Baum, 2016). However, these models are dependent on how the sampling is done to derive the above data. Statistical methods rely on certain properties of the landscape that are available at high resolution to be integrated into the model (Carrara, 1983; Chacón et al., 2006; van Westen et al., 2008) but the problem arises as these parameters in reality are not the contributing factors for the landslides but rather they are proxies for the driving characteristics. For example, geology type, which is one of the common parameters used in statistical frameworks, isn't a conditional factor for landslide occurrences as such but instead it is the shear strength and the bulk density of the lithology that makes a place more susceptible to landslides. Thus, in a statistically based model even though the input data is more accessible, due to the non-usage of the properties that govern the physics, the interpretability of the process from the model output often cumbersome or less straightforward than the physically-based counterpart (van Westen et al., 2006).

There have been numerous studies focused on understanding the differences in performances of the statistically based and the physically-based models (Guzzetti 2006; Yilmaz and Keskin, 2009; Canli et al., 2015). However, there are very few studies which have incorporated physical parameters into a statistical framework (Goetz et al., 2011; Pradhan et al., 2019). For instance, these previous studies had included the factor of safety distribution map to model the landslides. But an inclusion of other geotechnical and hydrological parameters into a statistical framework for landslide susceptibility modelling is yet to be done. The initial phase of this study aids in comprehending what are the important process driven and data driven parameters influencing landslide susceptibility prediction and the difference in the spatial trends that can be captured by them.

After this, delving into a broader perspective of the assumption that the parameters influencing the past and present landslides are most likely to influence the future occurrences (Furlani and Ninfo, 2015), an analysis on the temporal counterpart of landslide susceptibility is likely to give an understanding of the propagation of landslide occurrences. The ability of a statistical model to capture this temporal dependency was studied by Samia et al., (2018). It was observed that the previously existing landslides had an effect onto the subsequent landslide occurrences via a statistical analysis. An implementation of a statistical framework which is endowed with the capability for identifying the temporal effect for a region of high vegetation is yet to be done. This phase of the study would examine the capability of a statistical model to recognize the temporal dependence on the landslide susceptibility, if present, in the context of a tropical environment.

While the above two study phases concentrates on the landslide initiation, the last phase of this study extend the analyses by considering the extent of runout area of a landslide (or a combination of

landslides) per mapping unit. Certain studies have been carried out in the past where a statistical analysis is done on the evaluation of the landslide hazard based on the runout (Carrara et al., 1991) and analysis on the landslide occurrences and their runout distance (Devoli et al., 2009). This study specifically examines a statistical framework considering the percentage of landslide area per mapping unit, expecting to aid in understanding the extent of area that could be affected by a landslide.

## 1.2. Research Questions and Objectives

1. What are the differences among predictive spatial patterns provided by a purely physically-based model, a traditional statistical model with morphometric and thematic properties, and a hybrid model with a statistical algorithmic architecture but with a combination of the morphometric and thematic parameters together with predisposing factors obtained for/via physical simulations?

Obj. To test whether physically-based or statistical models or a hybrid between the two provide enough information to assess the landslide susceptibility in hurricane-driven disasters.

- i)* What are the available data inputs with regards to morphometric and thematic properties?
  - ii)* What were the parameters that OpenLISEM had considered for the simulation of slope instability?
  - iii)* How does the inclusion of the various parameters affect the susceptibility model?
  - iv)* What is the performance of each model? Is there a difference in their capability to estimate landslide susceptibility?
  - v)* Why does a difference occur and what are the aspects that models lack to capture in comparison to each other?
2. Is there an effect from the past to the future landslide occurrences on predictive power of landslide susceptibility model for a tropical region? Do landslides occur through time in the same location?

Obj. To implement an autoregressive model acting on a generalized additive model with its temporal component

- i)* Can a statistical framework identify the temporal dependency in densely forested areas where the regrowth of the vegetation is quite rapid?
3. What are the spatial characteristics of the landslide areas?

Obj. To extend a statistical framework to accommodate the prediction of percentage of landslide area per mapping unit

- i)* How and why are the influence of the conditioning factors different for the initiation and the depositional regions of a landslide?
- ii)* How can the landslide area model be integrated to landslide susceptibility models?

In order to answer the above-mentioned research questions, the most common implementation of statistical framework for landslide susceptibility, Generalized Linear Model (hereafter GLM; Nelder and Wedderburn, 1972; Atkinson et al., 1998; Reichenbach et al., 2018) has been tested in three separate experiments by extending the GLM framework in three different ways in this study. The GLM is essentially a multivariate statistical model that assumes that the effect of each covariate affecting the landslide susceptibility is linear. The first extension to the simpler GLM has been implemented by attaching to a GLM kernel a powerful variable selection tool called Least Absolute Shrinkage and Selection Operator (hereafter LASSO; Camilo et al., 2017; Lombardo et al., 2018). In the past years, for the implementation of variable selector in landslide susceptibility models a predominant variable selection routine corresponding to stepwise algorithm has been used. This stepwise algorithm has been demonstrated to be flawed and

extremely conservative (Harrell, 2015) whereas LASSO is a more reliable selector (Amato et al., 2019). The second experiment was conducted by extending the GLM framework to its more flexible counterpart, the Generalized Additive Model (hereafter GAM, Hastie and Tibshirani 1987; Goetz et al., 2011) which is able to account for non-linear effects, for instance, the use of ordinal properties, and the use of tools that would treat differently mapping units close and far in space or time. The third experiment featured a GLM used for modelling the extent to which a landslide would affect a given mapping unit, thus rather than the binary presence/absence landslide scenario (the typical input for a landslide susceptibility model), it attempts to model a continuous property. A description of how these were implemented to what study area and what were the outcomes is further detailed in this thesis.

### **1.3. Structure of this thesis**

Chapter 1 describes the background information, motivation and basic methodology and summarizes the main research questions and objectives of this study. The chapter 2 of the thesis describes the various geographical and topographical aspects of the study area. The detailed methodology for the study and the conditioning factors utilized in this study are presented in chapter 3. The results and discussions are summarized in chapter 4. Chapter 5 contains the fundamental conclusions and identifies both the limitations to the study as well as recommendations for future research.

## 2. STUDY AREA

Dominica is one of the Caribbean islands which lies between the French islands of Guadeloupe and Marie-Galante in the north and Martinique in the south (Figure 2.1). It is about 750 square kilometers in area with 335 streams and rivers, one of the largest boiling lakes in the world and nine active volcano spots (Lindsay et al., 2005). The island's capital is Roseau and is accessible via airways and waterways. Dominica has a population of 71625 (as of 2018 census) and its economy is predominantly reliant on agriculture. Though the high variations in the weather affects the crop production. The island lies in the hurricane region near the equator, making it prone to face various hurricanes. Over the past years, these hurricanes had triggered landslides (among other disasters) over the region ("Dominica | CHARIM," n.d.).

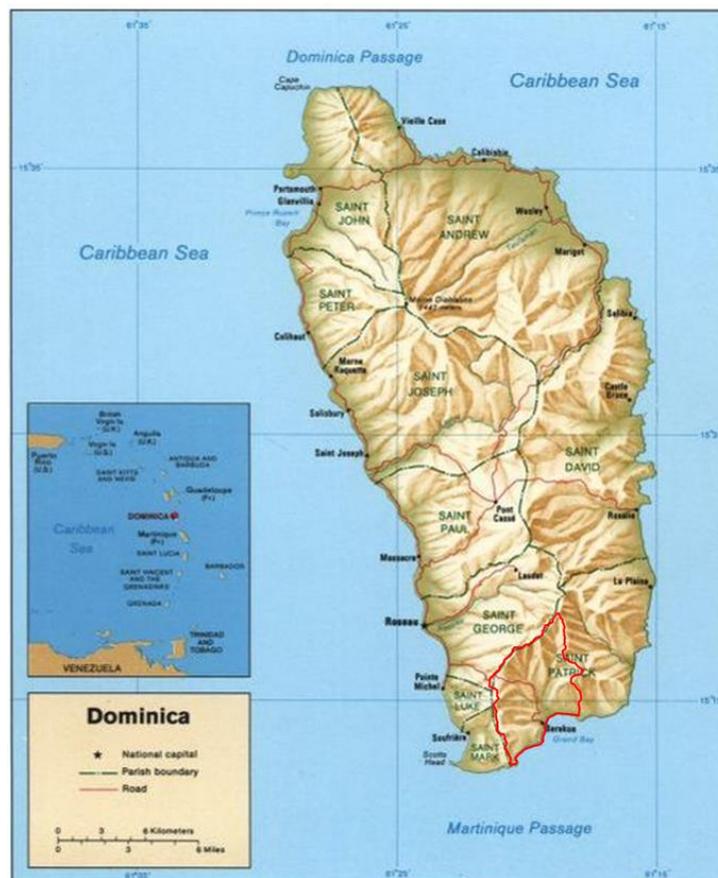


Figure 2.1 The relief map of the island (Central Intelligence Agency, 1990); the region of Grand Bay has been highlighted which would be the area of focus for first phase of the study.

### 2.1. Climatological Context of the Region

The island usually experiences a marine tropical climate with minimal seasonal variation. But the rainfall patterns have been varying due to the various hurricanes that have hit the region. Over the years, the island has been affected by tropical storms or hurricanes once in 2.5 years. Direct hurricane hits were occurring once in 9.87 years and the major hurricanes once in 16.44 years and these hurricanes were observed to have an average wind speed of 114mph.

Table 2.1 summarizes the various hurricanes and tropical storms that hit the island.

Table 2.1 Characteristics of the hurricanes and tropical storms that had triggered landslides in the island (The date denotes the day it hit the island)

<b>Year</b>	<b>Date</b>	<b>Event</b>	<b>Characteristic</b>
1806	Sep 20th	Hurricane	Triggered landslide and flooding in the capital, Roseau killing 131 people
1834	Sep 9th	Hurricane	A severe event triggering landslides (Category 4 or 5)
1916	Aug 28th	Hurricane	A strengthening tropical storm becomes a hurricane 85mph from the east
1926	Jul 24th	Hurricane	Triggered landslides and led to damage of roadways, electric and phone lines
1930	Sep 1st	Hurricane	90 mph winds hit the area from the east
1963	Sep 28th	Hurricane Edith	Winds of 80 mph triggering landslides and floods damaging the vegetation
1970	Aug 20th	Hurricane Dorothy	Mostly wind damage and affected north and east regions the worst.
1979	Aug 29th	Hurricane David	150 mph wind speed lasting 6 hours (Category 5)
1979	Sep 1st	Hurricane Frederick	Extended damage from the previous hurricane David
1980		Hurricane Allen	Category 1 hurricane which could have trigger landslides
1984	Nov 6th	Hurricane Klaus	Landslide at Bellevue Chopin
1987		Hurricane Emily	Triggered landslides
1988		Hurricane Gilbert	Landslides Matthieu and Layou River
1989	Sep 17th	Hurricane Hugo	Severe economic damage
1994	Sep 10th	Hurricane Debbie	Damages to the croplands
1995	Aug 27th	Hurricane Iris	Large landslides near Mathieu River.
1995	Sep 4th	Hurricane Marilyn	Hits from the SE with 80mph winds area reports sustained winds of 72mph for 10 minutes
1995	Sep 18th	Hurricane Luis	Large landslides near dam lake Mathieu.
1999	Nov 18th	Hurricane Lenny	Landslides in the north damaging the coastal infrastructures
2007	Aug 21st	Hurricane Dean	Major debris flow from the Soufriere Sulphur Springs' Upper Fumarole Area on the night of Hurricane Dean
2009	Sep 4th	Hurricane Erica	27 landslides along the roads
2010	Oct 31st	Hurricane Tomas	20 landslides along the roads
2011	Sep 28th	Hurricane Ophelia	84 landslides along the roads
2015	Aug 27th	Tropical storm Erika	The majority of damages were sustained in the transport sector (60 percent), followed by the housing sector (11 percent) and agriculture sector (10 percent). Approximately 7,229 impacted by the event in disaster declared areas

2017	Sep 18th	Hurricane Maria	Hits directly with 165 mph winds; 922mm of rain causing very heavy damage. According to media reports, the estimated damage total in Dominica is at least \$1.31 billion. The agricultural sector was essentially eliminated. The once-lush tropical island was effectively reduced to an immense field of debris. There was extensive damage to roads. Power, phone, and internet service were cut off, leaving the country almost incommunicado with the outside world. Wind speed of 133mph at Canefield Airport and 150mph at Douglas-Charles Airport.
------	----------	-----------------	--

## 2.2. Geological and Pedological Context of the Region

Dominica is a volcanic island with the oldest bedrocks of basalt deposits from volcanoes in the Miocene era. The central and southern island is mostly made up of younger Pleistocene deposits composed of ignimbrite and ash while the older deposits are in the eastern regions (Figure 2.2; Roobol and Smith, 2004). The tropical climate exposes the bedrock to deep weathering (DeGraff, 1991) and this leads to formation of the weathered volcanic soil. This soil weaker than the bedrock causes a loss in the soil shear strength with the exposure to high precipitation generating zones of slope instabilities (Walsh, 1982).

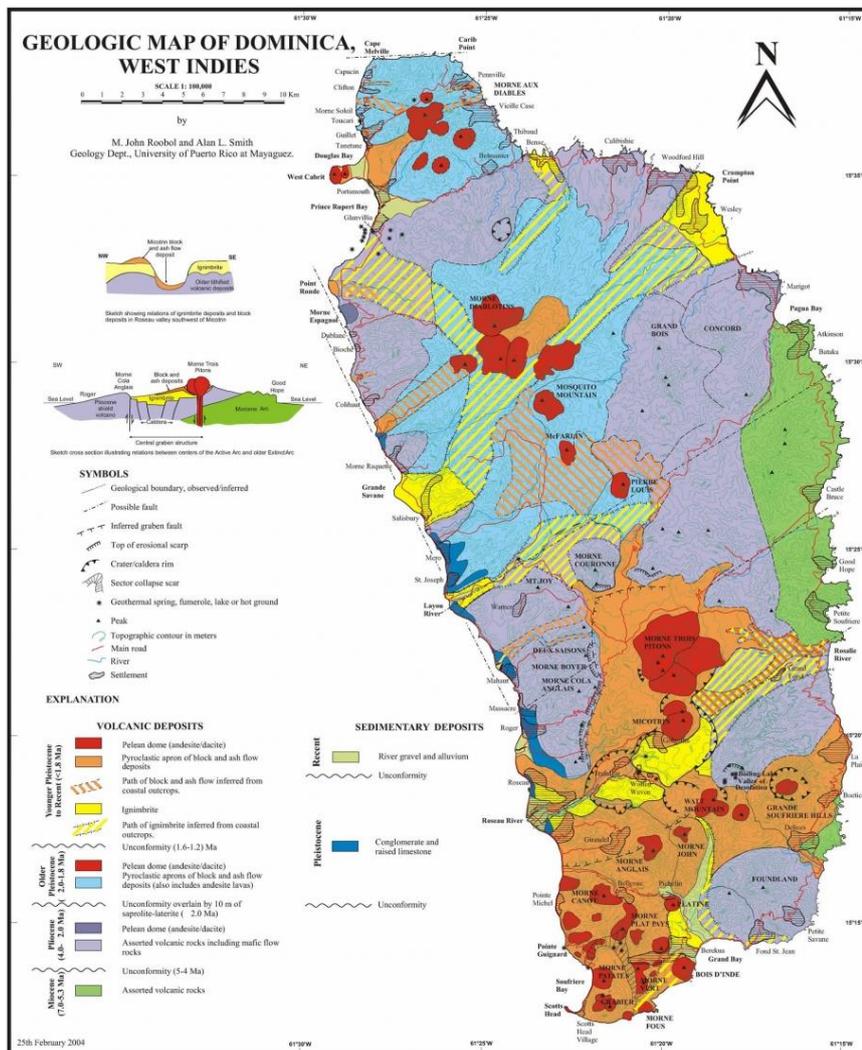


Figure 2.2 Geological map of Dominica (Roobol and Smith, 2004)

The island of Dominica majorly has four noticeable soils: allophanoid latosolics, allophanoid podzolics, kandoid soils and smectoid soils. While the allophanoid soils are formed in regions of high annual rainfall, the kandoid soils are formed in moderate rainfall and the smectoid are formed in the regions of low rainfall. The region due to the active volcano spots has fertile volcanic soil thereby increasing the growth of vegetation. Also, the clay soils of the region are high in porosity and can affect the groundwater flow and runoff (Rouse et al. 1986). The central island is made of allophanoid latosolics which is known to have a high permeability and low bulk density. These soils are generally stable unless there is a slopecut (Lang, 1967). Other prevalent soils in the area is the kandoid latosolics in the north-east and the young soils in the south west and west.

### 2.3. Landslide inventory of the Study Area

Throughout the years, there are five landslide inventories mapped for the whole region of Dominica. The landslides inventory made by DeGraff (1987) recorded 896 landslides which were probably caused by 1979 David, 1980 Allen and 1984 Klaus. The 1990 inventory by DeGraff mapped 187 landslides with possible triggering events: 1987 Emily, 1988 Gilbert and 1989 Hugo. The 1987 inventory was derived from the interpretation of the aerial photos from 1984 and the field investigation in 1986-Jan 1987. A digital map was available for this inventory but for the 1990 landslide inventory it was later digitized (van Westen, 2016). Both the inventories were mapped on a scale of 1:25000, while the 1987 map was classified the 1990 was not. The 2014 inventory by van Westen was generated by collection all available landslide reports but due to the low data reliability and availability, an inventory was mapped via pre- and post-event satellite image interpretation. This inventory was verified, and it reports 864 landslides classified based on the type. UNOSAT utilizing the image analysis of satellite imagery detected the presence of landslide and had reported 697 landslides caused by 2015 Erika (UNITAR-UNOSAT, 2015). However, the GIS data analysis shows 1554 landslide polygon, the 697 reported must have been generated automatically which probably merged neighbouring landslide areas (van Westen, 2016). This inventory did not have classification and was not verified. The 2017 inventory by van Westen mapped 10145 landslides triggered by 2017 Maria. This inventory had a detailed classification of the landslide type and had also mapped the parts of the landslide. This was the only inventory which had a proper scarp delineation. The Table 2.2 summarizes the five landslide inventories and the Figure 2.3 shows the maps of the landslide occurrences.

Table 2.2 Summary of the available landslide inventories for Dominica

<b>Year</b>	<b>1987</b>	<b>1990</b>	<b>2014</b>	<b>2015</b>	<b>2017</b>
<b>Inventory Characteristics</b>					
<b>Author</b>	DeGraff	DeGraff	van Westen	UNOSAT	van Westen
<b>Number of Landslides</b>	896	187	864	1554	10145
<b>Possible Triggering Events</b>	1979 David 1980 Allen 1984 Klaus	1987 Emily 1988 Gilbert 1989 Hugo	2007 Dean 2009 Erica 2010 Tomas 2011 Ophelia Heavy Rains (2011 and 2013)	2015 Erika	2017 Maria
<b>Classification</b>	Type Classification	No Classification	Type Classification	No Classification	Type and Part Classification

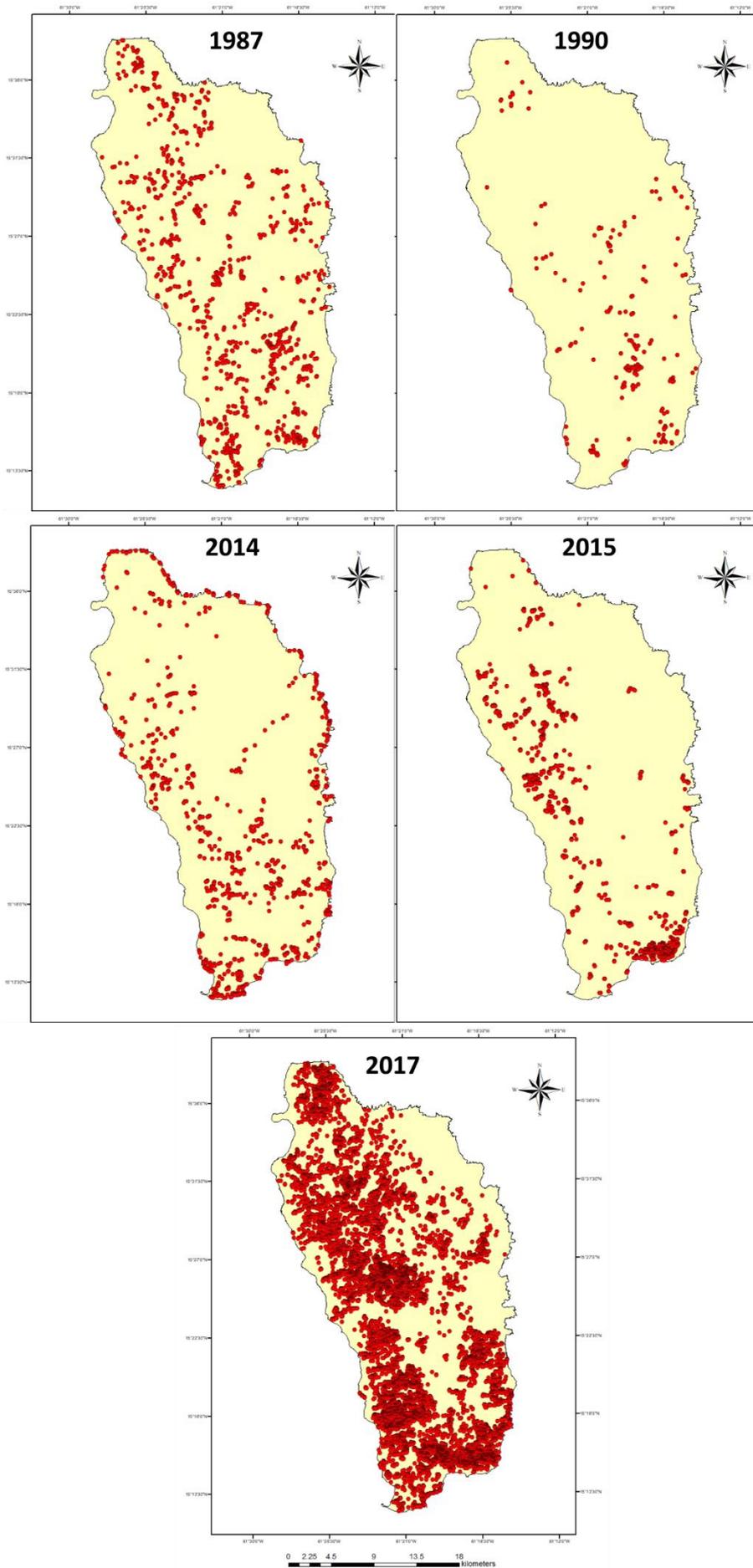


Figure 2.3 Landslide occurrences of various years

### 3. METHODOLOGY

The present work features a series of different models applied for various purposes when predicting landslide characteristics over space and time in Dominica. Every analytical step has been implemented in the R programming language which is widely used for statistical computations (Team R, 2014).

For an initial comparison of landslide susceptibility models built on the basis of terrain attributes and/or properties derived for and from physically-based models, a frequentist version of a binomial Generalized Linear Model framework (Atkinson et al., 1998; Brenning, 2005; Pourghasemi and Rahmati, 2018), implemented together with a powerful variable selection tool namely, Least Absolute Shrinkage And Selection Operator (hereafter LASSO; Camilo et al., 2017; Lombardo et al., 2017) has been used.

A Generalized Linear Model (hereafter GLM) is a well-established multivariate statistical model that can handle various exponential probability distributions (e.g., Gaussian, Bernoulli, Poisson, Gamma, etc.). Among these, the Bernoulli case is typically used whenever the target variable takes on two possible outcomes, which in landslide studies corresponds to the presence and absence of landslide occurrences over space and time (Chung et al., 1995; Stark and Guzzetti, 2009; Lombardo et al., 2018). A binomial GLM is also often referred to as Binary Logistic Regression (Hosmer and Lemeshow, 2000; Lombardo and Mai, 2018) and its formulation can be expressed as follows:

$$\eta P = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (1)$$

where,  $\eta$  is the logit link,  $P$  is the probability of landslide occurrence,  $\beta_0$  is the global intercept, and  $\beta_n X_n$  is the product between covariates and the corresponding estimated regression coefficients.

When the target variable and the outcome of a statistical model are expressed in the same scale, no transformation is required. However, for a binomial GLM the input is typically a series of binary instances whereas the output consists of a continuous probability spectrum. Because of this, a function that links the two input and output terms is required. Specifically, the logit link is used to move from odd-ratio to the probability scale.

