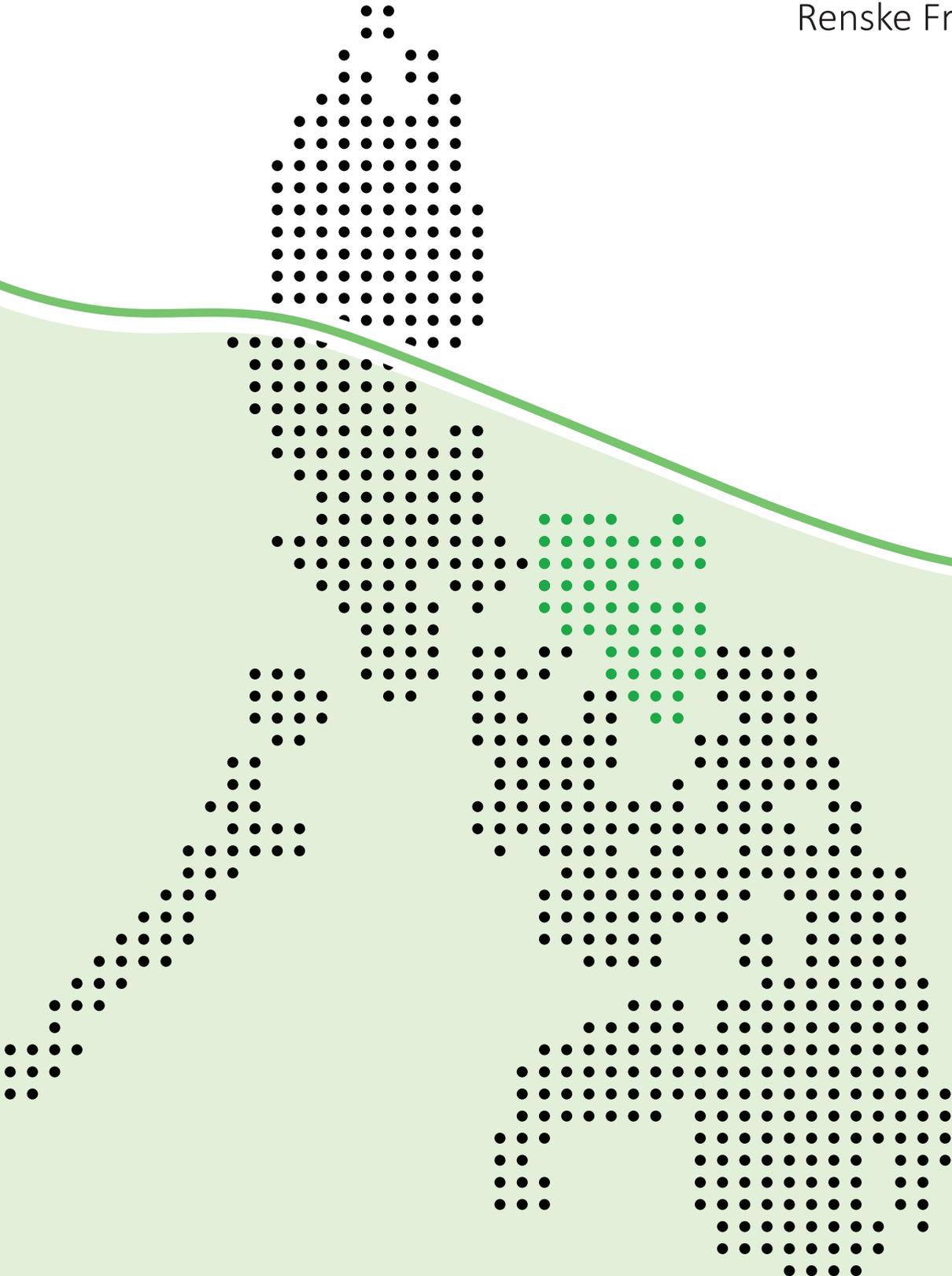


# Improving the reliability of an impact-based forecasting model

A case study for typhoons & landslides in the Philippines

Renske Free



# Improving the reliability of an impact-based forecasting model

A case study for typhoons & landslides in the Philippines

Renske Free

16-01-2022

To obtain the degree of Master of Science  
in Environmental Engineering  
with a specialization in Environmental Science.

To be defended publicly on January 24, 2022.

Student number: 4378024

Project duration: April 2021 – January 2022

Thesis committee:

Dr. T.A. Bogaard

Dr. M.J.C. van den Homberg (MBA)

Dr. F. Sperna Weiland

Dr. J.P. Aguilar-López

TU Delft, Chair

510, an initiative of the Netherlands Red Cross

Deltares

TU Delft

**Deltares**

 **TU Delft**

**510**



AN INITIATIVE OF  
THE NETHERLANDS  
RED CROSS

## ACKNOWLEDGEMENTS

---