From the previous equation one can derive the probability, i.e., the landslide susceptibility, by applying the following transformation:

$$P = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n}} \quad (2)$$

This model can feature both continuous and categorical properties but does not allow for more complex non-linearities. This GLM framework is the core of the analyses carried out in this study. To test the contribution of covariates to the susceptibility model, coming from different sources, a variable selection step has been added. Any frequentist GLM features a Maximum Likelihood step which is used to converge to the optimal parameter set and to enable subsequent statistical inference. The likelihood is given by:

$$l(\beta_0, \beta_1, \dots, \beta_n) = \sum_{i:y_i=1} \log \{(\pi(x_i))\} + \sum_{i:y_i=0} \log \{1 - (\pi(x_i))\} \quad (3)$$

where,  $x_i$  and  $\pi(x_i)$  represents the covariate value and the probability that the observation corresponds to a landslide presence in the  $i^{\text{th}}$  mapping unit. This model when utilized in the context of having a high number of covariates may exhibit complexity limiting the interpretability (Tibshirani, 1996). In order to overcome this, penalized logistic models are to be used and thus the integration of LASSO operator, which essentially penalizes the number of covariates in a model while assessing the predictive power at various covariates combinations (Camilo et al., 2017). LASSO penalizes the likelihood using the equation:

$$l^* = l - \lambda H \quad (4)$$

where,  $l$  is the likelihood that is being penalized by the inclusion of the terms  $H$  and  $\lambda$ .  $H$  controls parameter estimation while the  $\lambda$  acts to make sure that the likelihood and penalty are balanced. As  $\lambda$  increases, the domain of the regression coefficient is shrunk towards 0, therefore reducing the parameter space when the shrinkage reaches the zero value. The LASSO implementation in the glmnet R-package (Friedman et al., 2009; Team R, 2014), performs a 10-fold cross-validation step for each  $\lambda$  value. More specifically, for a given  $\lambda$ , the routine involves a random sampling scheme where 90% of the data is used for calibration and the complementary 10% is used for validation, thus providing a full description of the performance variability as the penalization increases.

Depending on the exponential family (Bernoulli/binomial, Poisson, Gamma, Gaussian, etc.) one chooses, the performance metrics for the cross-validation scheme changes. For instance, the simplest case for continuous properties is a Gaussian model, for which the performance is estimated via RMSE between observed and predicted estimates. As for the susceptibility context, where the model assumes that landslides are distributed over space according to a Bernoulli probability distribution, the most common metric consist of the Area Under the Curve (hereafter AUC), where the curve is a Receiver Operating Characteristic (hereafter ROC) one (Hanley and McNeil, 1982; Hosmer and Lemeshow, 2000; Gorsevski et al., 2006; Fagerland and Hosmer, 2012; Goetz et al., 2015). So, the same structure is maintained in the usage of LASSO in this study and the AUC over the 10% of validation data, ten times for each  $\lambda$  is measured. This means that for an array of 100  $\lambda$  values, a cross validation scheme is built featuring a total of 1000 replicates and thus informs of the mean behaviour and the associated uncertainty in model performances as the number of covariates decreases because of the penalty. Ultimately, one can choose the subset of the original covariates that offers the best performance with the minimum number of covariates. Here, this statistical framework is applied to three different scenarios: *i*) model considering the morphometric and thematic parameters; *ii*) model considering the physical parameters; *iii*) model that considers a combination of the two parametric datasets. Despite the relative complexity of this procedure, the computational times are still in the order of seconds to minutes for a matrix with approximately 500,000 elements (or pixels) and tens of covariates. In this study, the considered matrix consists of 344,973 pixels with 42 covariates for the combined model which considers both physically-based and the statistically based model parameters. The Figure 3.1 depicts the methodological flowchart for the first phase of the study.

However, two weaknesses affect this procedure. Firstly, a GLM is a linear model, therefore, if the functional relation between the susceptibility and any of the covariates does not respect the linearity assumption, the model will inevitably suffer or misrepresent these relations. Secondly, no matter the cross-validation scheme one chooses, the uncertainty is estimated by using a slightly varied version of the original dataset. Therefore, the uncertainty estimation is not an integral part of the model when it comes to produce susceptibility maps. However, a different modelling framework exist where both these weaknesses can be accounted for. For instance, an extension to the GLM framework is available to model relations other than the linear cases. This is commonly referred to as Generalized Additive Models (hereafter GAM; Brenning, 2008; Park and Chi, 2008; Pourghasemi and Rahmati, 2018) where continuous and categorical properties can be modelled in addition to ordinal ones, as well as other type of effects acting over space and time, also at the latent level (not directly expressed as a predisposing factor in the data). A GAM formulation, in analogy to the GLM one shown in Eq.1, can be summarized as follows:

$$\eta P = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + f_1 + f_2 + f_m \quad (5)$$

where the  $f$  terms can be any type of nonlinear functions. For instance, for a given ordinal property, e.g., the Slope Steepness, a nonlinear function could consist of a random walk of the first order (Bakka et al., 2018), which, for every class of slope, accounts for the ordinal dependency that exist between adjacent

classes. This option contrasts with the use of categorical properties, e.g., Geology, where each class is modelled independently from the other.

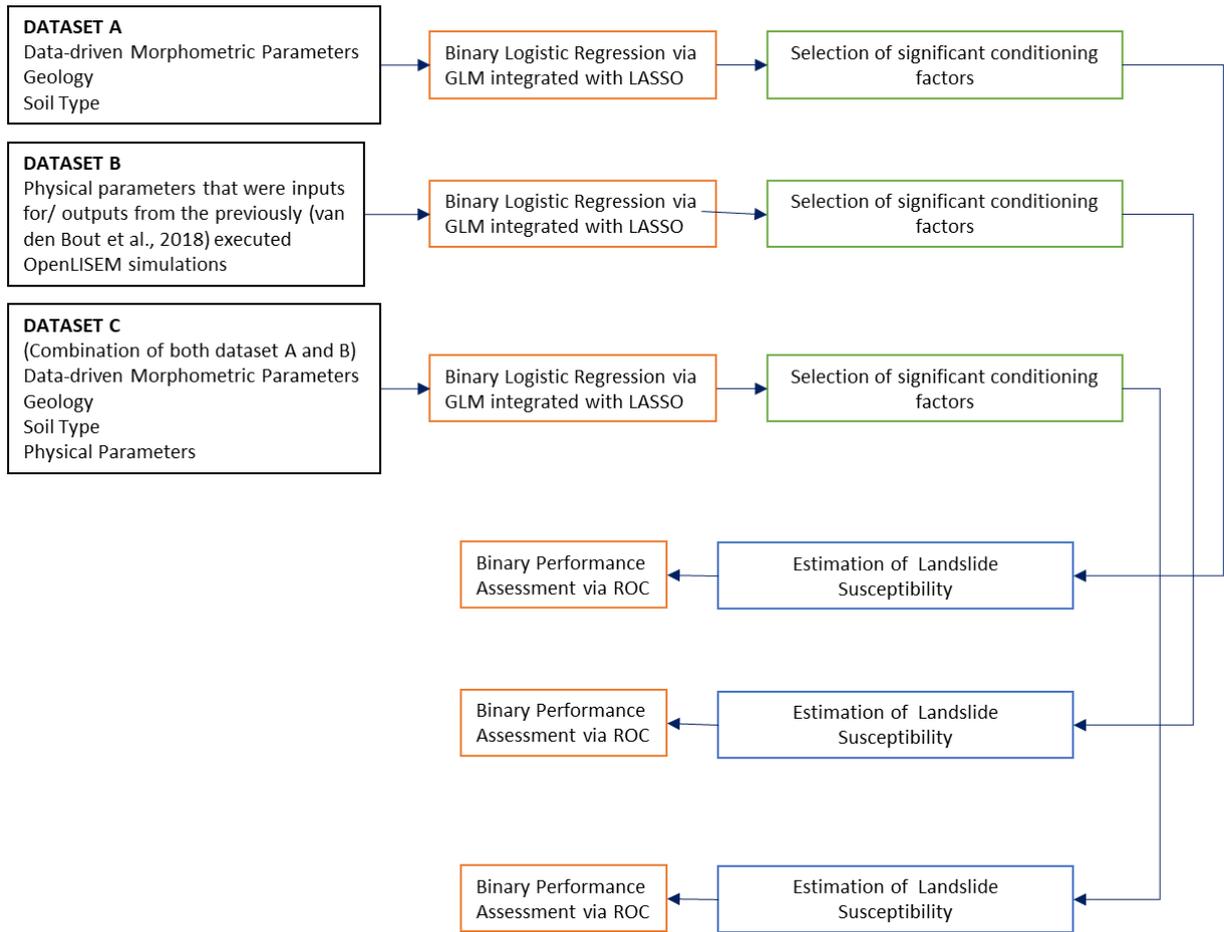


Figure 3.1 The Methodology Flowchart for First Phase of the Study

As for a more rigorous uncertainty estimation procedure, one can opt for a Bayesian formulation instead of its frequentist counterpart. In a frequentist formulation the uncertainty is typically estimated as a separate analytical step from the actual fit. In fact, on the one end, one builds a specific model fitted to the entirety of the data. And on the other hand, one separately implements large number of resampling/bootstrapping routines to estimate the potential uncertainty of a given dataset. The Bayesian formulation provides the same information but obtains the model estimate and its associated uncertainties at the same time. For instance, in Bayesian modelling, every component of the model is estimated with a distribution, which is not derived from the same type of resampling schemes mentioned above for the 10-fold cross validation routines used in the LASSO implementation. For this reason, a Bayesian GAM may offer a more complete overview of the susceptibility in a given area, possibly leading to better performance because of the higher flexibility of a GAM with respect to GLM. This greater flexibility and richness in information (uncertainty) comes with a toll which usually inflated the computational times. One solution in the literature is to use the Integrated Nested Laplacian Approximation (hereafter INLA; Bakka et al., 2018) instead of the most common option found in Markov Chain Monte Carlo (hereafter MCMC; Zhou et al., 2003) studies, where a very large combination of parameters is tested. INLA is fully implemented in R and offers performance in the same order of MCMCs, with lower computational costs. However, it still requires longer computational times than the simpler frequentist GLM initially mentioned. To bypass this issue, a different mapping unit has been chosen in two phases of the thesis. For models that require a fine spatial partition (10m pixel resolution), a simpler frequentist GLM with a LASSO step for variable selection is

utilized. As for the whole island of Dominica, a Slope Unit partition (Alvioli et al., 2016) has been opted in order to reduce the data size and execute more complex model architectures.

More specifically, because Dominica has a multitemporal landslide inventory for the whole island, a GAM with a temporal model component has been implemented (Figure 3.2). This is commonly referred to as Autoregressive model, which models the whole landslide data and links each temporal inventory accounting for the existence of residuals between observed (presence/absence) and predicted landslides (susceptibility) temporally adjusting for local deviations per Slope Unit. This is done under the assumption that some of these residuals could be due to path-dependency effects acting at the Slope Unit level.

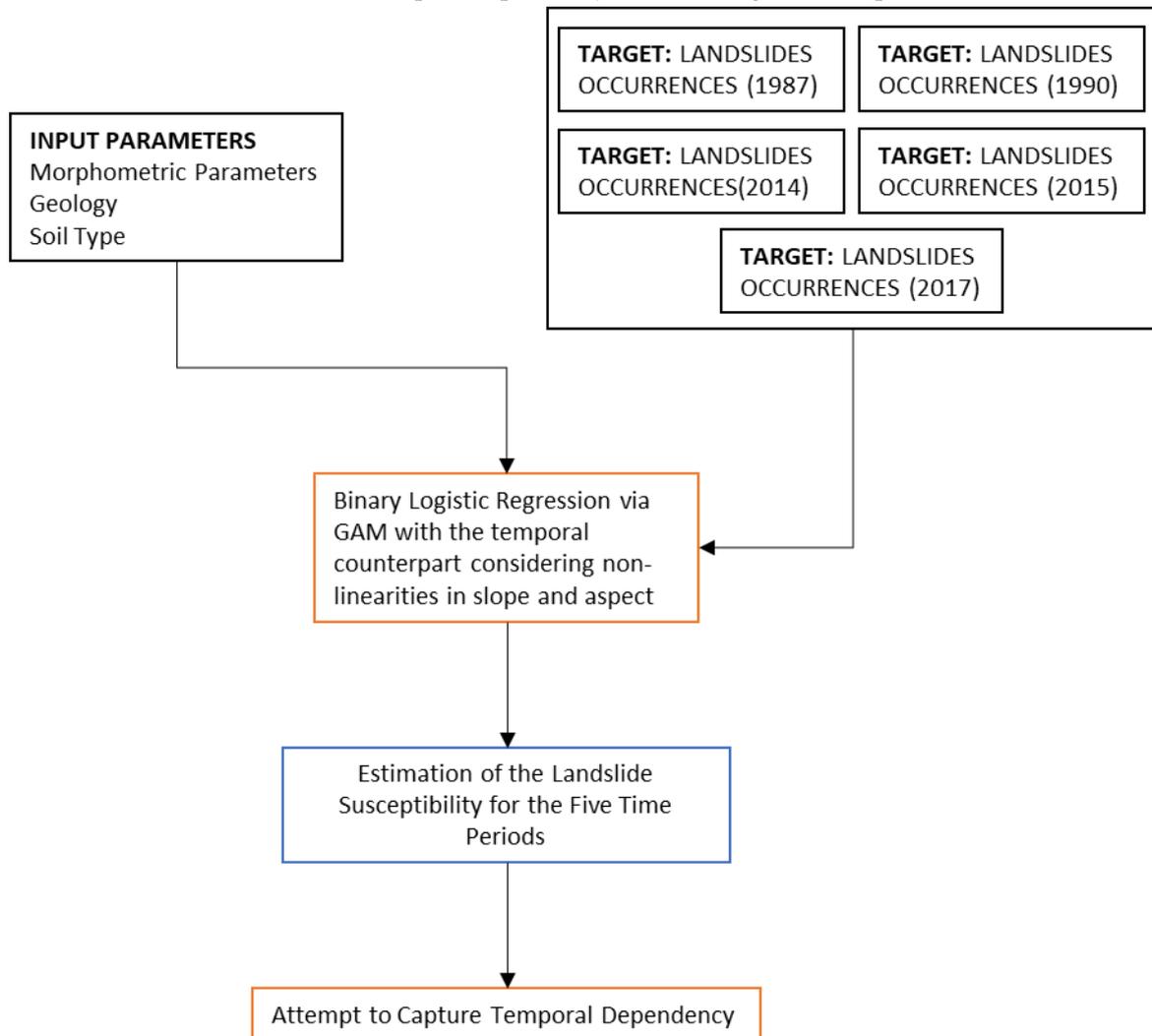


Figure 3.2 The Methodology Flowchart for the Second Phase to investigate the Temporal Dependency on Landslide Susceptibility

Finally, another modelling tool has been used too, this time disregarding the presence/absence scheme discussed before and focusing on some spatial characteristics linked to landslide areas. Specifically, the following were computed:

- 1- The sum of all landslide areas in a given Slope Unit.
- 2- Divide the sum by the extent of the Slope Unit the landslides belonged to.
- 3- Express this ratio in percentage to convey how much of a given Slope Unit failed because of landslides in each multi-temporal case.

As a result, the target variable obtained is a continuous one and cannot be modelled with a binomial GLM. Therefore, for continuity with the modelling framework explained above, the same frequentist GLM framework has been implemented but by using a Gaussian likelihood instead of a Bernoulli (Refice and Capolongo, 2002; Zhou et al., 2003; Pourghasemi and Rahmati, 2018). To measure the model performance, instead of the AUC, the metrics that are typical of continuous properties, namely Mean Absolute Error, Mean Square Error and the actual error measured as the difference between observed and predicted estimates, is used. The methodological flowchart for this model is given in the Figure 3.3.

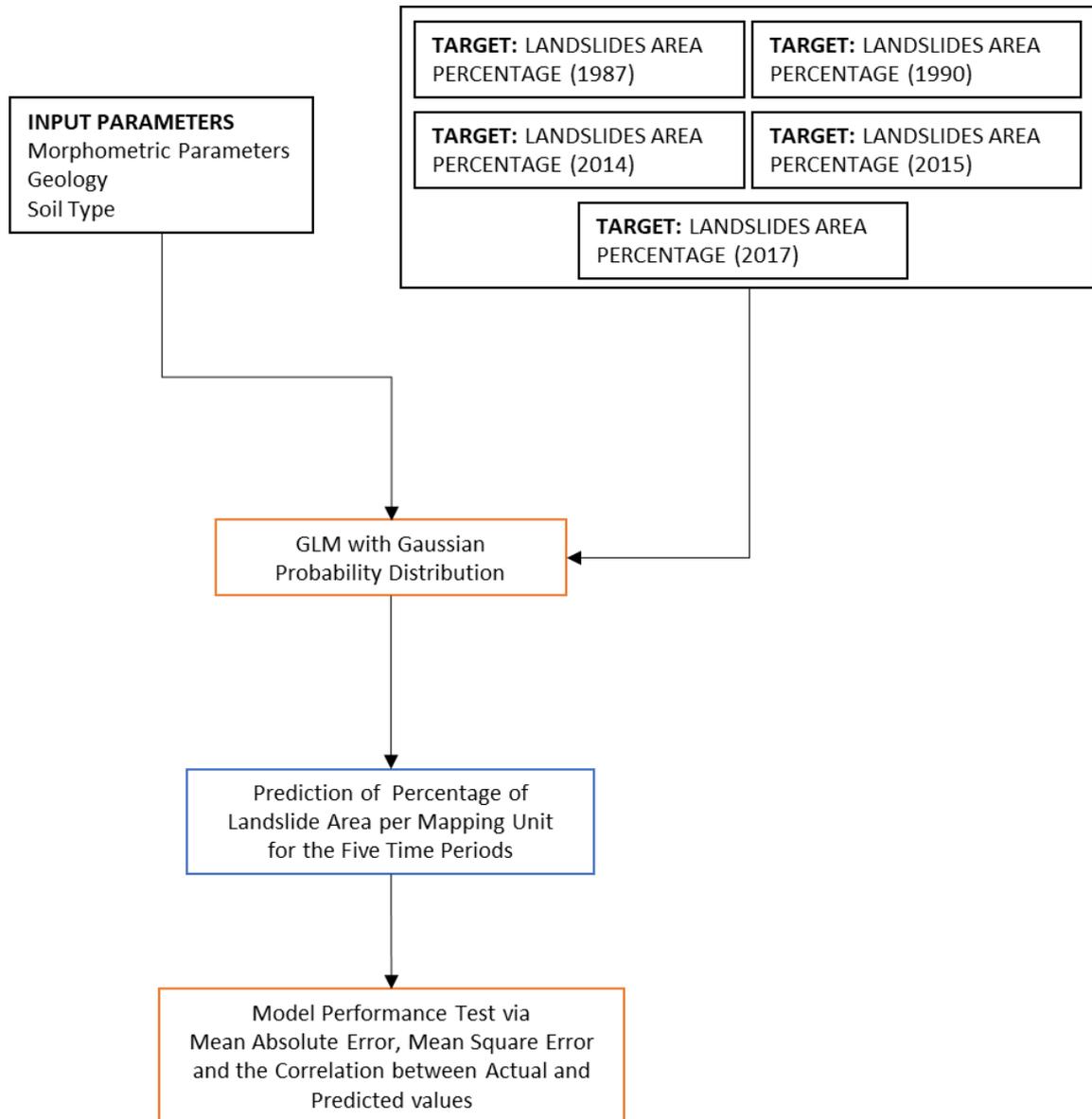


Figure 3.3 The Methodology Flowchart for the Final Phase of the Study where an Initial analysis is done on the Spatial Patterns of the Landslide Run-out Areas

### 3.1. Landslide Conditioning Factors for the various Models of this Study

As mentioned above, this study considers five models out of which the three GLM used for the initial phase will be executed at a pixel-level for the region of Grand Bay in Dominica. In the succeeding phases of the study, the statistical frameworks will be implemented at a slope unit level for the entire region of Dominica. The various parameters that would be an input to these models are summarized in the following sections based on their mapping unit (pixel or slope unit).

### 3.1.1. Parameters for the Models at Pixel-Level

This phase aims to understand how the spatial characteristics of the landslide susceptibility changes with respect to the input dataset, specifically, whether inclusion of physical parameters is relevant. In this study, the physical parameters are obtained from previous study in the region of Grand Bay by van den Bout et al., 2018. An open-source OpenLISEM hazard model considering the Saint-Venant water flow, catchment surface and sub-surface hydrology, Pudasaini two-phase generalized debris flow equations and iterative slope failure method, was developed for the simulations of the impact of hurricane Maria. OpenLISEM simulates the effect of rainfall on a landscape by utilizing various spatial data like the DEM, soil characteristics, land use and infrastructures. The OpenLISEM model was simulated in a data environment of resolution 10 meters and had various hydrological and geotechnical input parameters. The input parameters which were utilized in the OpenLISEM model were interpolated from the data obtained via field observations. These field observations were coarse and other available remote sensing data was taken advantage of to obtain a continuous spatial pattern. From these input parameters, only certain spatially distributed factors were considered for the statistical framework, i.e., the data which was specific only for the outlet point was not taken into account. From the simulations of the OpenLISEM, relevant factors were utilized in the statistically based model. This model results were validated by van den Bout et al. (2018) by investigating the eradication, exposure and damage to the region post-Maria and scrutinize the change in elevation pre- and post-Maria. The specific parameters that were derived from these simulations are mentioned below.

The statistical framework aims to model the landslide susceptibility of the region with the landslide initiation points, particularly, the centroid of the scarp parts of the landslides caused by the hurricane Maria is considered as the target variable. 822 landslides were triggered in the region of Grand Bay after the hurricane. The initiation of these landslides would represent the presence in the target variable in the binomial model. From the OpenLISEM parameters, bulk density, saturated hydraulic conductivity (hereafter  $K_{sat}$ ), water porosity, Manning's n, pore pressure, soil depth, soil moisture, land cover, normalized difference vegetation index, leaf area index, local drainage density, maximum flowing height and safety factor are integrated into the statistical model and the relevance of the physical parameters is studied.

Bulk density is the volumetric density of the soil, that is, it is the weight of a soil in a given volume. A higher bulk density curbs the length of the roots (Blake and Hartge, 1986). Soil texture is the main factor which affects this parameter. Higher bulk density would generally mean sandy soils. Soil with high bulk density generally lead to shallow landslides. The ability of the pores of a saturated soil to transmit water and is the coefficient that defines the relation between hydraulic gradient of Darcy's Law and the water movement speed is  $K_{sat}$ . Darcy's law characterizes how a fluid would flow in a porous medium.  $K_{sat}$  depends on the various horizons of the soil (Ziegler et al., 2004, Zimmermann et al., 2006). An increase in  $K_{sat}$  of the topsoil destabilizes a slope and increases the landslide susceptibility. It is a coefficient which denotes the surface roughness, that is, the amount of friction a flowing water experiences in the region. It has been computed by Robert Manning in 1889 (Yen, 1992) and was selected based on the various guidelines (Arcement and Schneider, 1989). This coefficient affects the flow velocity, the higher the Manning's n the lower the velocity of flowing water. This reduced velocity of the flow leads to accumulation of the debris resulting in extensive damage and high landslide susceptibility. Soil depth denotes the depth from the topsoil to the bedrock. Landslides occur when the shear stress is more than a threshold value for a slope failure (Terzaghi, 1962) and soil depth play a role in affecting this shear stress. The local drainage density is the ratio of the length of the channels to the total area of the mapping unit and it is vastly dependent on the physical and climatological characteristics of the catchment area. Slope instabilities are frequently caused by

adverse ground water. A good drainage system aid in altering the ground water conditions and reduce the water potential of the pores, thus, stabilizing shallow translational landslides (Stanic, 1984).

The vegetation parameters that were included were the vegetation cover, normalized difference vegetation index (hereafter NDVI) and the leaf area index. The vegetation cover represents the ratio of the area of the mapping unit with vegetation to the total area of the mapping unit. While this gives an understanding of the area covered with vegetation, NDVI gives on insight on what type of vegetation. It is calculated according to the work by Kogan with the National Oceanic and Atmospheric Administration (1994). Negative NDVI denotes water, NDVI close to zero denotes barren, low positive values (0.2 to 0.4) denotes grasslands and shrubs and the higher values of NDVI denotes rainforests. Leaf area index is the projected area of leaves per unit of surface land (Waring and Running, 2010). This is yet another way to represent the flora of the region. Usually leaf area index lies between 0 to higher than 6, where 0 denotes barren land and the higher values denote dense forests. The taller and denser forests have an increased root length which binds the various layers of the soil (O'Loughlin, 1984). Thus, a higher value of the vegetation indices means less landslide susceptibility.

The above parameters are relatively uniform over time, but even certain dynamic physical parameters were considered in this study like the initial soil moisture, soil water potential, pore pressure and maximum flowing height. These parameters were derived based on the effects of the rainfall caused by the hurricane Maria. While the initial soil moisture denotes the water content of the soil, the soil water potential denotes how much more moisture could be absorbed by the soil. An increase in soil moisture increases the weight of the soil and thus it is proportional to the shear and normal stress. Another parameter that affects the soil moisture is the pore pressure which denotes the pressure of the water held between the pores of the soil. An increase in the pore pressure decreases the resistive shear of the soil on a slope by decreasing the degree of effective stress (Stanic, 1984). This decrease in the stress leads to an increase in the slope instabilities. The maximum flowing height denotes the vertical displacement of the water due to the increased precipitation caused by the hurricane.

Another very relevant parameter utilized is the factor of safety. This denotes the ratio between the driving and the resisting forces on the surface. For a slope to be instable the driving force should be greater (Romani et al., 1972). This is the slope instability outcome from the OpenLISEM simulation considering the effects of the hurricane Maria. The simulation had considered the iterative slope failure method which determines the failure from the bottom to the top of the landslide iterating in the direction of upslope. Certain assumptions like that the failure surface is always parallel to the surface gradient and the propagation of the landslide can happen only in the upslope direction. This means that the OpenLISEM model is more compatible towards modelling transitional landslides rather than rotational. These are the physical parameters that are being used in this study.

Respecting the data environment of the physically-based model, the morphometric parameters are generated at the same scale. These parameters are frequently used in statistical models for landslide susceptibility. These include elevation, slope, Northness and Eastness of aspect, plan curvature, profile curvature, relative slope position, topographic wetness index (hereafter TWI), stream power index (hereafter SPI) and distance to streams. With these the thematic properties of geology and soil type is also considered.

The elevation is one of the frequently utilized conditioning factors as it is easily available and, in this study, the Lidar data is utilised. More than 80% of the area contains the lidar data only the small regions

in the north west of Grand Bay. For this area, the SRTM data is used and the DEM is merged respecting the hydrological setting of the region. The part which does not have the lidar data contains only 60 landslide initiation points out of the 822 landslides that have occurred. From this elevation, a major and relevant conditioning factor, slope is derived. The slope has been calculated via the method proposed by Zevenbergen and Thorne (1987). The slope is one of the approaches to parameterize the effect of gravity, thus it is a predominant factor in the prediction of landslides. Another derivative of the elevation is the aspect which represents the direction in which the slope faces. It has been initially calculated according to Zevenbergen and Thorne (1987). It is generally accepted that for shallow landslides the aspect is a proxy for strata attitude but also of wet and dry soils because of the exposure to sunlight (Peng et al., 2014; Lombardo et al., 2016; Zhang et al., 2016). However, the aspect is a circular variable and therefore requires to be modelled non-linearly, which is traditionally done by binning the aspect [0,360) range into finite number of classes and considering each one of them independently i.e. assuming that there is no relationship between the classes. However, an alternative exists where the cyclic signal of the aspect is decomposed into two simpler components, to be used as linear properties. This is the case for Northness and Eastness. While Northness is the cosine of aspect (in radians), the Eastness is the sine of aspect. Another parameter is the relative slope position which is calculated as the elevation of the cell relative to the elevation of the ridge i.e., the cell it flows up to and the cell it flows down to, the valley (Freeman, 1991; Böhner and Selige, 2006). The next parameter is the curvature which is a derivative of the slope and denotes the rate of change in the direction of the flow (Heerdegen and Beran, 1982). This factor is decoupled into two: the plan curvature, which is perpendicular to the direction of slope and the profile curvature which is parallel to the direction of slope. These factors help in understanding whether the flow converges or diverges on the face of the slope.

The TWI is a steady state wetness index as a function of slope and the upstream area which quantifies the topographic influence on hydrological processes (Beven and Kirkby, 1979). It is given by the natural logarithm of the ratio between upslope area (or flow accumulation) and the tangent of slope (in degrees). This index is a measure of how much water is potentially retained as a function of the morphometric conditions of the region. Whereas the SPI quantifies the erosive power of streams (Hack, 1973) and this is given by the product of the upslope area (or flow accumulation) and the tangent of slope (in degrees). It is a relevant conditioning factors as the high slopes which have moderate level of erosion are more susceptible to landslides i.e., the increase in erosion increases the mass of the debris and thus intensify the extent of exposure due to a landslide (Bartarya and Valdiya, 1989). Another factor which would affect the slope stability is the proximity to the streams. The regions near to the streams are exposed to a river undercut due to the impact of the flow of water in the slope of the surface which in turn generates a slope instability (Gokceoglu and Aksoy, 1996). The topography of the region is Grand bay is flat only in the coastal regions thus all the channels that are considered in this study flows through uneven terrain.

The thematic properties of geology type and the soil type are the categorical parameters in this study. These act as a proxy for many physical parameters, for example, the type of soil is a proxy for the soil characteristics like the saturated hydraulic conductivity, soil moisture, water potential and so on. Over the years, the lithological and pedological information has been passed through a statistical framework for modelling landslide susceptibility. The region of Grand Bay consists of 10 geology (Geo 1: Young Pleistocene Volcanics; Geo 2: Young Pleistocene Pelean Domes; Geo 3: Ignimbrite on Young Pleistocene; Geo 4: Young Pleistocene Craters; Geo 5: Young Pleistocene Ignimbrites; Geo 6: Pleistocene apron of block and ash; Geo 7: Pleistocene Pelean Domes; Geo 8: Recent River Gravel and Aluminium; Geo 9: Ignimbrite on Pliocene Volcanics; and Geo 10: Pliocene Volcanics) and 6 soil type classes (Soil 1:

Allophanoid Latosolics; Soil 2: Kandoid Latosolics; Soil 3: Protosols; Soil 4: Skeletal; Soil 5: Young Soils; and Soil 6: Unclassified).

In order to comprehend the relevance of various factors for landslide susceptibility mapping, the statistical framework is executed with three different covariate sets: A) morphometric parameters, geology and soil type; B) physical parameters and C) combination of all the above mentioned parameters. The Table 3.1 summarises the characteristics of the various conditioning factors utilized in this study. The maps of the conditioning factors are enclosed in Appendix II.

Table 3.1 Characteristics of the Conditioning Factors (Median for continuous data with unique values and Mean for the continuous data is tabulated)

<b>CONDITIONING FACTORS</b>	<b>CHARACTERISTICS</b>	<b>MINIMUM</b>	<b>MEAN/ MEDIAN</b>	<b>MAXIMUM</b>
Bulk Density	Continuous parameter having 11 unique values	1.274	1.367	1.771
Saturated Hydraulic Conductivity	Continuous parameter having 11 unique values	1.512	16.909	16.909
Manning's n	Continuous parameter having 9 unique values	0.050	0.150	0.204
Soil Depth	Continuous parameter	0.000	211072	860978
Soil Moisture	Continuous parameter having 11 unique values	0.310	0.424	0.467
Water Potential	Continuous parameter having 9 unique values	0.167	0.181	0.207
Pore Pressure	Continuous parameter	0.000	0.037	0.041
Total Infiltration	Continuous parameter	0.000	0.037	0.041
Vegetation Cover	Continuous parameter having 6 unique values	0.010	0.900	0.900
NDVI	Continuous parameter having 6 unique values	0.005	0.767	0.767
Leaf Area Index	Continuous parameter having 7 unique values	0.010	5.756	5.756
Local Drainage Density	Continuous parameter having 8 unique values	1.000	3.000	9.000
Maximum Flowing Height	Continuous parameter	0.000	0.169	4.638
Factor of Safety	Continuous parameter	0.819	6.987	1000.000
Elevation	Continuous parameter	0.000	376.000	1118.700
Slope	Continuous parameter	0.000	30.574	83.260
Northness	Continuous parameter	-1.000	-0.205	1.000
Eastness	Continuous parameter	-1.000	0.162	1.000
Plan Curvature	Continuous parameter	-26.674	0.000	8.001
Profile Curvature	Continuous parameter	-0.538	0.000	0.266

Topographic Wetness Index	Continuous parameter	-0.538	4.868	20.185
Stream Power Index	Continuous parameter	0.000	136981.613	694936.979
Relative Slope Position	Continuous parameter	0.000	0.499	1.000
Distance to Streams	Continuous parameter	0.000	128.000	813.900
Geology Type	Categorical parameter of 10 classes	N/A	N/A	N/A
Soil Type	Categorical parameter of 6 classes	N/A	N/A	N/A

### 3.1.2. Parameters for the Models at Slope Unit Level

For the investigation on temporal component, the same mapping unit is required to record repeated landslide occurrence over time. The simplest option at hand would be either to create a squared lattice with a coarser resolution although this would cut through some landscape features, hence it would not respect the geomorphology of the area. A different but geomorphologically-sound solution would be to compute Slope Units (Alvioli et al., 2016). In this study, the *r.slopeunits* has been used to compute the slope units of the area. Specifically, they were generated with the parameters, namely, minimum surface area of the slope units to be 100,000 m<sup>2</sup> (the area that the slope unit partition would try to converge) the initial flow accumulation area threshold to be 1,000,000 m<sup>2</sup> (the initial size of the catchment subdivision from which the *r.slopeunits* model tries to fit the above mentioned slope unit sizes) and the circular variance of terrain aspect as 0.35 (parameter that constraints the homogeneity of the exposition per slope units; the closer to 0, the less the aspect variance that would be accepted; the closer to 1, the more *r.slopeunits* would accept a very large variance in the aspect). After this, the generated slope units are checked whether they respect the topographic setting of the region and is then edited accordingly. The whole island of Dominica is divided into 3318 slope units (Appendix IV shows the slope units map of the island).

Thus, to accommodate the study on the temporal dependence and model a multi-temporal landslide susceptibility, the GAM is implemented at a slope unit level for the whole island. The presence/absence of the landslide initiation for every time period is the target variable for this model. As the investigation is for the whole island, due to the data availability, the conditioning factors used for this are elevation, slope, aspect, plan curvature, profile curvature, topographic wetness index, stream power index, distance to streams, geology and soil type. The previously mentioned methods are utilized for the derivation of these parameters.

All these parameters except the slope and the aspect are used linearly. In order to resample the pixel-level continuous data, the mean and standard deviation of the factors in the particular slope unit are computed using zonal statistics. In the case of the categorical conditioning factors (geology and soil type), the ratio of the area of the specific class to the total area of the slope unit is computed.

The GAM considered the same non-linear function with random walk of the first order (Bakka et al., 2018) for both the mean slope and aspect but treated aspect as a cyclic covariate. The mean slope was grouped into 20 classes and it was handled as a non-linear covariate because there exists a relation between the classes of mean slope steepness. For instance, the class 5-10 degrees contains smaller values than the class 10-15 degrees and it contains larger values than the class 0-5 degrees, implying a sorted structure in a reclassified slope factor (or any other ordinal factor). This information is of value for landslide susceptibility assessment because adjacent classes should behave more similarly than far away classes. For the conditioning factor aspect, it was classified into 16 classes of interval 22.5 degrees and the majority of the class in the

particular slope unit was considered. This reclassified factor has been also modelled to account for adjacent-class dependence, but in a cyclic structure where also the first (337.5-360 degrees) and last (0-22.5 degrees) classes are constrained to behave similarly.

In the final phase of the study, where it is attempted to model the landslide area percentage per mapping unit, the GLM is implemented with a Gaussian probability distribution as the target variable is a continuous data. Similar to the previous phase, the covariate dataset is the elevation, slope, aspect, plan curvature, profile curvature, topographic wetness index, stream power index, distance to streams, geology, and soil type. This is also implemented at a slope unit level and all the parameters are handled as linear covariates. Since there are no non-linearity considerations, the classes of the aspect are treated independently. For this, all the five landslide inventories are examined and the percentages of landslide area per slope unit are predicted for every time period.

## 4. RESULTS AND DISCUSSIONS

The statistical analysis had been carried out in order to identify the significance of the various conditioning factors (mentioned in the section 3.1) for modelling spatial landslide susceptibility, multi-temporal landslide susceptibility and percentages of landslide area per mapping unit. For the spatial landslide susceptibility, the morphometric, thematic, and physical parameters were examined, and the results have been summarized in the section 4.1. By utilizing the five landslide inventories, the study attempted to model the multi-temporal landslide susceptibility by deliberating on the ability of a statistical model being able to capture the temporal dependence (Section 4.2). Finally, a model to predict the percentage of landslide area per mapping unit had been implemented and the results are outlined in section 4.3.

### 4.1. Importance of the covariates among morphometric, thematic, and physical parameters for the landslide susceptibility modelling

For this analysis, three different models are put on scrutiny in a statistical framework. The first model is with an input dataset, “Dataset A” that has the most frequent data driven conditioning factors that are utilized for prediction of landslide susceptibility. The second model regards the input dataset, “Dataset B”, which has the physical parameters i.e., the inputs and the intermediate outputs of the simulations from OpenLISEM (carried out for the region by van den Bout et al., 2018). The third model has an input dataset, “Dataset C”, including both the first and the second dataset. The LASSO variable selector has been utilized to identify the preferable conditioning factors for landslide susceptibility mapping.

The Figure 4.1 shows the results of the LASSO variable selector. The lambda ( $\lambda$ ) denotes the regularization parameter and the top axis of the plot denotes the number of covariates and the vertical axis plots the AUC values. It can be recognised that AUC value does not significantly vary as certain covariates are removed; this says that those covariates produce no change in the predictability thereby implying their role to be negligible with respect to the overall susceptibility model. The model with dataset A had an AUC value of 0.75 and had significant 15 conditioning factors, the model based on dataset B had an AUC value of 0.72 and had 13 relevant covariates and the combined model had an AUC value of 0.77 and had 33 predictive parameters. From these AUC values, all the three models are acceptable and good (Hosmer and

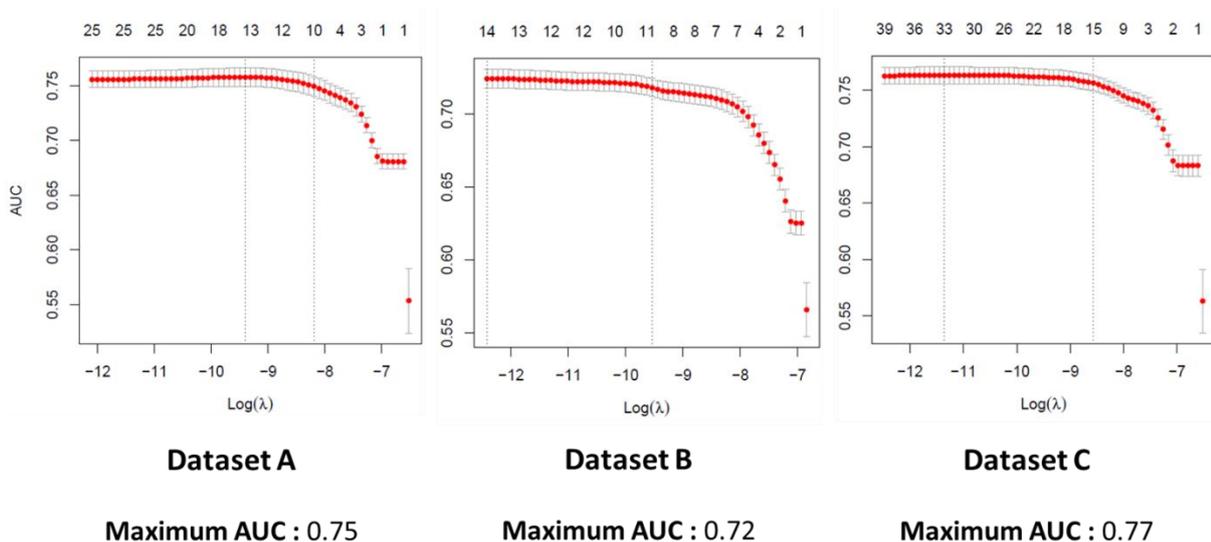


Figure 4.1 The results of the variable selector in the models with Dataset A, Dataset B and Dataset C

Lemeshow, 2000), but it is observed that the model with dataset C has better estimations of the landslide susceptibility.

In order to understand this difference in the AUC values, the regression coefficients obtained for the different models are observed. The Table 4.1 and Table 4.2 summarizes the different regression coefficients in dataset A and dataset B and its corresponding coefficient obtained in dataset C (Appendix I has the information of the types of soil and geology in Grand Bay). It is observed that in the scenarios (dataset A and dataset C), geology type 3 (Ignimbrite on Young Pleistocene); geology type 9 (Ignimbrite on Pliocene Volcanic); soil type 3 (Protosols) and soil type 6 (Unclassified) has no significance. This is due to the fact that types of geology/soil are in small areas where the recorded number of landslides is very minimal, that by reducing the proportion of the landslide occurrences in the particular class close to zero.

Table 4.1 Regression Coefficients of the conditioning factors of Dataset A and Dataset C

<b>CONDITIONING FACTORS</b>	<b>DATASET A</b>	<b>DATASET C</b>
Elevation	-0.7027	-0.6708
Slope	0.6459	0.6158
Northness	-0.3246	-0.3779
Eastness	-0.1451	-0.1682
Plan Curvature	0.0000	0.0030
Profile Curvature	0.1150	0.1342
Topographic Wetness Index	0.0000	-0.0160
Stream Power Index	0.0000	0.0038
Relative Slope Position	0.1610	0.2408
Distance to Streams	-0.1444	-0.3041
Geo 1	0.3290	0.2910
Geo 2	0.1466	0.0747
Geo 3	0.0000	0.0000
Geo 4	0.0000	-0.1733
Geo 5	-0.6878	-0.5826
Geo 6	-0.2678	-0.5171
Geo 7	0.0000	-0.0056
Geo 8	0.0000	0.2420
Geo 9	0.0000	0.0000
Geo 10	0.0000	-0.5556
Soil 1	0.0032	0.0251
Soil 2	0.0000	-0.0507
Soil 3	0.0000	0.0000
Soil 4	-0.0012	-0.0455
Soil 5	-0.3279	-0.6605
Soil 6	0.0000	0.0000

Table 4.2 Regression Coefficients of the conditioning factors of Dataset B and Dataset C

<b>CONDITIONING FACTORS</b>	<b>DATASET B</b>	<b>DATASET C</b>
Elevation	-0.5290	-0.6708
Slope	0.3913	0.6158

Bulk Density	-0.3234	0.0000
Saturated Hydraulic Conductivity	0.2708	0.3973
Manning's n	0.7754	0.4846
Soil Depth	-0.0203	0.0824
Soil Moisture	0.0000	0.0000
Water Potential	-0.5561	-0.3367
Pore Pressure	-0.0699	0.0636
Total Infiltration	0.0000	0.0000
Vegetation Cover	0.0000	-0.0310
NDVI	-0.0250	-0.0140
Leaf Area Index	-0.5686	-0.3362
Local Drainage Density	-0.2925	0.0000
Maximum Flowing Height	-0.4319	-0.2954
Factor of Safety	-0.1476	-0.0391

In the case of models with dataset B and dataset C, the total infiltration and soil moisture have a zero regression coefficient in the landslide susceptibility prediction. This observation is peculiar as it was expected that these two parameters have if not high but at least a moderate level of significance. When the spatial patterns of the above two parameters were studied, it was noticed that the region had 11 unique values for soil moisture which had a minimum of 0.31 and maximum 0.47 and for the parameter of total infiltration it was a continuous data but the changes of the values over the space was very minor (values differing in the 3rd and 4th decimal places). This observation of insignificant spatial variability over the region accounts for the non-relevance for the landslide susceptibility model of the study area.

The other conditioning factors were identified to have a significance in either one scenario (either in the dataset A or C and in the dataset B or C) or both the scenario (in the exclusive dataset A or B as well as the combined dataset C). The Figure 4.2 and 4.3 shows the plot of the covariates' (excl. geology and soil type) regression coefficients.

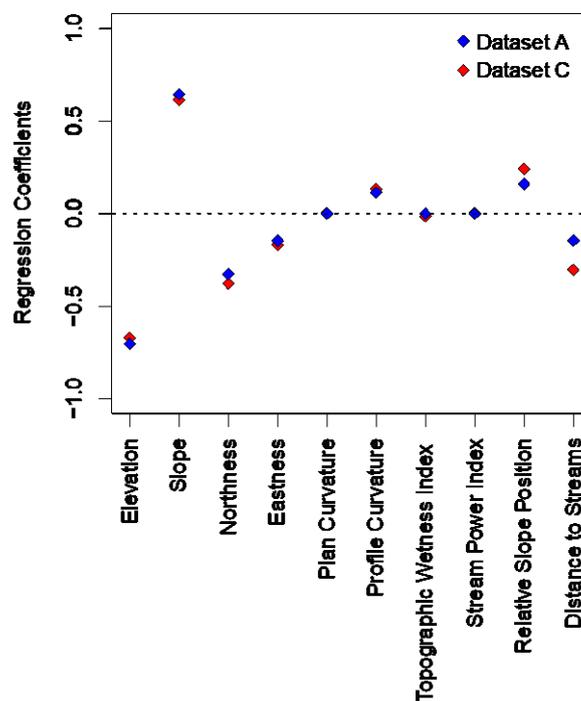


Figure 4.2 The Regression Coefficients of the continuous covariates of Dataset A and Dataset C

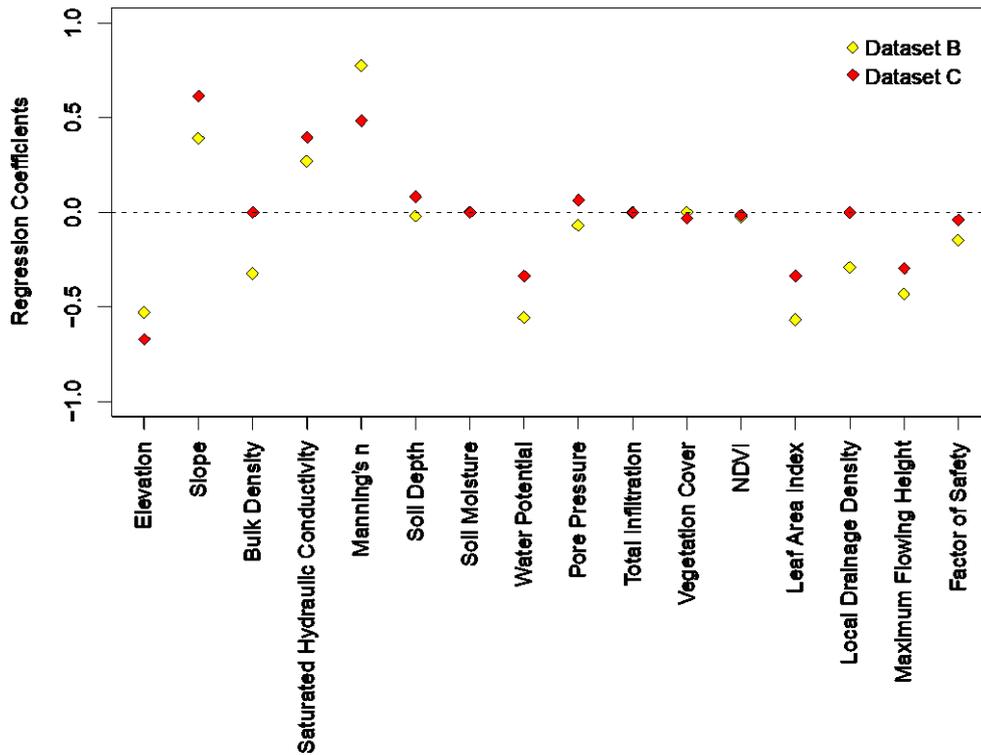


Figure 4.3 The Regression Coefficients of the covariates of Dataset B and Dataset C

The elevation and slope are common factors in the dataset A and B. The elevation has a higher relevance in scenario A than scenario B, while in the combined dataset, the elevation coefficient is quite close to the average of the two coefficients obtained in the other scenarios. In all the scenarios, positive relevance is recognized for the slope but the importance of this factor in the dataset B is lesser. The combined dataset has its coefficient for slope closer to the coefficient obtained in the dataset A. It is observed that while the slope has a positive coefficient, the elevation has a negative coefficient, this might be due to the fact that the derived parameters from the elevation have a positive relevance and this affects its regression coefficient. Explicitly for the region of Grand Bay, the low-lying coastal regions recorded more landslide occurrences, hence, the negative relevance.

According to the direction of watershed, most of the slopes of the region were facing the south and south east. Also, due to the direction of the hurricane Maria hitting the island from the direction of south west, in both datasets A and C it is observed that the Northness has higher negative coefficients than the Eastness, and, the dataset C has a slightly greater importance to the two. It is noticeable that the plan curvature does not a high significance, it has zero importance in the dataset A and its coefficient is closer to zero in dataset C. The profile curvature on the other hand is given a positive relevance in both the cases and is more significant in the combined dataset. The component parallel to the slope (profile) is able to capture the spatial patterns of the landslide susceptibility more than the perpendicular component (plan).

The topographic wetness index and stream power index has zero importance in the dataset A, while it is observed that they are of relevance in the combined dataset but the magnitude of their coefficients denote that they are not of high significance for the predictive model. The relative slope position is recognized to have a positive influence and the distance to streams is observed to have a negative influence in both the scenarios. But it is noticeable that the significance of these two have almost doubled in the combined dataset. The distance to streams captures more of the spatial patterns rather than SPI. Also, it can

be observed that the first derivatives (slope, relative slope position) and the second derivative (profile curvature) of elevation has a positive significant regression coefficient thus making the elevation have a negative influence on the susceptibility model.

While the bulk density obtains a significant negative coefficient in the dataset B, it is given zero importance in the combined dataset. This means that certain parameters from dataset A which is included in the combined dataset bear the information relevant to the bulk density thereby cancelling out the significance of bulk density. The saturated hydraulic conductivity is observed to have a positive impact on the prediction of landslide susceptibility. The significance of this parameter is higher in the combined dataset. This positive relevance of the  $K_{sat}$ , which is contradicting the reality, is due to the influence of the soil water potential and the pore pressure.

While the significance of Manning's  $n$  in the dataset B is the highest, it reduces in relevance in the combined dataset as the combined dataset includes certain terrain elements which would also account for the characteristics of Manning's  $n$ . This is mainly due to the allocation of Manning's  $n$  based on various land uses for the OpenLISEM simulations. This factor indirectly shows the influence of the natural and manmade infrastructures of the region. The lower values of Manning's  $n$  denote the barren areas thus, it has a high positive relevance for the susceptibility model.

The soil depth shows certain relevance to the predictive ability of the model, it has been observed that the soil depth has a negative coefficient in the case of dataset B, but it has a positive and slightly higher coefficient in the combined dataset. The water potential is a significant conditioning factor in both the scenarios. Though its relevance is reduced in the combined dataset. The pore pressure has a similar magnitude of coefficient in both cases but its impact changes as in the model with dataset B it is given a negative coefficient, whereas, in the combined case it is given a positive coefficient. The local drainage density loses its significance in the combined dataset, but in the model with dataset B, it is noticed that it has a negative influence. The influences of these soil characteristics in the model with dataset B being negative is due to the high positive relevance of the  $K_{sat}$ . In the case of combined dataset C, the inclusion of the various soil and geology types is playing a role in the significance of the physical soil parameters.

The vegetation cover and NDVI have no significant relevance in both the scenarios. While in dataset B, the vegetation cover is given zero importance, the NDVI is given a negative importance close to zero and in the combined dataset, the vegetation cover and NDVI is given an almost zero importance. But the leaf area index is highly significant in the model with dataset B. In both cases, it is noticed that the leaf area index has a negative coefficient, but its relevance gets reduced in the combined dataset. The vegetation cover and NDVI were observed to have only 5 unique values and in those 3 values were closer to each other whereas the leaf area index had 7 classes with varied distribution. Thus, the statistical framework selects the leaf area index as the most relevant.

The maximum flowing height in both the scenarios endures a negative impact on the landslide susceptibility. Its relevance decreases in the combined dataset. This was a continuous parameter with a range of 0 to 4.64. The factor of safety is the slope instability parameter obtained from the OpenLISEM simulations. The factor of safety that was simulated from the OpenLISEM model and it was rescaled as every other parameter to check the relevance in the susceptibility modelling. If the factor of safety has a large regression coefficient means the statistical framework has failed in modelling the landslide susceptibility. An almost zero negative coefficient of this parameter is observed denoting that the other

conditioning factors used in the model have captured the spatial variations of the landslide susceptibility and thus making the factor of safety less relevant.

After this, the susceptibility maps were generated both in a pixel mapping unit utilizing the regression coefficients obtained via the variable selection (Figure 4.4; Appendix III has two (of the many) of the instances where the model with dataset B overestimated or underestimated). The ROC curves of the pixel-level susceptibility models inform on the capability of the model to predict unstable condition and it is recognized that the combined model performs slightly better than the model based of dataset A and this performs better than the model with dataset B (Figure 4.5).

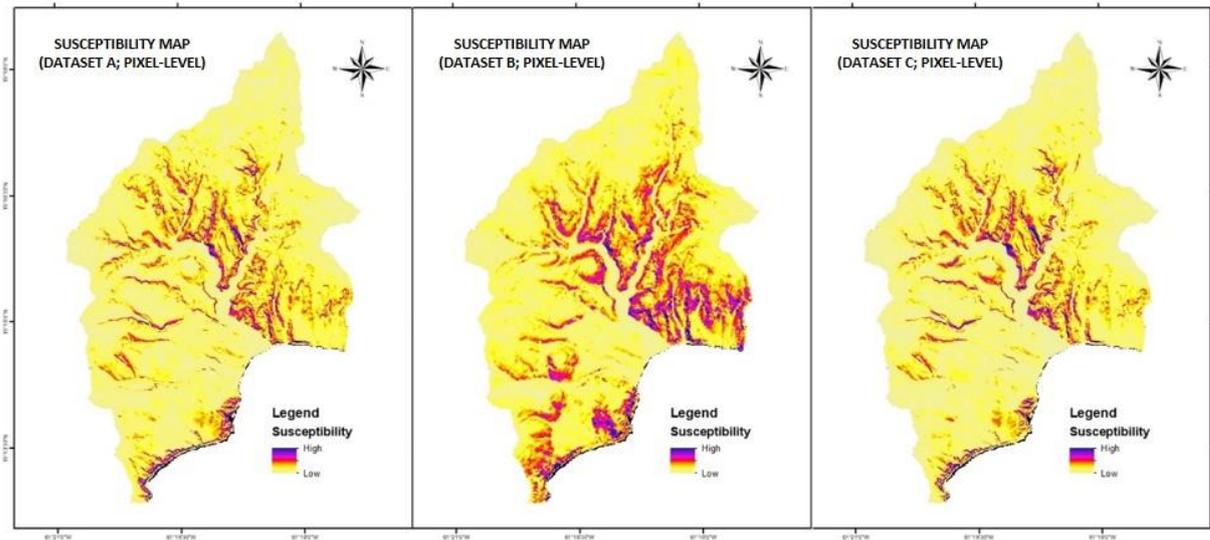


Figure 4.4 Landslide Susceptibility obtained from the three Models

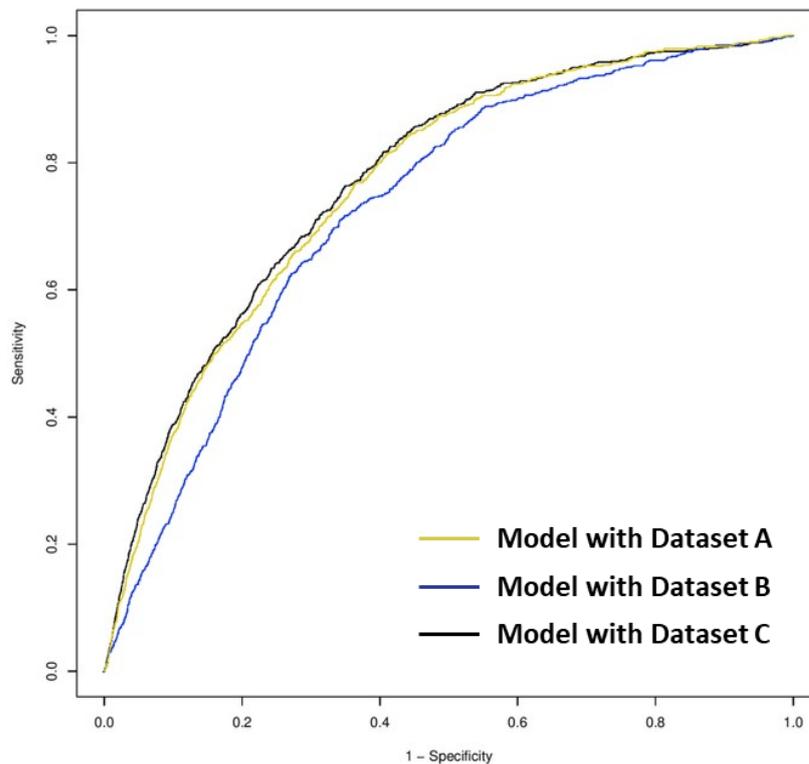


Figure 4.5 ROC curves of the Models

In order to get a clear picture of what spatial patterns of the models A and B are captured by the combined model C, the scatterplot between the predicted values of the models is graphed. The Figure 4.6 show the scatterplot between the combined model and the model with dataset A; model with dataset C and with dataset B at pixel-level, respectively. It was observed that it has a similarity of almost 90% with the first model but had only a 41% with the second model. Even though that the combined model majorly illustrates on the relevance from dataset A, it does capture certain elements from dataset B making it have a slightly better performance than the model with the frequent data driven parameters (dataset A). One of possible reasons for the low performance of the model with the physical parameters must be due to the low spatial variability and the coarse sampling size from which the corresponding input parameters were interpolated. Thus, the combined model leans towards the morphometric parameters which had sufficient variability in the values over the regions with presence/absence of landslide occurrences. Hence, the increased performance of the susceptibility models.

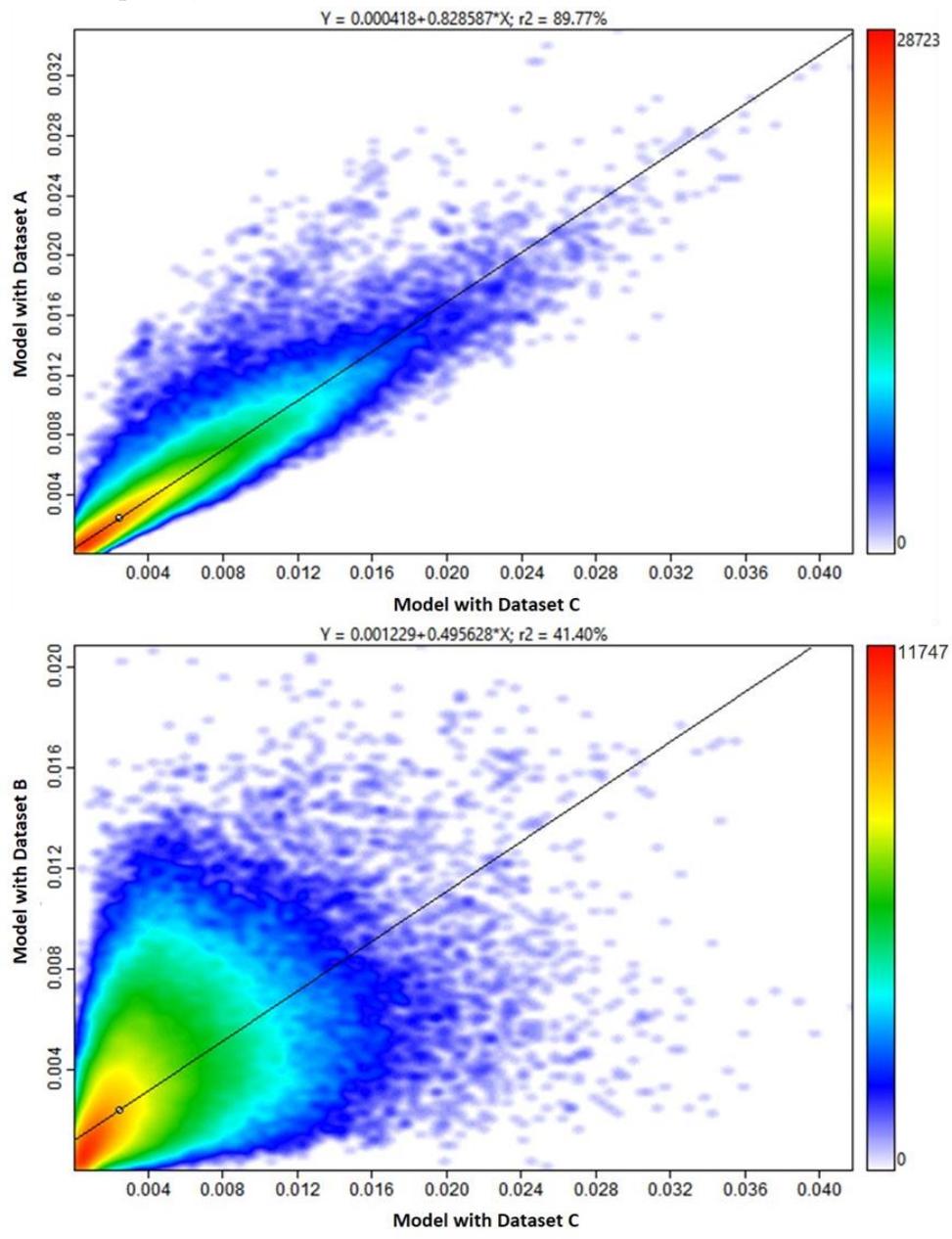


Figure 4.6 Scatterplot between the Predicted Values of the Model with dataset C and those of the Model with dataset A and B

#### 4.2. Attempting to capture temporal effects in the landslide susceptibility.

For this phase of the research, the mapping unit used was slope units and multi-temporal dataset was created. Rather than the GLM utilized for the previous study, a GAM framework with non-linearity functions for slope steepness and the aspect is considered. An autoregressive model acts on this GAM, which also integrates the temporal aspect as a non-linear function. Initially, the significance of the conditioning factors on the temporal dependency was examined. This phase as mentioned in section 2.7 has the mean and standard deviation of the continuous covariates and the area ratio of the categorical covariates as the parameters for the model.

The mean of topographic wetness index, stream power index and distance to streams, and, the standard deviation of plan curvature and stream power index have no significance i.e., their coefficient is zero. As the standard deviation denotes the variability of the parameter over the slope unit, it is observed to have some significance in the case of TWI and distance to streams. Fifteen out of twenty classes of geology show no significance and eight out sixteen soil types have no relevance to the temporal dependency. The regression coefficients of the significant covariates are plotted in Figure 4.7.

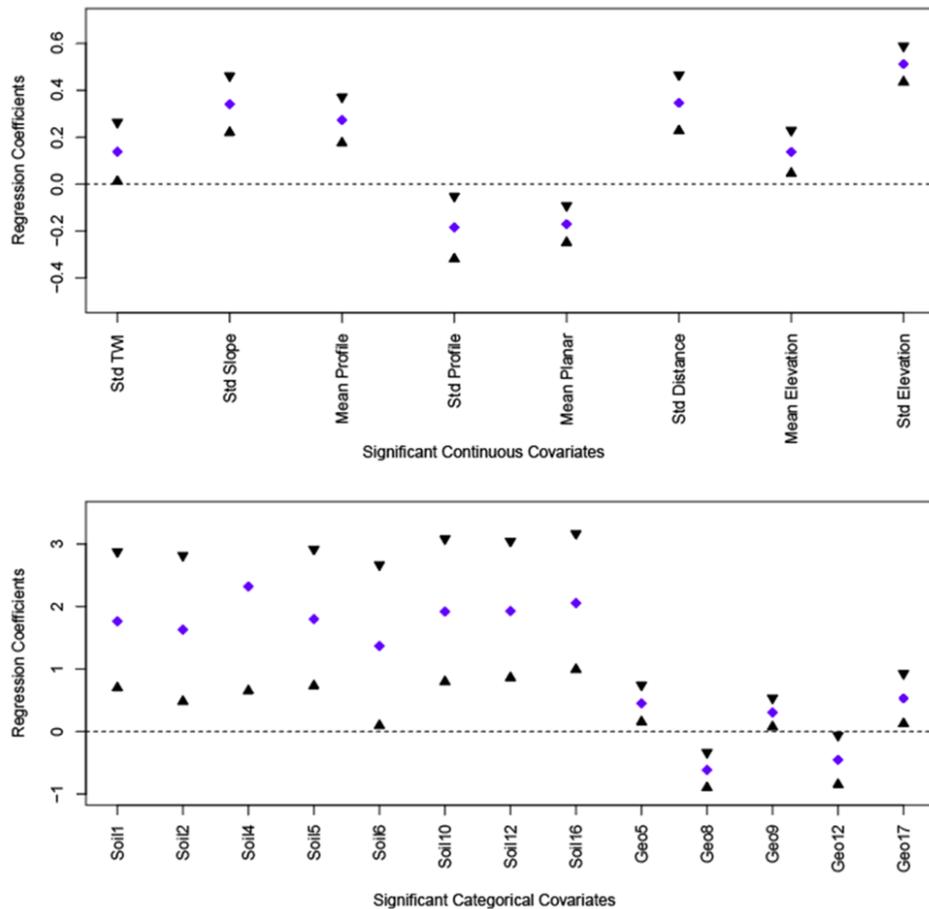


Figure 4.7 The Regression Coefficients of the Significant Covariates in the Temporal Dependency Model (Appendix I has the geology types and soil types of Dominica)

The mean slope grouped into 20 classes was considered as a non-linear factor. It was observed that the lower classes of slope had negative and a low magnitude coefficient and the higher classes had positive and relatively higher magnitude coefficients (Figure 4.8). Slope up to 30 degrees demonstrates a negative coefficient and the rate of increase in the magnitude of coefficient from slope of 0 degrees to 30 degrees is

lesser than that from slope of 30 degrees to above. The standard deviation of slope is yet another significant parameter and it positively influences in the estimation of temporal dependency.

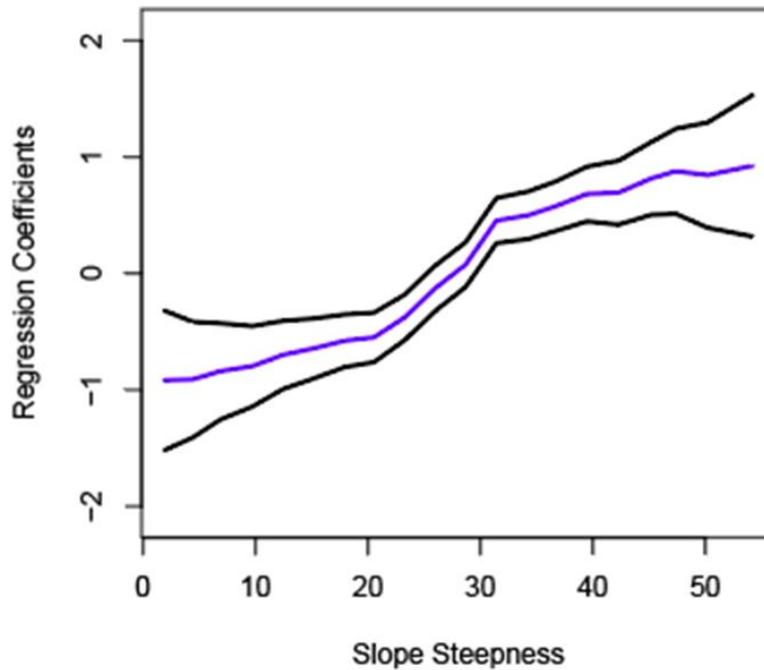


Figure 4.8 The Regression Coefficient of the Classes of the Slope Steepness

The aspect is another non-linear factor. The aspect ranges from 0 to 360 where both 0 and 360 denote the north direction, therefore, the coefficient for aspect of 0 and 360 should coincide. Thus, the aspect is integrated into the GAM with a cyclic behaviour. From the Figure 4.9, it can be observed that there is a high peak in the region of 90 to 100 and a low peak in the 250 to 300. The high peak is at the aspect with southward direction and this is because of the hurricane Maria of 2017 affecting the major of the southeast regions of the island. From the landslide inventory map (Figure 2.3), it can be observed that the west parts of the island have been affected less over the period of time thus the low peak in the aspect of west direction.

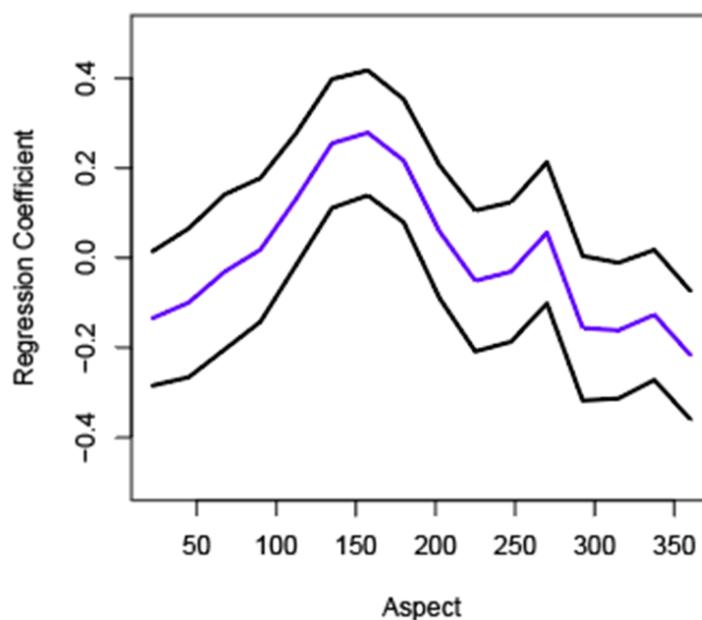


Figure 4.9 The Regression Coefficient of the Classes of the Aspect

The mean and standard deviation of the elevation is having a positive coefficient and the standard deviation of elevation is the most relevant parameter. The mean plan curvature is of a negative coefficient and is of low relevance. The mean of the profile curvature has a positive coefficient while its standard deviation is given a negative coefficient. The profile curvature is more relevant than the plan curvature in this context. The standard deviation of topographic wetness index is given a coefficient close to zero and is not of high importance for temporal dependency. Another significant parameter is the standard deviation of distance to streams which has a high positive influence. The relevant classes of geology are Ignimbrite on Pliocene Volcanic, Miocene Volcanics, Older Pleistocene Volcanics, Pleistocene Pelean Domes and Young Pleistocene Craters, and the soil types are Allophanoid Latosolics, Allophanoid Podzolics, Hydrogenic Group, Kandoid Latosolics, Protosols, Skeletal and Young Soils.

It is observable that the influences of the various morphometric parameters are varying in this model from that of the previous pixel-based model. This is due to the inclusion of the entire island as the study area which perpetually increases the spatial variability of the conditioning factors with respect to the landslide occurrences. Also, in this phase the spatial trends of all the landslide inventories are utilized unlike the previous where it was solely based on the landslides triggered by hurricane Maria.

With the regression coefficients obtained above, the mean susceptibility and the uncertainty for every time period was generated (Figure 4.10, Figure 4.11, Figure 4.12, Figure 4.13, Figure 4.14 and Figure 4.15). It can be recognized that the south-eastern and the western region of the island is moderately to highly susceptible to landslides over the years. It can be inferred that the uncertainty is quite high for these susceptibility models. This is because the statistical framework is highly data dependent on the number of presences of landslide initiations. The model for the year 2017 has a less uncertainty compared to the other four. Also, the dataset as such is relatively small in dimension. But, even in such a case if occurrence of landslides were consistent in the same spatial setting, then the model could have learnt better and the uncertainty would have been reduced. But it was not the scenario in the study area, thus, even though the model was able to capture the spatial patterns with some uniformity for the susceptibility it had a high uncertainty.

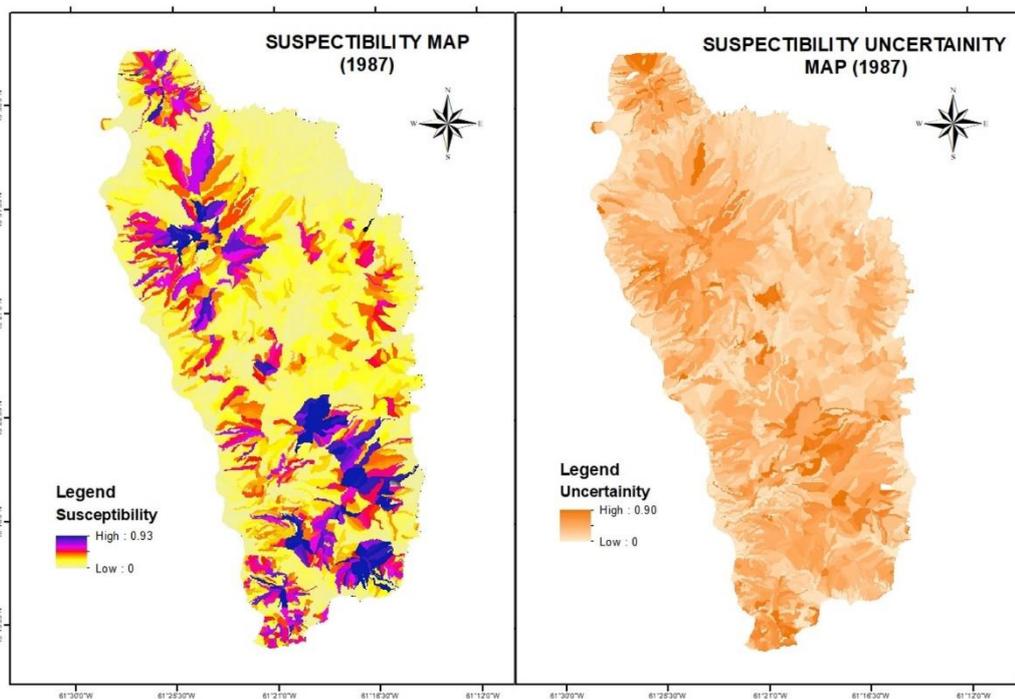


Figure 4.10 Landslide Susceptibility and its Uncertainty Maps of the year 1987

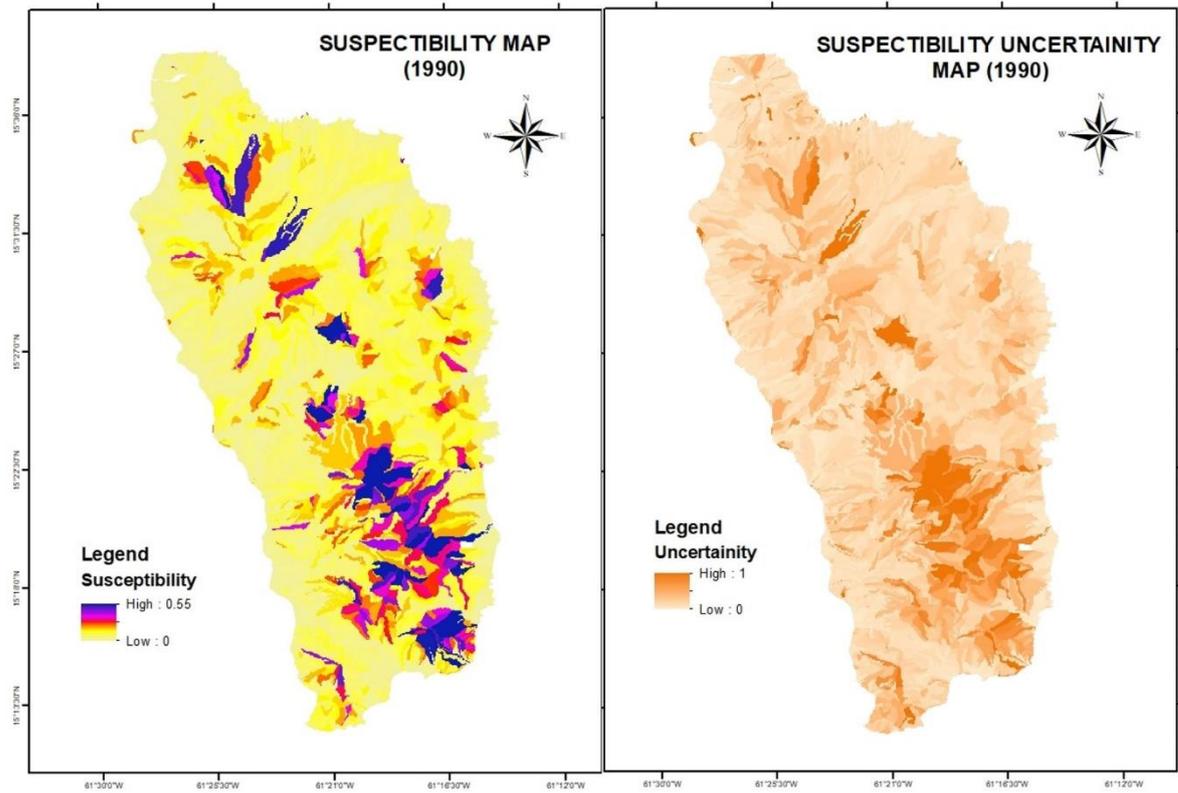


Figure 4.11 Landslide Susceptibility and its Uncertainty Maps of the year 1990

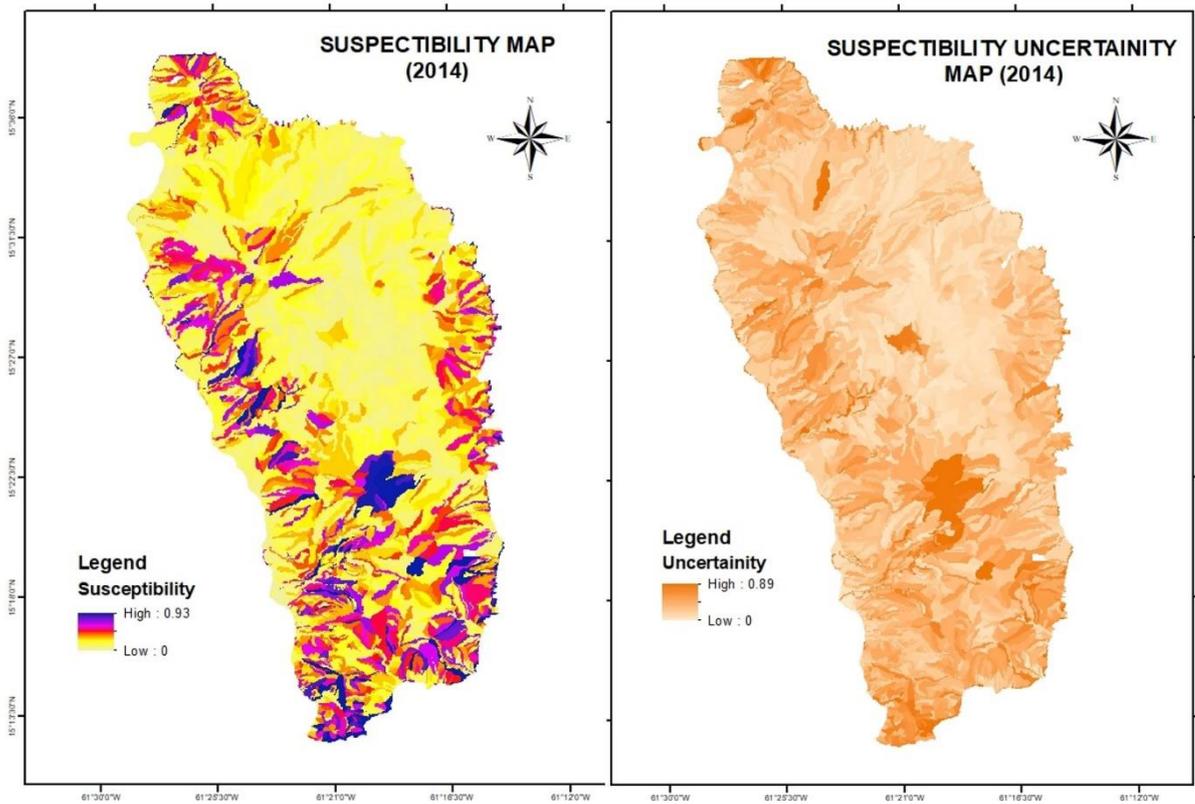


Figure 4.12 Landslide Susceptibility and its Uncertainty Maps of the year 2014

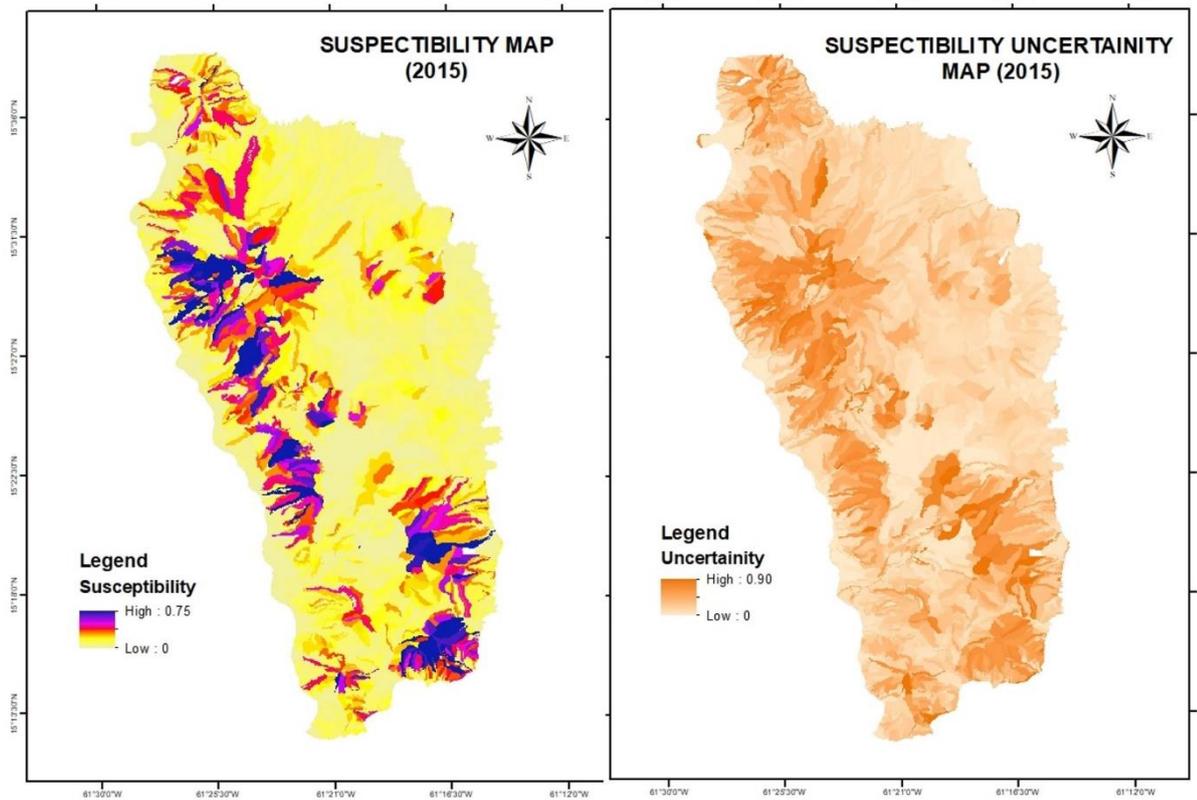


Figure 4.13 Landslide Susceptibility and its Uncertainty Maps of the year 2015

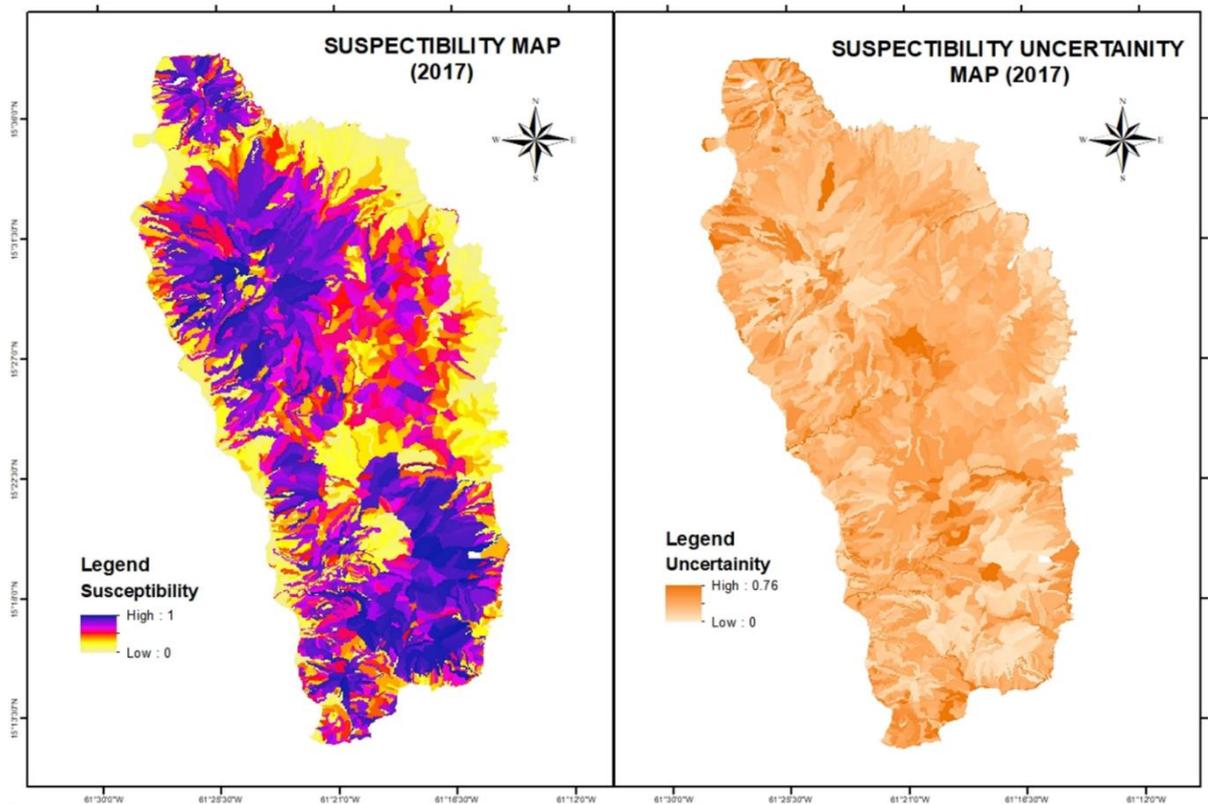


Figure 4.14 Landslide Susceptibility and its Uncertainty Maps of the year 2017

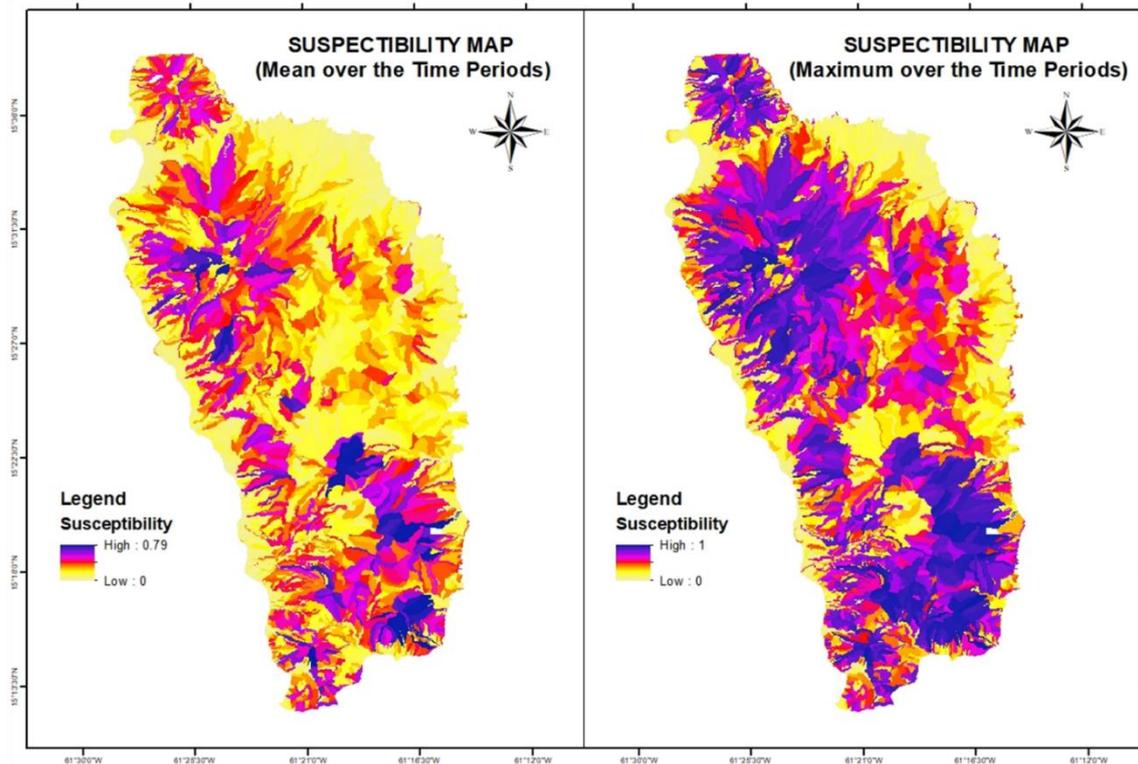


Figure 4.15 Mean and Maximum of Landslide Susceptibility over the years

Then the latent temporal effect was analysed. Certain slope units show continuously increasing or decreasing susceptibility values but the rate of change of these values are very minimal incapacitating the model to point out a definite temporal dependency. It was recognized that no clear temporal dependence exists in the study area, among successive multi-temporal landslide occurrences (Figure 4.16). This might be due to the fact that there is a spatial variability of the triggering events over the period of years, thus the model is dominated by the spatial trends rather than temporal ones.

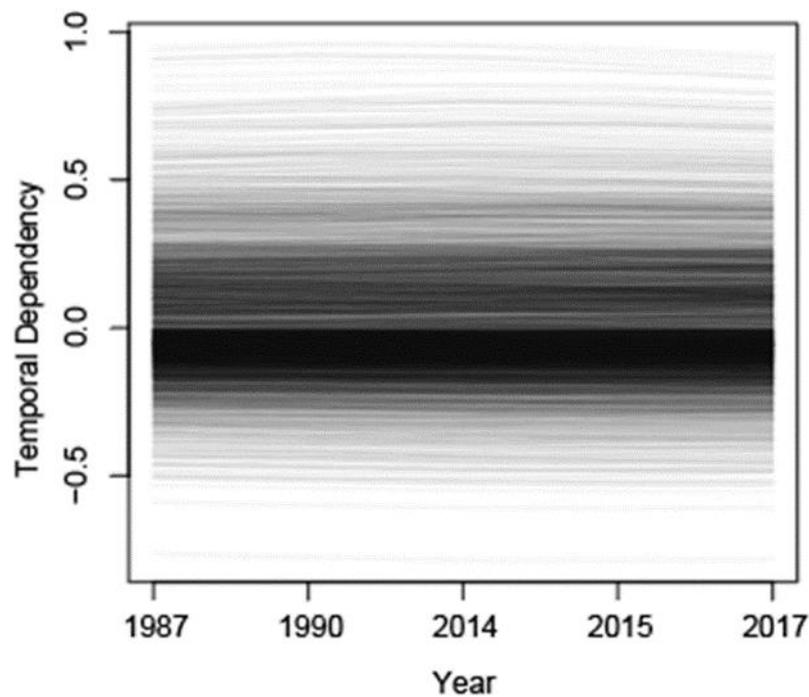


Figure 4.16 The Temporal Dependencies generated for the Various Slope Units

### 4.3. Spatial patterns of the percentage of landslide area per mapping unit over the years

Since the temporal dependence coefficient does not exist rather than a GAM the statistical framework of a GLM will be used for the prediction of percentages of landslide area per slope unit. Unlike the previous two where the target variable is binary, this model will take into account the percentage of area of the slope unit affected by the landslide. The same conditioning factors are utilized for this and the regression coefficients of different factors for the different time periods is given in the Figure 4.17.

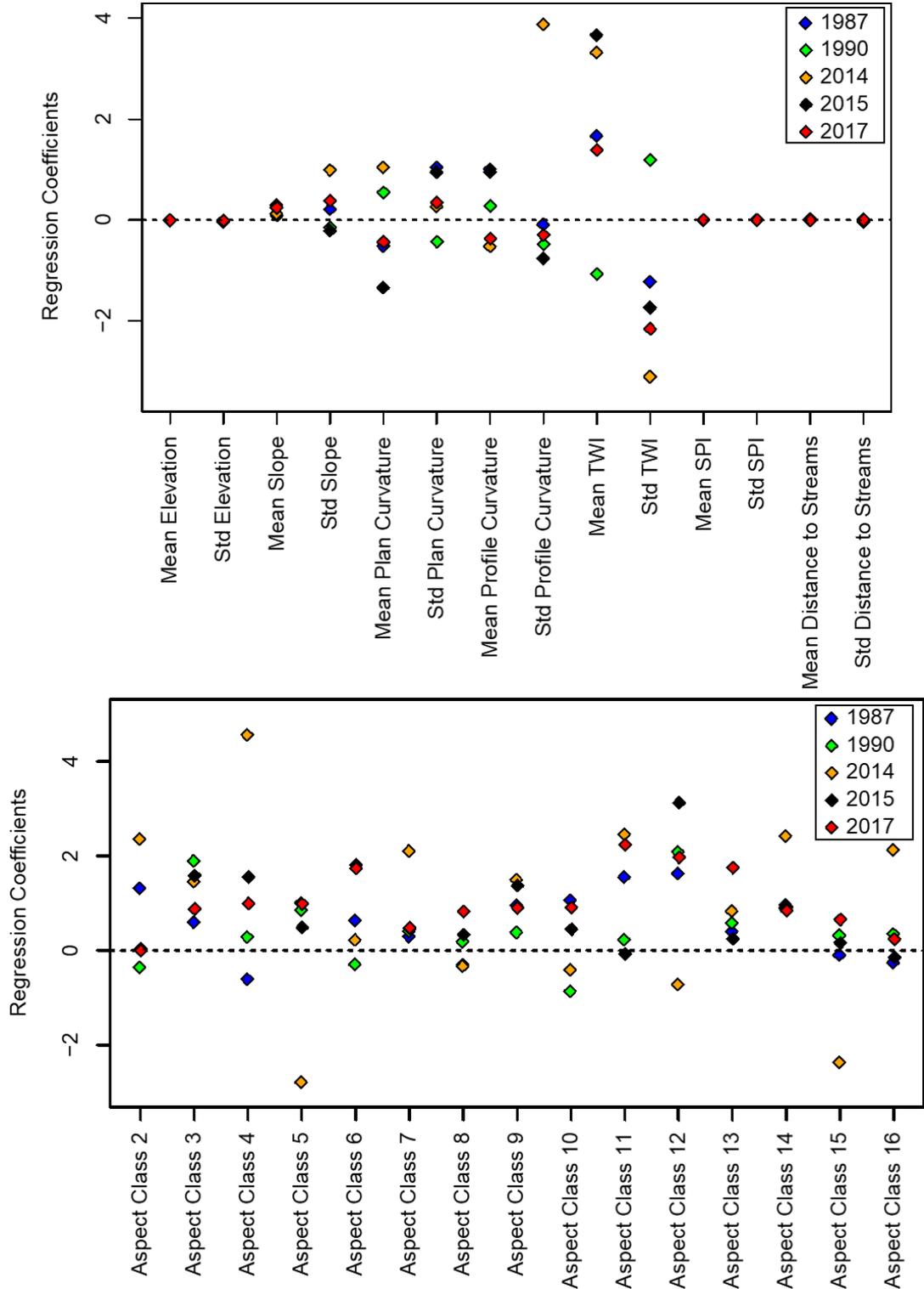


Figure 4.17 The Regression Coefficients of the Covariates (excl. geology and soil type) in Models for the Landslide Area Percentages of the years

The regression coefficient of the standard deviation of the elevation in a slope unit is higher than that of the mean elevation. And the mean and standard deviation of the slope steepness have an acceptable relevance and the coefficient of slope standard deviation is lower. This is because if the terrain is undulating and has a consistent decreasing slope, the landslide extent is larger. The plan and profile curvature are one of the most relevant parameters for this model. As these parameters aid in understanding whether the flow converges or diverges on the slope surface. The surfaces which have concave plan and convex profile curvature would increase the extent of the flow, thereby, increasing the area of landslides. The most relevant parameter in this model observed to be the topographic wetness index. The direction of the flow of a landslide is identified by the spatial patterns of TWI as it is a function of the upstream area per unit length perpendicular to the flow direction and the slope. Thus, while in the investigation of the scarp region of landslide, the TWI has no significance, for the analysis of depositional area of a landslide, it is significant. The stream power index and the distance to streams show no or less significance in this study.

Initially, with these factors, the landslide area is predicted for the slope units which had a landslide. Then the model performance was examined using Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). Finally, for the slope units which actually had zero landslides, the landslide area percentages were predicted. The landslide area percentage maps for the five time periods is given in the Figure 4.18, Figure 4.19, Figure 4.20, Figure 4.21 and Figure 4.22. From the MAE and RMSE values (which is calculated on the slope units which had an actual landslide percentage), the models have predicted the values considerably.

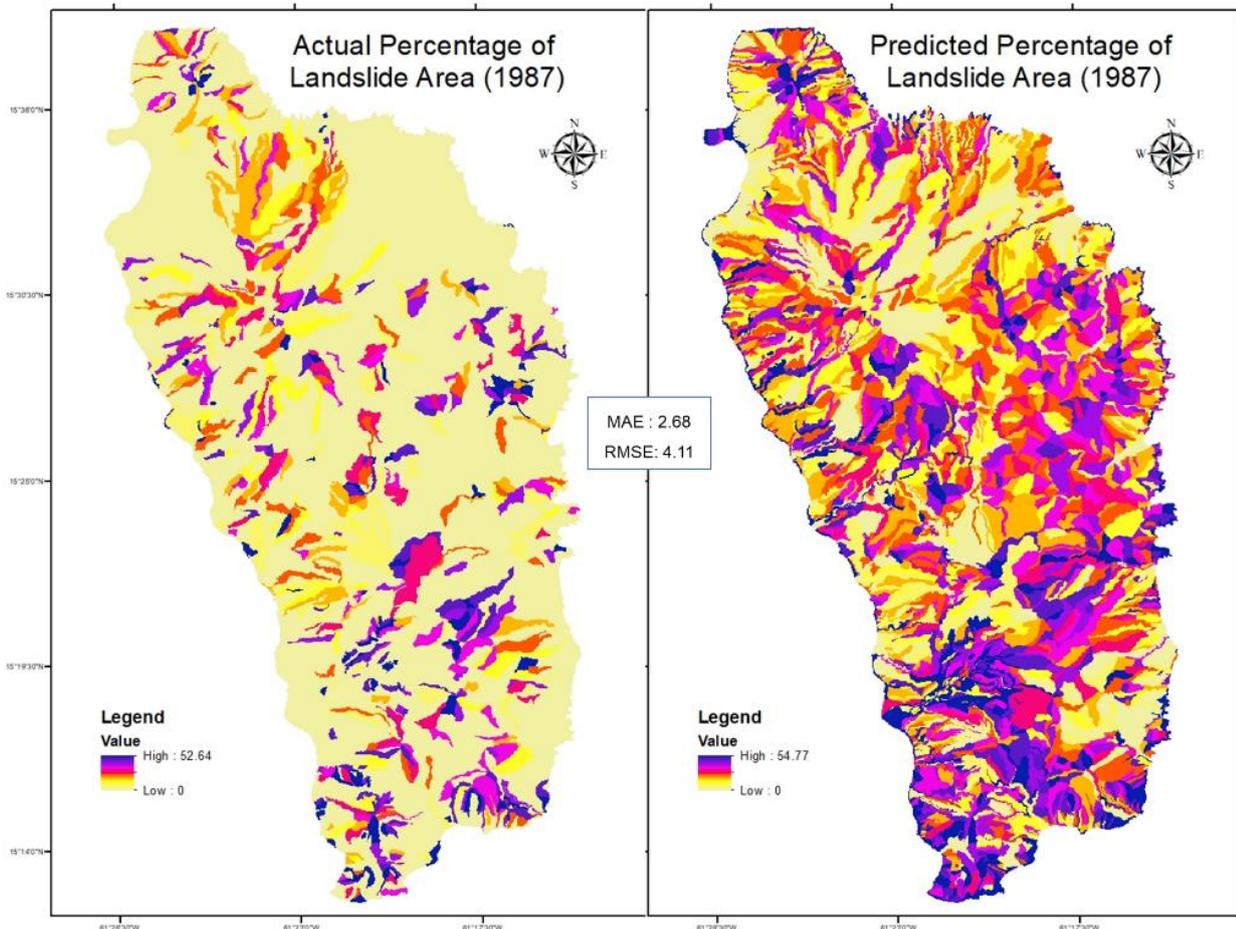


Figure 4.18 The Actual and the Predicted Percentages of the Landslide Areas of 1987

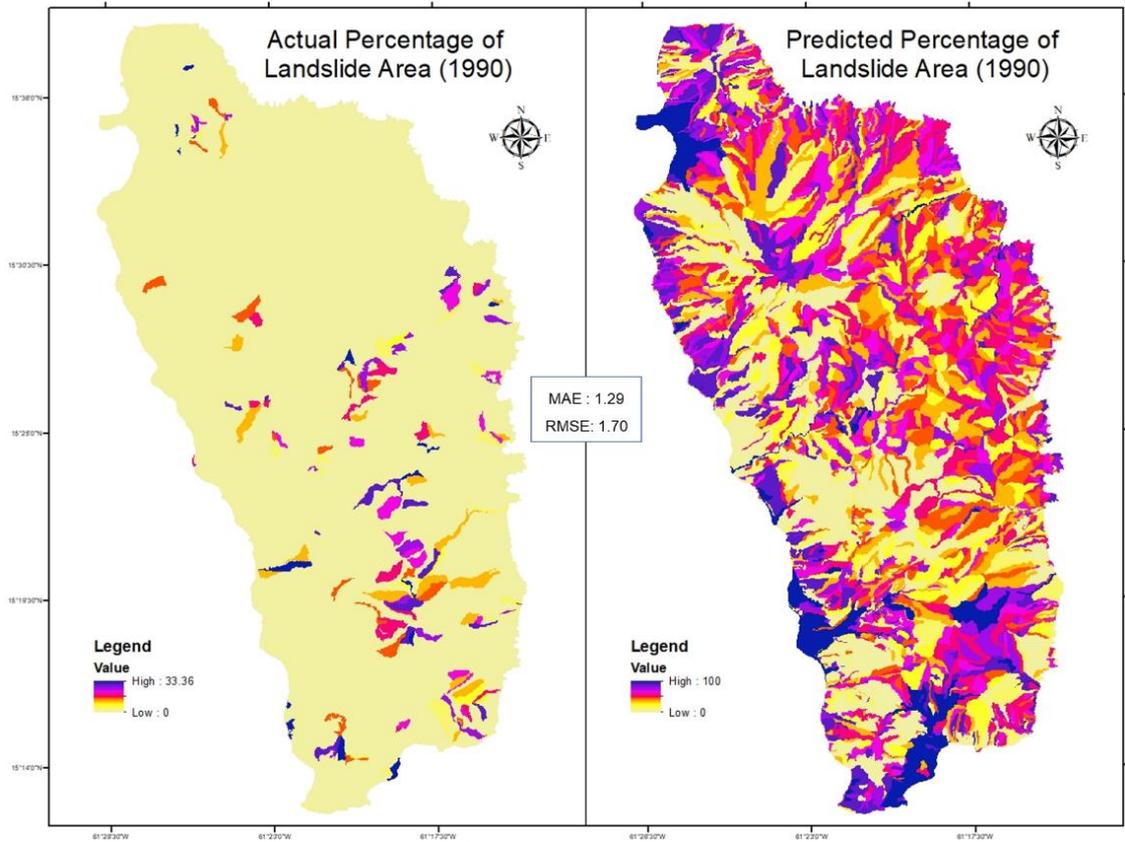


Figure 4.19 The Actual and the Predicted Percentages of the Landslide Areas of 1990

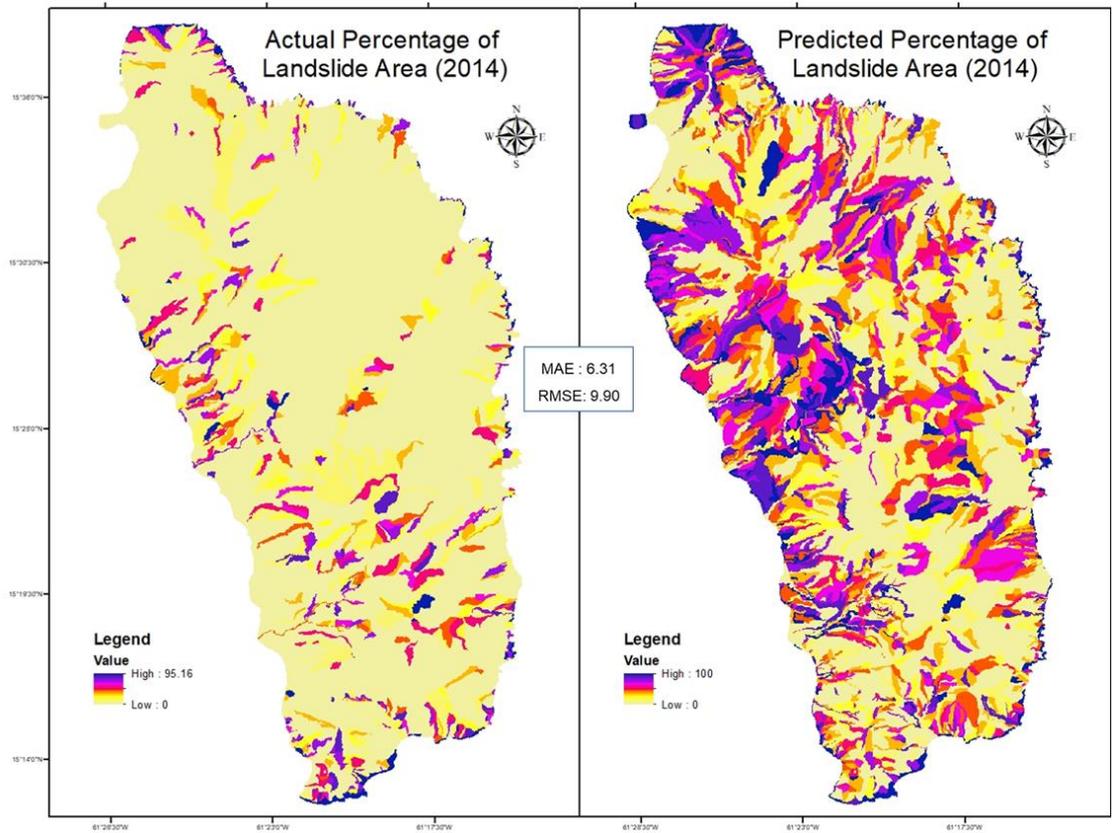


Figure 4.20 The Actual and the Predicted Percentages of the Landslide Areas of 2014

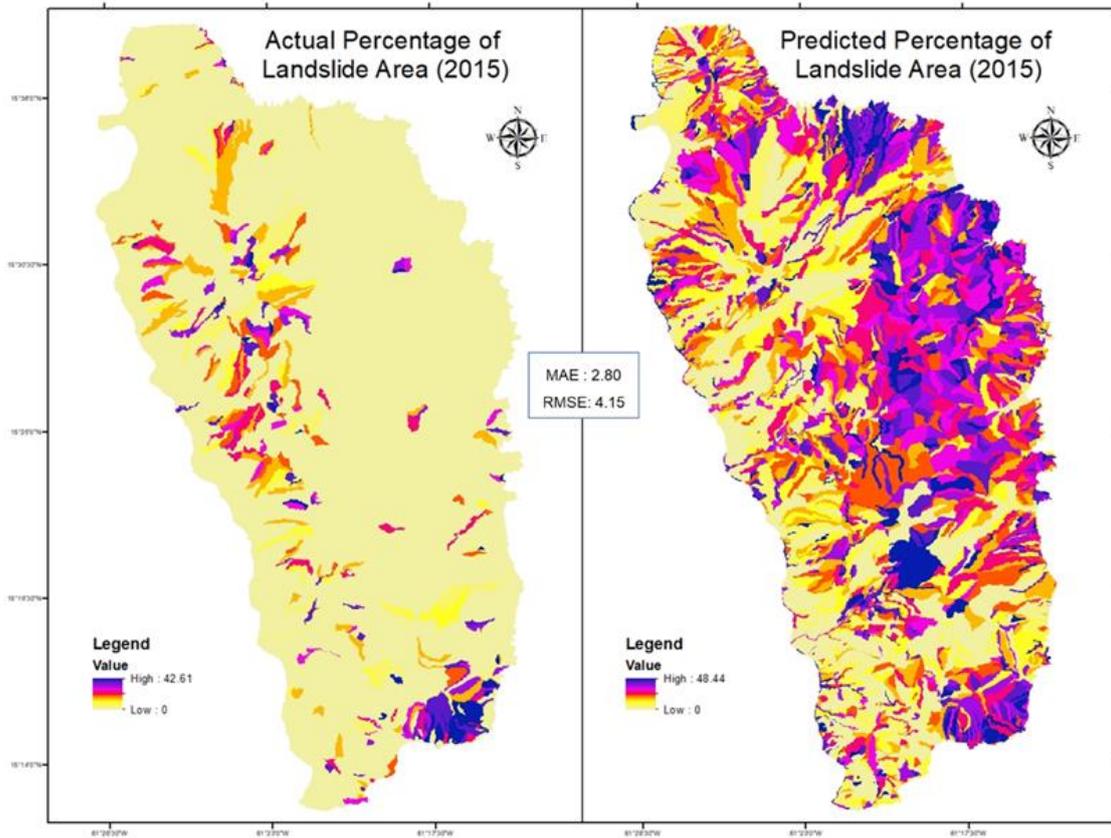


Figure 4.21 The Actual and the Predicted Percentages of the Landslide Areas of 2015

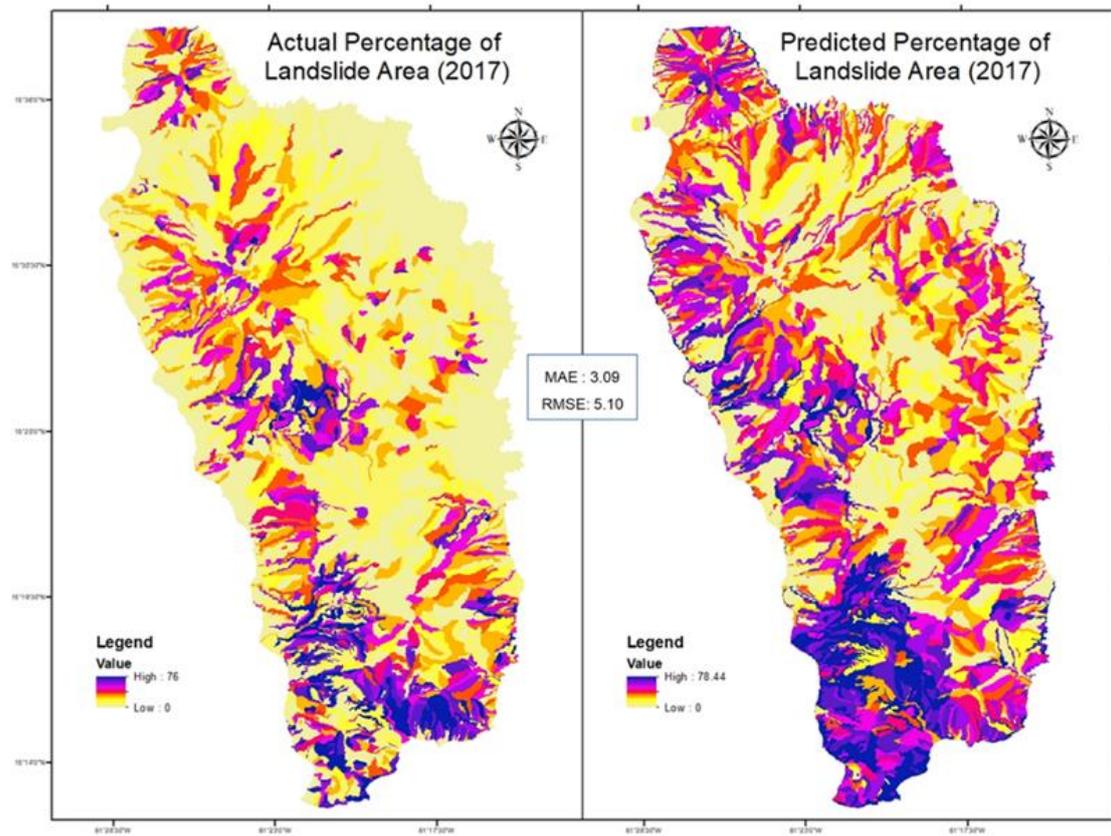


Figure 4.22 The Actual and the Predicted Percentages of the Landslide Areas of 2017

While the above maps included the predictions for the slope units with zero percentages, the scatterplot between the actual and predicted was made for the slope units which had an actual landslide area percentage (Figure 4.23). The shown relations indicate that the r-squared of models 2017, 2015 and 1987 had a low level of variance explained, whereas models 2014 and 1990 were able to explain the variance slightly better. While all the models have a satisfactory Pearson correlation coefficient, the models of 2014 and 1990 have a significantly higher correlation.

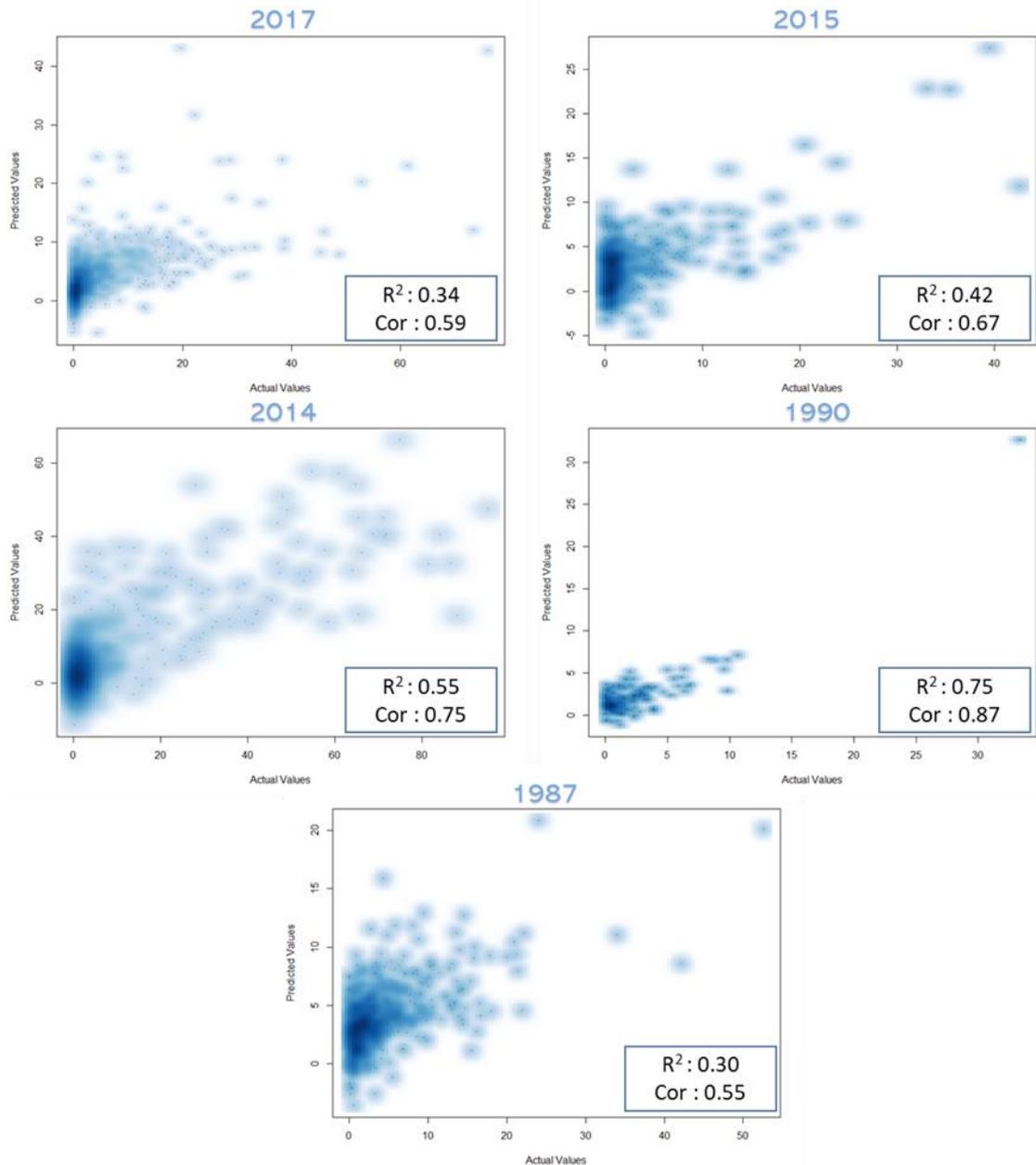


Figure 4.23 The Scatterplots between the Actual and the Predicted Percentage of Landslide Area per Slope Unit

The Pearson correlation coefficients denote whether the increase/decrease in the actual value is reflected on the predicted value. Whereas the r-squared informs on how much variability can be captured. All the models are able to mimic the increase/decrease but when it comes to capturing the variability, they are still 50% on average. These moderate level of values of the r-squared could be due to the data dependency of the model on the spatial characteristics of the conditioning factors with respect to the landslide area in the particular slope unit.

## 5. CONCLUSIONS AND RECOMMENDATIONS

From the initial phase of the study where, a statistical analysis was carried out on the relevance of the different conditioning factors like the morphometric parameters, geology type, soil type and the physical parameters, it was observed that many physical parameters have less influence when it comes to the landslide susceptibility modelling. It is observed that the morphometric data driven parameters contribute far more than the physically properties. It is safe to say that few process-driven factors have an equal relation with the geomorphological DEM-derived factors, for e.g., the slope and relative slope position of geomorphometric and the Manning's  $n$  and saturated hydraulic conductivity of physically-based seemed to have a relatively high positive relevance when it comes to susceptibility mapping. But the majority of the significant physical parameters in the exclusive model have less significance in the combined model. Though, certain characteristics from the physical parameters were captured by the model which considered all the above-mentioned conditioning factors, the performance did not increase significantly in comparison to the model that considered only the morphometric parameters, geology type and soil type.

This disparity among the spatial characteristics of the estimated landslide susceptibility is likely due to the uniformity of the values of the physical parameters. The physical parameters as such were computed via empirical formulas and spatial interpolation from a very coarse sampling data. This spatial invariability over the region weakened the ability of the model to recognize the relationship of the conditioning factor on the landslide susceptibility (Belsley, 1993; Harrell, 2015). Another limitation of this study was the consideration of a small study area, this was done in order to accommodate the physical parameters that were simulated previously by van den Bout et al. (2018). But this further lessened the changes in spatial tendencies. This restricted the competency of the statistical framework to assess the relevance (Peduzzi et al., 1995; Harrell, 2015) of process-driven properties. For instance, in the dataset matrix that has been used for this, only the particular initiation of the landslide occurrences is considered. This means the dataset has only 0.24% of the pixels (822 out of 344,973) corresponding to the landslide presences. Therefore, when process-driven parameters are exceedingly smooth across space, the model may have little spatial variability to derive functional statistical relations. Conversely, the higher level of details of terrain attributes may have led to better estimates in the data-driven context.

From the above study it was clear that the spatial invariability was the major concern which led to the moderate level of performance. In order to overcome this, in the future studies, the statistical frameworks could be applied to a larger extent of area which would have some recognizable changes in the physical setting of the region, customarily decreasing the consistency of the physical parameters which was observed in this study, or, to access more detailed geotechnical and hydrological survey data across space. Another approach for procuring a better performing susceptibility model would be to focus on an extension of the covariate set. In this study, the dataset does not directly apply any climatological parameters, which should be included in the analysis (Crozier, 2010). Also, from the lidar data that is available, a land use classification mapping (Dubayah and Drake, 2000; Yan et al., 2015) could have been done which might have shown a significance in the susceptibility modelling.

In the second phase of this study, a statistical framework was tested for its ability to perceive the temporal dependency of the past landslide occurrences on the landslide susceptibility. The landslide susceptibility was mapped for the five time periods, but a relatively high uncertainty was observed as the models rely on the size of the dataset and spatial characteristics of the input parameters. When the trends on how the landslide susceptibility changes over time was analysed, no clear sign of temporal effects was identified for this region. In the region of Dominica, it is observed that the triggering events largely vary over space with respect to previous or subsequent hurricane realizations. For instance, the hurricanes that

triggered the landslides in 1987 hit the west and the south; in 1990 landslides were triggered in the central region; in 2014 the landslides recorded were in the coastal regions and the south; in 2015 the west of the island and the 2017 Maria was a severe event which affected most of the areas of the island. This spread of landslide occurrences made the model more sensitive to spatial trends rather than temporal effects if present at all.

As mentioned above, statistical frameworks are sensitive to the data, the landslide presences hardly followed a regular pattern over the years. In this study, the trend of the susceptibility which was modelled with respect to the presence/absence was checked for the particular slope unit. Maybe an analysis on the number of landslides that had occurred could have given a better insight on the temporal dependence. Also, since Dominica as such is a tropical region with varying climatic conditions there might be significant changes to the terrain over the period of 30 years. This study uses the spatial characteristics of the recent elevation data for estimating the landslide susceptibility for all the periods.

In the last phase of the study, a statistical framework was applied for predicting the percentage of landslide area per slope unit. For this the whole of the landslide was taken into consideration and some parameters like the topographic wetness index, plan curvature and the profile curvature, which had less significance in the previous landslide susceptibility model based on the landslide initiation, exhibited a high significance in the landslide depositional area analysis. A good accuracy had been observed via the mean absolute error and root mean square error, and, it can be said that the integration of the landslide area might increase the performance of the landslide susceptibility model. Rather than the presence/absence schema, the scarp delineation could be used.

The landslide area percentage models though reflected on the increase/ decrease of the percentages adequately, they were unable to efficiently capture the variance. This issue may be due to the choice of the model. In fact, the analysis on the percentages has been implemented in a GLM framework assuming that the percentage distribution behaves like a Gaussian process. This model choice has inevitably brought some issues. In fact, a Gaussian model may predict negative values or values more than 100 for the target variable we considered, although they should not exist for a property whose domain is constrained between 0 and 100. Here, we addressed this issue by converting all negative values to zero and all values greater than 100 to 100. Nevertheless, this is an approximation and further experiments are certainly required. For instance, models that are constrained to produce results between 0 and 100 already exist, e.g., beta regression (Ferrari and Cribari-Neto, 2004; Schmid et al., 2013). These could be implemented to better model the percentage of failed slope units.

Also, even if a beta regression model would perform well, a clear problem would still remain. In fact, the definition of hazard should contextually feature the prediction of “where”, “when” and “how destructive” a population of landslides may be. However, separately computing the susceptibility (be it purely spatial or spatio-temporal) and the percentage, will not address the requirement of contextually feature the information reported above. In fact, being the susceptibility and the percentage models built on the basis of the same covariate set, they cannot be combined (multiplied) by definition because they will not be independent from each other. Therefore, a valid solution to the problem could consist of running joint-probability models, where the spatial or spatio-temporal probability of landslide occurrence is directly linked, within the same model architecture, to the associated landslide size characteristic or percentage per slope unit.

## LIST OF REFERENCES

---

- Abella, E. A. C., & Van Westen, C. J. (2008). Qualitative landslide susceptibility assessment by multicriteria analysis: a case study from San Antonio del Sur, Guantánamo, Cuba. *Geomorphology*, 94(3-4), 453-466.
- Alvioli, M., & Baum, R. L. (2016). Parallelization of the TRIGRS model for rainfall-induced landslides using the message passing interface. *Environmental Modelling & Software*, 81, 122-135.
- Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F., & Guzzetti, F. (2016). Automatic delineation of geomorphological slope units with r. slopeunits v1. 0 and their optimization for landslide susceptibility modeling. *Geoscientific Model Development*, 9(11), 3975.
- Amato, G., Eisank, C., Castro-Camilo, D., & Lombardo, L. (2019). Accounting for covariate distributions in slope-unit-based landslide susceptibility models. A case study in the alpine environment. *Engineering geology*, 260, 105237.
- Anagnostopoulos, G. G., & Burlando, P. (2012). An Object-oriented computational framework for the simulation of variably saturated flow in soils, using a reduced complexity model. *Environmental modelling & software*, 38, 191-202.
- Anagnostopoulos, G. G., Faticchi, S., & Burlando, P. (2015). An advanced process-based distributed model for the investigation of rainfall-induced landslides: The effect of process representation and boundary conditions. *Water Resources Research*, 51(9), 7501-7523.
- Arcement, G. J., & Schneider, V. R. (1989). Guide for selecting Manning's roughness coefficients for natural channels and flood plains.
- Atkinson, P., Jiskoot, H., Massari, R., & Murray, T. (1998). Generalized linear modelling in geomorphology. *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Group*, 23(13), 1185-1195.
- Bakka, H., Rue, H., Fuglstad, G. A., Riebler, A., Bolin, D., Illian, J., ... & Lindgren, F. (2018). Spatial modeling with R-INLA: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(6), e1443.
- Bartarya, S. K., & Valdiya, K. S. (1989). Landslides and erosion in the catchment of the Gaula River, Kumaun Lesser Himalaya, India. *Mountain Research and Development*, 405-419.
- Baum, R. L., Savage, W. Z., & Godt, J. W. (2008). TRIGRS-A Fortran program for transient rainfall infiltration and grid-based regional slope-stability analysis, version 2.0 (No. 2008-1159). US Geological Survey.
- Belsley, D. A. (1993). Conditioning Diagnostics Collinearity and Weak Data in Regression. *JOURNAL-OPERATIONAL RESEARCH SOCIETY*, 44, 88-88.
- Beven, K. J., & Kirkby, M. J. (1979). A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Journal*, 24(1), 43-69.
- Blake, G. R., & Hartge, K. H. (1986). Bulk density. *Methods of soil analysis: Part 1 Physical and mineralogical methods*, 5, 363-375.
- Böhner, J., & Selige, T. (2006). Spatial prediction of soil attributes using terrain analysis and climate regionalisation.
- Borga, M., Dalla Fontana, G., Da Ros, D., & Marchi, L. (1998). Shallow landslide hazard assessment using a physically based model and digital elevation data. *Environmental geology*, 35(2-3), 81-88.
- Bout, B., Lombardo, L., van Westen, C. J., & Jetten, V. G. (2018). Integration of two-phase solid fluid equations in a catchment model for flashfloods, debris flows and shallow slope failures. *Environmental modelling & software*, 105, 1-16.
- Brabb, E. E. (1985). Innovative approaches to landslide hazard and risk mapping. In *International Landslide Symposium Proceedings, Toronto, Canada* (Vol. 1, pp. 17-22).
- Brenning, A. (2005). Spatial prediction models for landslide hazards: review, comparison and evaluation.
- Brenning, A. (2008). Statistical geocomputing combining R and SAGA: The example of landslide susceptibility analysis with generalized additive models. *Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie*, 19(23-32), 410.
- Camilo, D. C., Lombardo, L., Mai, P. M., Dou, J., & Huser, R. (2017). Handling high predictor dimensionality in slope-unit-based landslide susceptibility models through LASSO-penalized Generalized Linear Model. *Environmental modelling & software*, 97, 145-156.
- Campbell, R. H. (1973). *Isopleth map of landslide deposits, Point Dume Quadrangle, Los Angeles County, California; an experiment in generalizing and quantifying areal distribution of landslides* (No. 535). US Geological Survey.
- Canli, E., Thiebes, B., Petschko, H., & Glade, T. (2015). Comparing physically-based and statistical landslide susceptibility model outputs-a case study from Lower Austria. *Vienna, Austria*.

- Carrara, A. (1983). Multivariate models for landslide hazard evaluation. *Journal of the International Association for Mathematical Geology*, 15(3), 403-426.
- Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V., & Reichenbach, P. (1991). GIS techniques and statistical models in evaluating landslide hazard. *Earth surface processes and landforms*, 16(5), 427-445.
- Carrara, A., Cardinali, M., Guzzetti, F., & Reichenbach, P. (1995). GIS technology in mapping landslide hazard. In *Geographical information systems in assessing natural hazards* (pp. 135-175). Springer, Dordrecht.
- Central Intelligence Agency, U. S. (1990). Perry-Castañeda Map Collection: Maps of the Americas. Retrieved from <https://legacy.lib.utexas.edu/maps/americas.html>
- Chacón, J., Irigaray, C., Fernandez, T., & El Hamdouni, R. (2006). Engineering geology maps: landslides and geographical information systems. *Bulletin of Engineering Geology and the Environment*, 65(4), 341-411.
- Chung, C. J. F., Fabbri, A. G., & Van Westen, C. J. (1995). Multivariate regression analysis for landslide hazard zonation. In *Geographical information systems in assessing natural hazards* (pp. 107-133). Springer, Dordrecht.
- Crozier, M. J. (2010). Deciphering the effect of climate change on landslide activity: A review. *Geomorphology*, 124(3-4), 260-267.
- DeGraff, J. V. (1985). Using isopleth maps of landslide deposits as a tool in timber sale planning. *Bulletin of the Association of Engineering Geologists*, 22(4), 445-453.
- DeGraff, J. V. (1987). Landslide hazard on Dominica, West Indies-Final Report. *Organization of American States, Washington, DC*.
- DeGraff, J. V. (1990). Post-1987 Landslides on Dominica. *West Indies: An Assessment of Landslide Hazard Map Reliability and Initial Evaluation of Vegetation Effect on Slope Stability, Submitted to The Commonwealth of Dominica and the O.A.S.*
- DeGraff, J. V. (1991). Determining the significance of landslide activity: Examples from the Eastern Caribbean. *Caribbean Geography*, 3(1), 29.
- Devoli, G., De Blasio, F. V., Elverhøi, A., & Hoeg, K. (2009). Statistical analysis of landslide events in Central America and their run-out distance. *Geotechnical and Geological Engineering*, 27(1), 23-42.
- Dominica | CHARIM. (n.d.). Retrieved October 24, 2019, from <http://www.charim.net/dominica/information>
- Dubayah, R. O., & Drake, J. B. (2000). Lidar remote sensing for forestry. *Journal of Forestry*, 98(6), 44-46.
- Fagerland, M. W., & Hosmer, D. W. (2012). A generalized Hosmer–Lemeshow goodness-of-fit test for multinomial logistic regression models. *The Stata Journal*, 12(3), 447-453.
- Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of applied statistics*, 31(7), 799-815.
- Freeman, T. G. (1991). Calculating catchment area with divergent flow based on a regular grid. *Computers & Geosciences*, 17(3), 413-422.
- Friedman, J., Hastie, T., & Tibshirani, R. (2009). glmnet: Lasso and elastic-net regularized generalized linear models. *R package version*, 1(4).
- Furlani, S., & Ninfo, A. (2015). Is the present the key to the future?. *Earth-Science Reviews*, 142, 38-46.
- Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F., & Reichenbach, P. (2008). Comparing landslide inventory maps. *Geomorphology*, 94(3-4), 268-289.
- Glade, T., & Crozier, M. J. (2005). Landslide hazard and risk: concluding comment and perspectives. *Landslide hazard and risk*. Wiley, Chichester, 767-774.
- Goetz, J. N., Guthrie, R. H., & Brenning, A. (2011). Integrating physical and empirical landslide susceptibility models using generalized additive models. *Geomorphology*, 129(3-4), 376-386.
- Goetz, J. N., Brenning, A., Petschko, H., & Leopold, P. (2015). Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. *Computers & geosciences*, 81, 1-11.
- Gökceoglu, C., & Aksoy, H. Ü. S. E. Y. İ. N. (1996). Landslide susceptibility mapping of the slopes in the residual soils of the Mengen region (Turkey) by deterministic stability analyses and image processing techniques. *Engineering Geology*, 44(1-4), 147-161.
- Gorsevski, P. V., Gessler, P. E., Foltz, R. B., & Elliot, W. J. (2006). Spatial prediction of landslide hazard using logistic regression and ROC analysis. *Transactions in GIS*, 10(3), 395-415.
- Guzzetti, F. (2006). Landslide hazard and risk assessment.

- Guzzetti, F., Carrara, A., Cardinali, M., & Reichenbach, P. (1999). Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology*, 31(1-4), 181-216.
- Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., & Ardizzone, F. (2005). Probabilistic landslide hazard assessment at the basin scale. *Geomorphology*, 72(1-4), 272-299.
- Hack, J. T. (1973). Stream-profile analysis and stream-gradient index. *Journal of Research of the us Geological Survey*, 1(4), 421-429.
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1), 29-36.
- Hansen, A. (1984). Landslide hazard analysis. *Slope instability*, 523-602.
- Hansen, A., Franks, C. A. M., Kirk, P. A., Brimicombe, A. J., & Tung, F. (1995). Application of GIS to hazard assessment, with particular reference to landslides in Hong Kong. In *Geographical Information Systems in assessing natural hazards* (pp. 273-298). Springer, Dordrecht.
- Harrell Jr, F. E. (2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer.
- Hastie, T., & Tibshirani, R. (1987). Generalized additive models: some applications. *Journal of the American Statistical Association*, 82(398), 371-386.
- Heerdegen, R. G., & Beran, M. A. (1982). Quantifying source areas through land surface curvature and shape. *Journal of Hydrology*, 57(3-4), 359-373.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression*, John Wiley & Sons. New York.
- Kogan, F. N. (1994). NOAA plays leadership role in developing satellite technology for drought watch. *Earth Observation Magazine*, 11, 1405-1409.
- Lang, D. M. (1967). Soil and land-use surveys No. 21. Dominica. *Soil and land-use surveys No. 21. Dominica*.
- Lindsay, J. M., Trumbull, R. B., & Siebel, W. (2005). Geochemistry and petrogenesis of late Pleistocene to Recent volcanism in Southern Dominica, Lesser Antilles. *Journal of Volcanology and Geothermal Research*, 148(3-4), 253-294.
- Lombardo, L., & Mai, P. M. (2018). Presenting logistic regression-based landslide susceptibility results. *Engineering geology*, 244, 14-24.
- Lombardo, L., Bachofer, F., Cama, M., Märker, M., & Rotigliano, E. (2016). Exploiting Maximum Entropy method and ASTER data for assessing debris flow and debris slide susceptibility for the Giampilieri catchment (north-eastern Sicily, Italy). *Earth Surface Processes and Landforms*, 41(12), 1776-1789.
- Lombardo, L., Opitz, T., & Huser, R. (2018). Point process-based modeling of multiple debris flow landslides using INLA: an application to the 2009 Messina disaster. *Stochastic environmental research and risk assessment*, 32(7), 2179-2198.
- Malamud, B. D., Turcotte, D. L., Guzzetti, F., & Reichenbach, P. (2004). Landslide inventories and their statistical properties. *Earth Surface Processes and Landforms*, 29(6), 687-711.
- Montgomery, D. R., & Dietrich, W. E. (1994). A physically based model for the topographic control on shallow landsliding. *Water resources research*, 30(4), 1153-1171.
- Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3), 370-384.
- Nilsen, T. H., & Brabb, E. E. (1977). Slope stability studies in the San Francisco Bay region, California. *Reviews in Engineering Geology-GSA*, 3, 235-243.
- O'Loughlin, C. L. (1984). Effectiveness of introduced forest vegetation for protection against landslides and erosion in New Zealand's steepplands. In *Symposium on effects of forest land use on erosion and slope stability* (pp. 275-280). University of Hawaii.
- Park, N. W., & Chi, K. H. (2008). Quantitative assessment of landslide susceptibility using high-resolution remote sensing data and a generalized additive model. *International Journal of Remote Sensing*, 29(1), 247-264.
- Peduzzi, P., Concato, J., Feinstein, A. R., & Holford, T. R. (1995). Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. *Journal of clinical epidemiology*, 48(12), 1503-1510.
- Peng, L., Niu, R., Huang, B., Wu, X., Zhao, Y., & Ye, R. (2014). Landslide susceptibility mapping based on rough set theory and support vector machines: A case of the Three Gorges area, China. *Geomorphology*, 204, 287-301.

- Pourghasemi, H. R., & Rahmati, O. (2018). Prediction of the landslide susceptibility: Which algorithm, which precision?. *Catena*, 162, 177-192.
- Pradhan, A. M. S., Kang, H. S., Lee, J. S., & Kim, Y. T. (2019). An ensemble landslide hazard model incorporating rainfall threshold for Mt. Umyeon, South Korea. *Bulletin of Engineering Geology and the Environment*, 78(1), 131-146.
- Refice, A., & Capolongo, D. (2002). Probabilistic modeling of uncertainties in earthquake-induced landslide hazard assessment. *Computers & Geosciences*, 28(6), 735-749.
- Reichenbach, P., Galli, M., Cardinali, M., Guzzetti, F., & Ardizzone, F. (2005). Geomorphologic mapping to assess landslide risk: concepts, methods and applications in the Umbria Region of central Italy. *Landslide Risk Assessment*. John Wiley, Chichester, 429-468.
- Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., & Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*, 180, 60-91.
- Rigon, R., Bertoldi, G., & Over, T. M. (2006). GEOTop: A distributed hydrological model with coupled water and energy budgets. *Journal of Hydrometeorology*, 7(3), 371-388.
- Romani, F., Lovell Jr, C. W., & Harr, M. E. (1972). Influence of progressive failure on slope stability. *Journal of Soil Mechanics & Foundations Div*, 98(Proc Paper 9349).
- Roobol, M. J., & Smith, A. L. (2004). Geologic Map of Dominica, West Indies: Geology Department, University of Puerto Rico at Mayaguez. Electronic document.
- Rouse, W. C., Reading, A. J., & Walsh, R. P. D. (1986). Volcanic soil properties in Dominica, West Indies. *Engineering Geology*, 23(1), 1-28.
- Samia, J., Temme, A., Bregt, A., Wallinga, J., Guzzetti, F., Ardizzone, F., & Rossi, M. (2017). Characterization and quantification of path dependency in landslide susceptibility. *Geomorphology*, 292, 16-24.
- Samia, J., Temme, A., Bregt, A. K., Wallinga, J., Stuiver, J., Guzzetti, F., ... & Rossi, M. (2018). Implementing landslide path dependency in landslide susceptibility modelling. *Landslides*, 15(11), 2129-2144.
- Schmid, M., Wickler, F., Maloney, K. O., Mitchell, R., Fenske, N., & Mayr, A. (2013). Boosted beta regression. *PLoS one*, 8(4), e61623.
- Simoni, S., Zanotti, F., Bertoldi, G., & Rigon, R. (2008). Modelling the probability of occurrence of shallow landslides and channelized debris flows using GEOTop-FS. *Hydrological Processes: An International Journal*, 22(4), 532-545.
- Stanić, B. (1984). Influence of drainage trenches on slope stability. *Journal of Geotechnical Engineering*, 110(11), 1624-1636.
- Stark, C. P., & Guzzetti, F. (2009). Landslide rupture and the probability distribution of mobilized debris volumes. *Journal of Geophysical Research: Earth Surface*, 114(F2).
- Team R (2014). R: A Language and Environment for Statistical Computing <http://www.R-project.org>.
- Terlien, M. T., Van Westen, C. J., & van Asch, T. W. (1995). Deterministic modelling in GIS-based landslide hazard assessment. In *Geographical information systems in assessing natural hazards* (pp. 57-77). Springer, Dordrecht.
- Terzaghi, K. (1962). Stability of steep slopes on hard unweathered rock. *Geotechnique*, 12(4), 251-270.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.
- UNITAR-UNOSAT (2015). Landslide Affected Areas in Southeastern Dominica. Product ID: 2281 – English, Published: 1 Oct, 2015, GLIDE: TC-2015-000119-DMA. [http://www.unitar.org/unosat/node/44/2281?utm\\_source=unosat-unitar&utm\\_medium=rss&utm\\_campaign=van den Bout, B., Emtehani, S., Brenner, F., Dibaba, M., Orr, A., Westen, C. J. V., & Jettens, V.](http://www.unitar.org/unosat/node/44/2281?utm_source=unosat-unitar&utm_medium=rss&utm_campaign=van%20den%20Bout,%20Emtehani,%20Brenner,%20Dibaba,%20Orr,%20Westen,%20C.%20J.%20V.,%20&utm_term=Jettens%20V) (2018). Physically-based multi-hazard modelling and its influence on predicted hazard and risk for Dominica. *AGUFM, 2018*, NH33D-1031.
- van Westen, C. J., Van Asch, T. W., & Soeters, R. (2006). Landslide hazard and risk zonation—why is it still so difficult?. *Bulletin of Engineering geology and the Environment*, 65(2), 167-184.
- Van Westen, C. J., Castellanos, E., & Kuriakose, S. L. (2008). Spatial data for landslide susceptibility, hazard, and vulnerability assessment: an overview. *Engineering geology*, 102(3-4), 112-131.
- Van Westen, C. J. (2016). National scale landslide susceptibility assessment for Dominica. *The World Bank Caribbean Handbook on Risk Information Management Technical Report*, 91.
- Varnes, D. J. (1984). *Landslide hazard zonation: a review of principles and practice* (No. 3).

- Walsh, R. P. D. (1982). A provisional survey of the effects of Hurricanes David and Frederic in 1979 on the terrestrial environment of Dominica, West Indies. *Swansea Geographer*, 19, 28-35.
- Waring, R. H., & Running, S. W. (2010). *Forest ecosystems: analysis at multiple scales*. Elsevier.
- Wu, W., & Sidle, R. C. (1995). A distributed slope stability model for steep forested basins. *Water resources research*, 31(8), 2097-2110.
- Yan, W. Y., Shaker, A., & El-Ashmawy, N. (2015). Urban land cover classification using airborne LiDAR data: A review. *Remote Sensing of Environment*, 158, 295-310.
- Yen, B. C. (Ed.). (1992). *Channel flow resistance: centennial of Manning's formula*. Water Resources Publication.
- Yilmaz, I., & Keskin, I. (2009). GIS based statistical and physical approaches to landslide susceptibility mapping (Sebinkarahisar, Turkey). *Bulletin of Engineering Geology and the Environment*, 68(4), 459-471.
- Zevenbergen, L. W., & Thorne, C. R. (1987). Quantitative analysis of land surface topography. *Earth surface processes and landforms*, 12(1), 47-56.
- Zhang, M., Cao, X., Peng, L., & Niu, R. (2016). Landslide susceptibility mapping based on global and local logistic regression models in Three Gorges Reservoir area, China. *Environmental Earth Sciences*, 75(11), 958.
- Zhou, G., Esaki, T., Mitani, Y., Xie, M., & Mori, J. (2003). Spatial probabilistic modeling of slope failure using an integrated GIS Monte Carlo simulation approach. *Engineering Geology*, 68(3-4), 373-386.
- Ziegler, A. D., Giambelluca, T. W., Tran, L. T., Vana, T. T., Nullet, M. A., Fox, J., ... & Evett, S. (2004). Hydrological consequences of landscape fragmentation in mountainous northern Vietnam: evidence of accelerated overland flow generation. *Journal of Hydrology*, 287(1-4), 124-146.
- Zimmermann, B., Elsenbeer, H., & De Moraes, J. M. (2006). The influence of land-use changes on soil hydraulic properties: implications for runoff generation. *Forest ecology and management*, 222(1-3), 29-38.

## APPENDIX I

### Geology Types of Grand Bay

- Geo 1: Young Pleistocene Volcanics
- Geo 2: Young Pleistocene Pelean Domes
- Geo 3: Ignimbrite on Young Pleistocene
- Geo 4: Young Pleistocene Craters
- Geo 5: Young Pleistocene Ignimbrites
- Geo 6: Pleistocene apron of block and ash
- Geo 7: Pleistocene Pelean Domes
- Geo 8: Recent River Gravel and Aluminium
- Geo 9: Ignimbrite on Pliocene Volcanics
- Geo 10: Pliocene Volcanics

### Soil Types of Grand Bay

- Soil 1: Allophanoid Latosolics
- Soil 2: Kandoid Latosolics
- Soil 3: Protosols
- Soil 4: Skeletal
- Soil 5: Young Soils
- Soil 6: Unclassified

### Geology Types of Dominica

- Geo 1: Block Ash flow on Young Pleistocene
- Geo 2: Block Ash on Pliocene
- Geo 3: Conglomerate and raised limestone
- Geo 4: Ignimbrite on Old Pleistocene
- Geo 5: Ignimbrite on Pliocene volcanic
- Geo 6: Ignimbrite on Young Pleistocene material
- Geo 7: Ignimbrites Block Ash Pliocene
- Geo 8: Miocene volcanics
- Geo 9: Older Pleistocene volcanics
- Geo 10: Pleistocene apron of block and ash
- Geo 11: Pleistocene Craters
- Geo 12: Pleistocene Pelean Domes
- Geo 13: Pliocene Craters
- Geo 14: Pliocene Pelean dome
- Geo 15: Pliocene volcanics
- Geo 16: Recent River gravel and alluvium
- Geo 17: Young Pleistocene Craters
- Geo 18: Young Pleistocene Ignimbrites
- Geo 19: Young Pleistocene Pelean Dome
- Geo 20: Young Pleistocene volcanics

## **Soil Types of Dominica**

Soil 1: Allophanoid Latosolics

Soil 2: Allophanoid podzolics

Soil 3: Beach Sand

Soil 4: Hydrogenic Group

Soil 5: Kandoidlatosolics

Soil 6: Kandoidlatosols

Soil 7: Other Clay Latosolics

Soil 8: Phytogenic Group

Soil 9: Pond

Soil 10: Protosols

Soil 11: Shingle

Soil 12: Skeletal

Soil 13: Smectoid Clay Soils

Soil 14: Soufriere

Soil 15: Unclassified

Soil 16: Young Soils

## APPENDIX II

Maps of the various conditioning factors used in the study for Grand Bay

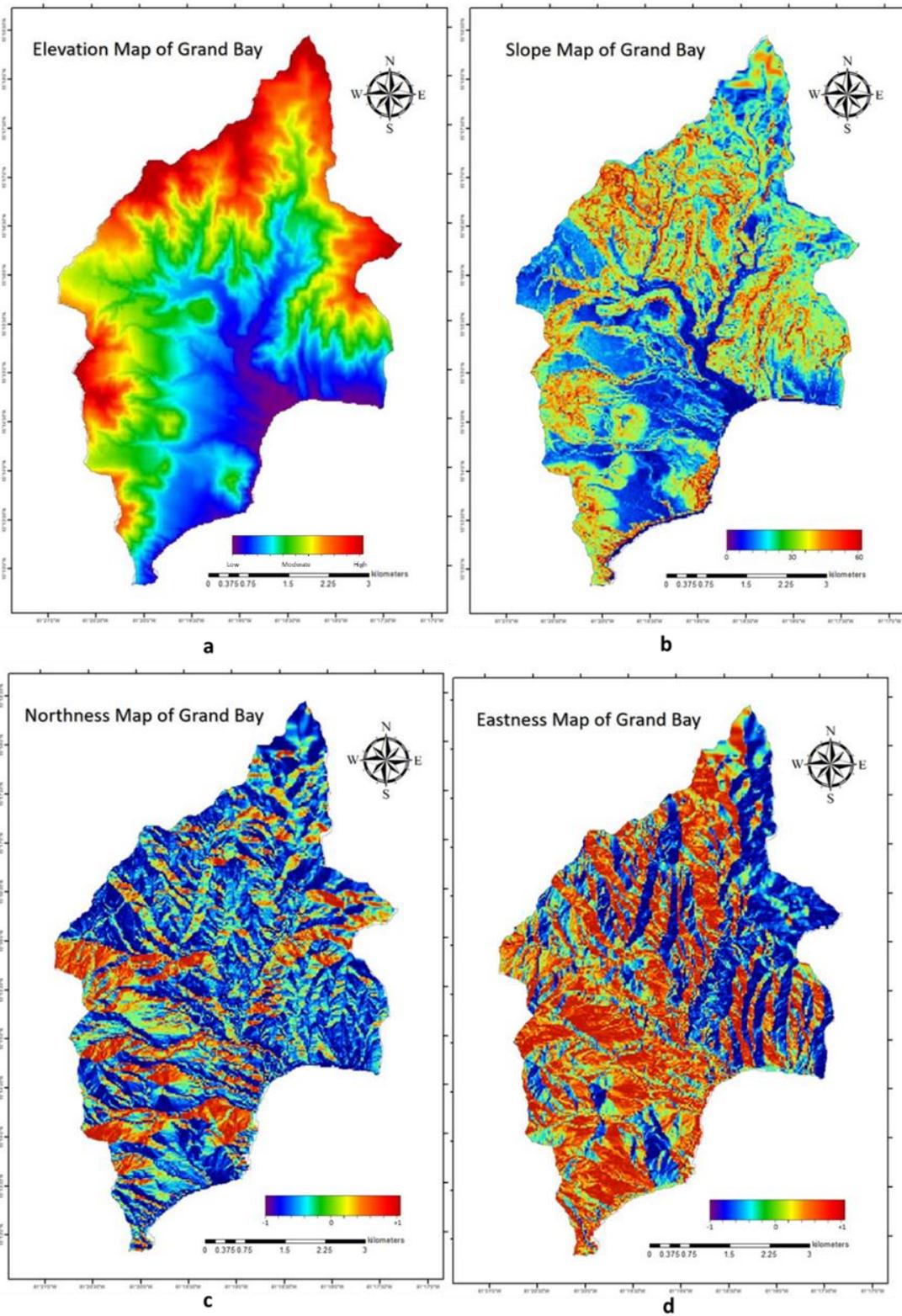


Figure 1. a) Elevation map; b) Slope map (in degrees); c) Northness map; d) Eastness map of Grand Bay

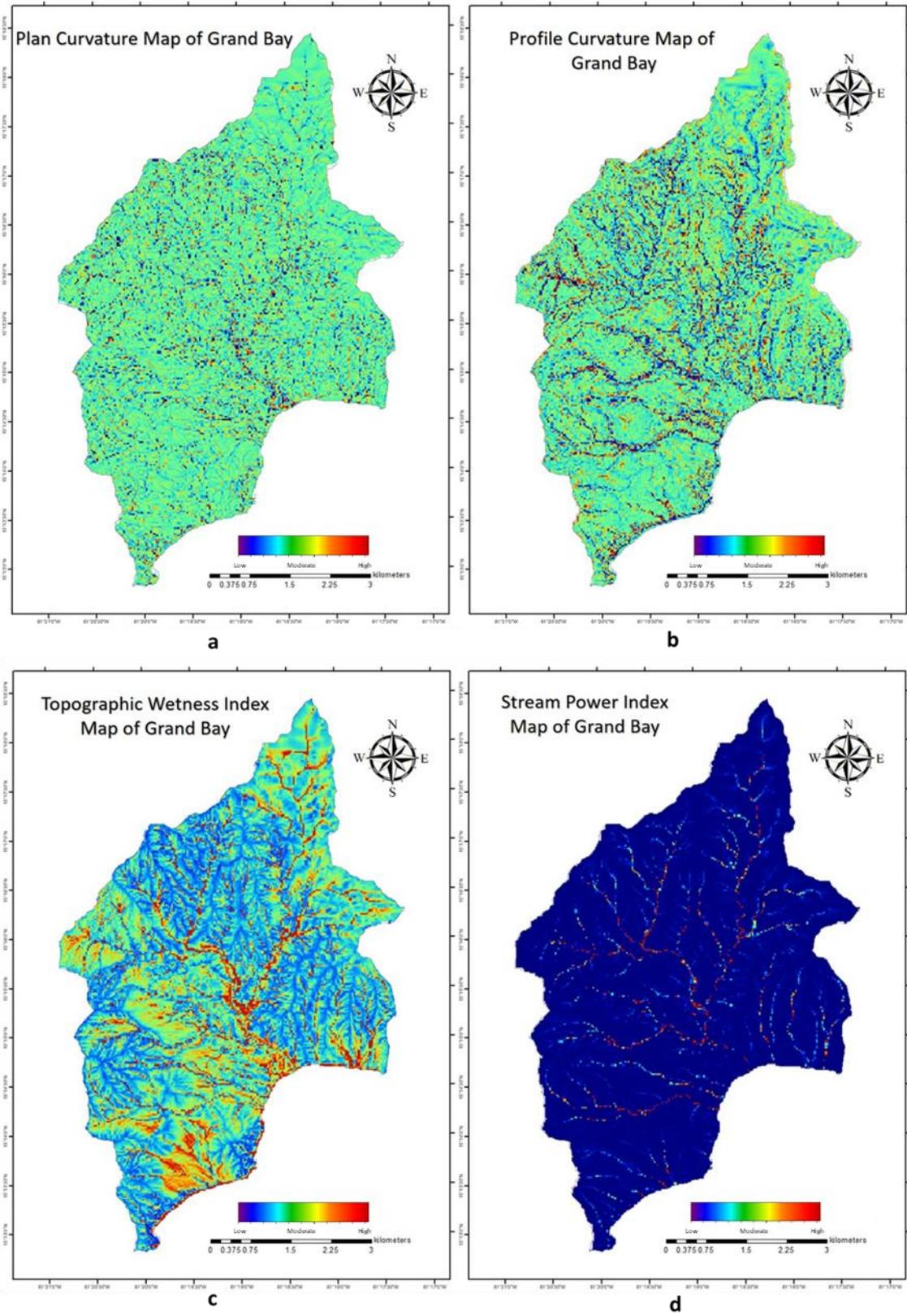


Figure 2. a) Plan curvature map; b) Profile curvature map; c) Topographic wetness index map; d) Stream power index map of Grand Bay

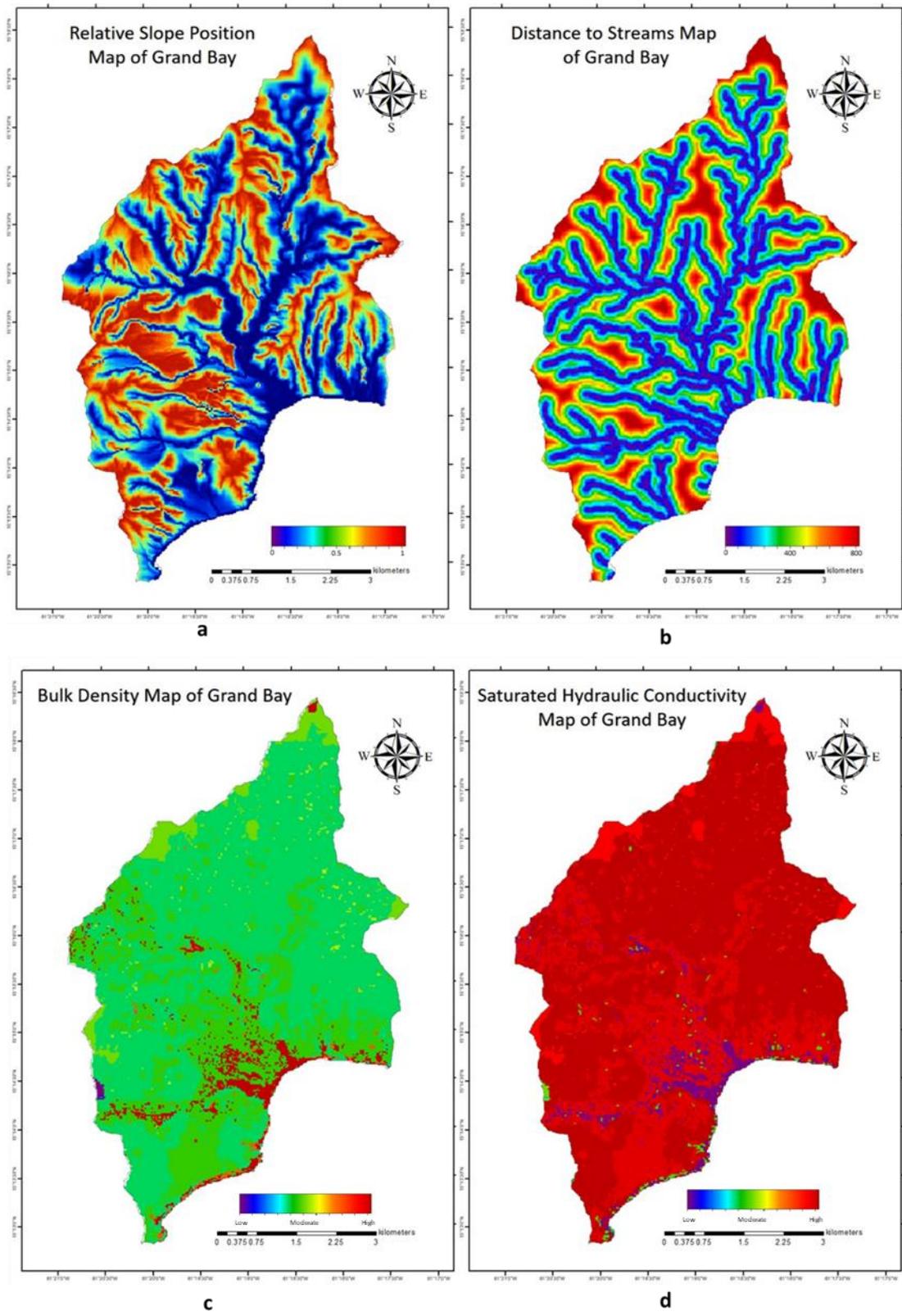


Figure 3. a) Relative slope position map; b) Distance to streams map (in meters); c) Bulk density map (in g/cm<sup>3</sup>); d) Saturated hydraulic conductivity map of Grand Bay

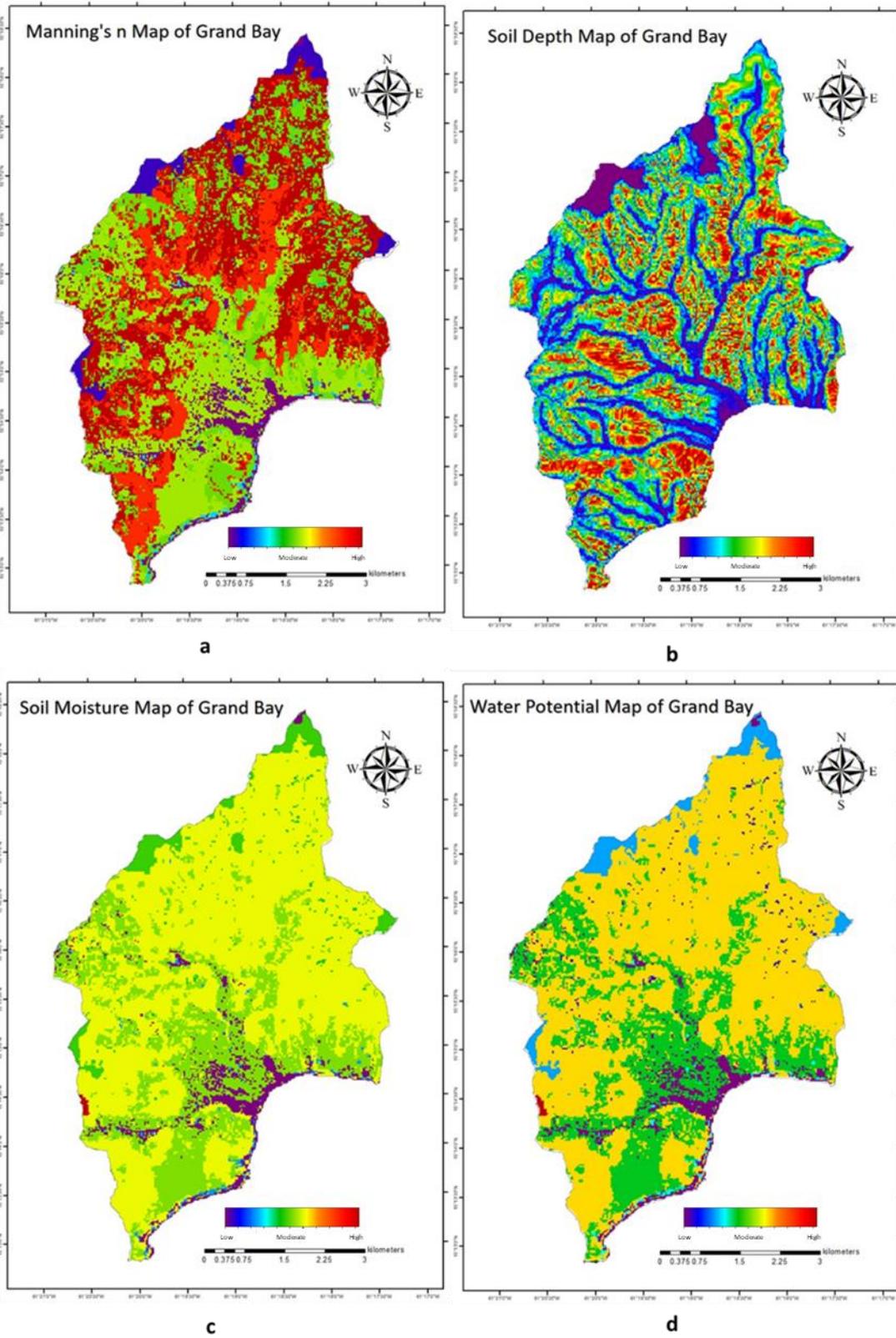


Figure 4. a) Manning's n map; b) Soil depth map; c) Soil moisture map; d) Water potential map of Grand Bay

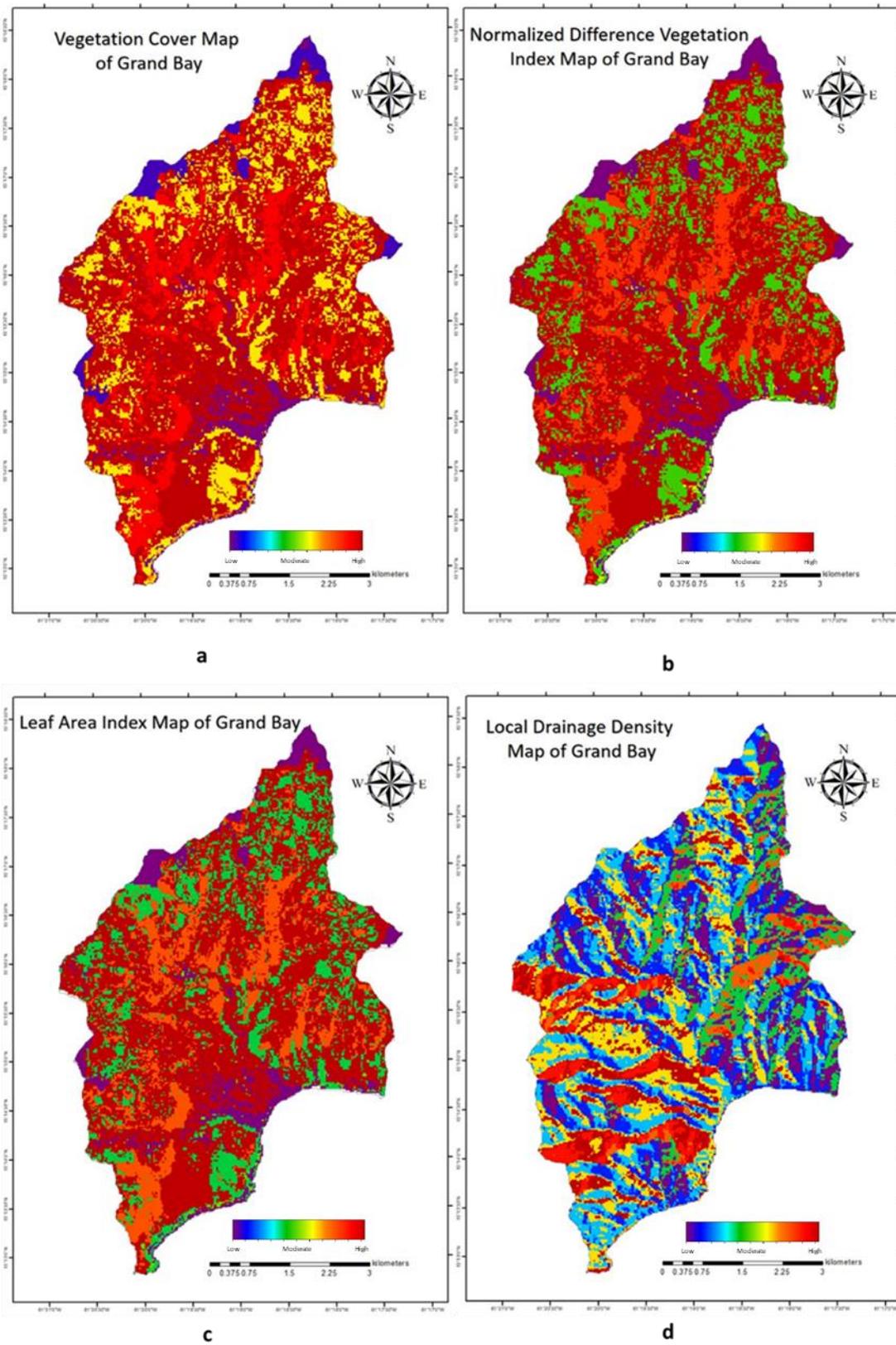


Figure 5. a) Vegetation cover map; b) Normalized difference vegetation index map; c) Leaf area index map; d) Local drainage density map of Grand Bay

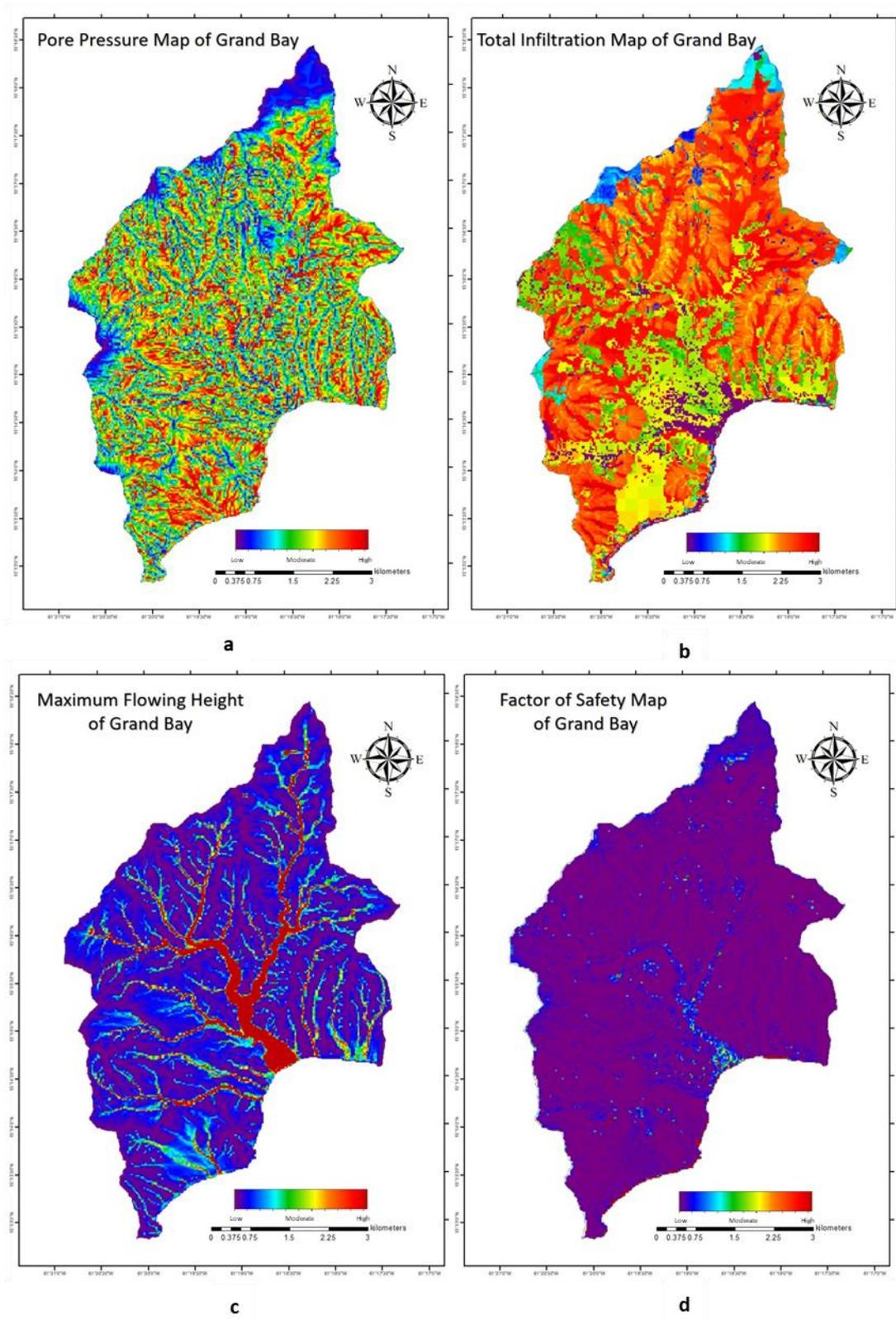
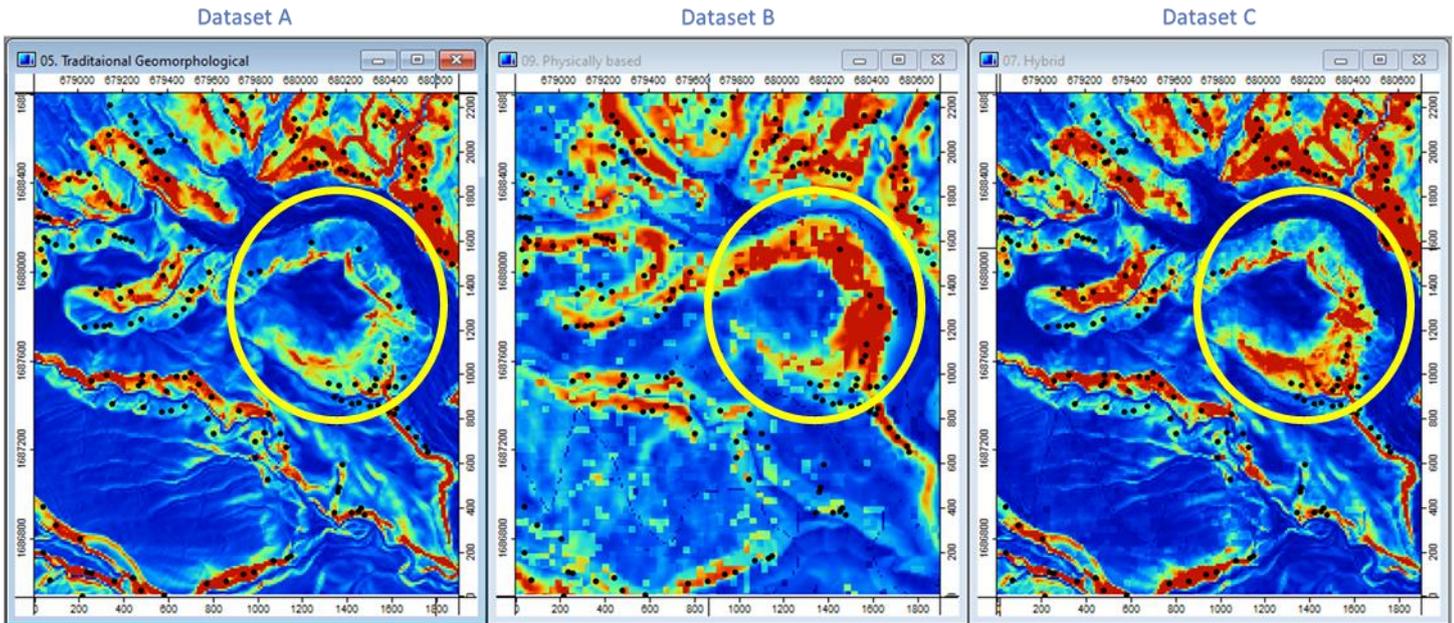


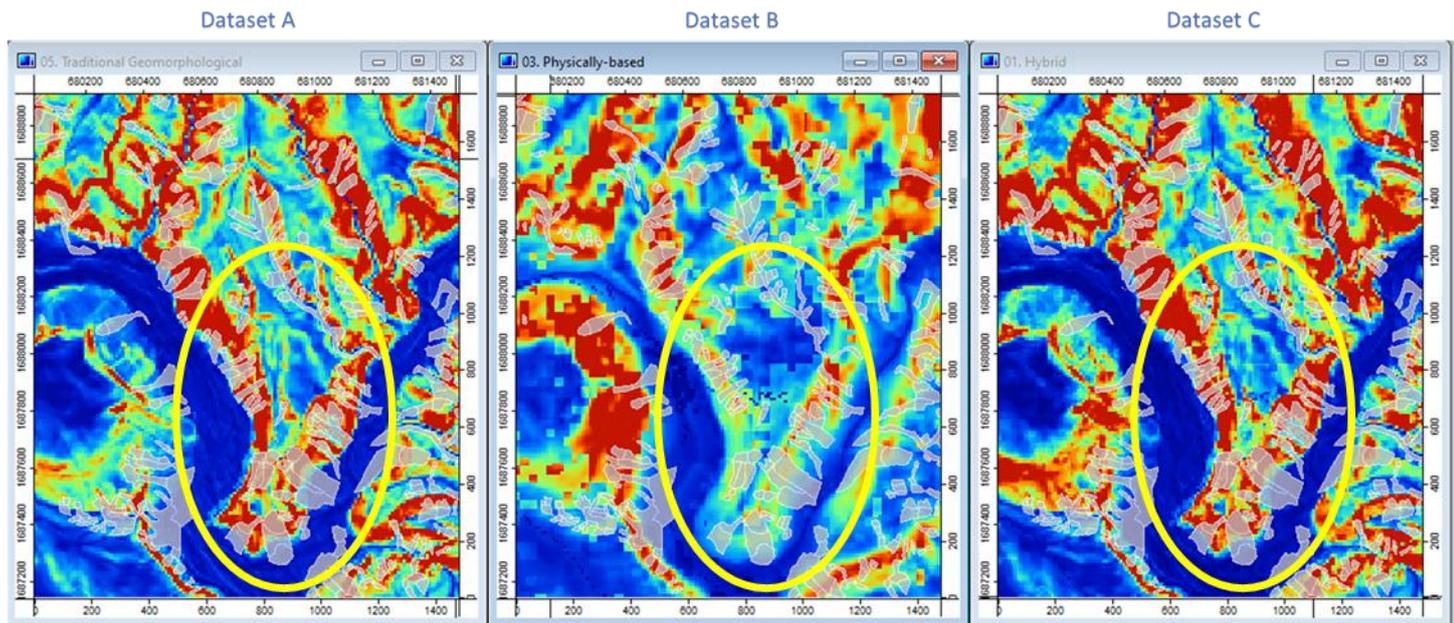
Figure 6. a) Pore pressure map; b) Total infiltration map; c) Maximum flowing height map; d) Factor of safety map of Grand Bay

### APPENDIX III

Scenarios where the model with dataset B performance was visually noticeable



An example of where the model with dataset B overestimates



An example of where the model with dataset B underestimates



In certain instances, the model with the dataset B has overestimated/underestimated the susceptibility. In the figures above, the black points represent landslide initiation and the polygons the landslides. It can also be observed that only little information is captured from dataset B in the model with dataset C.

APPENDIX IV

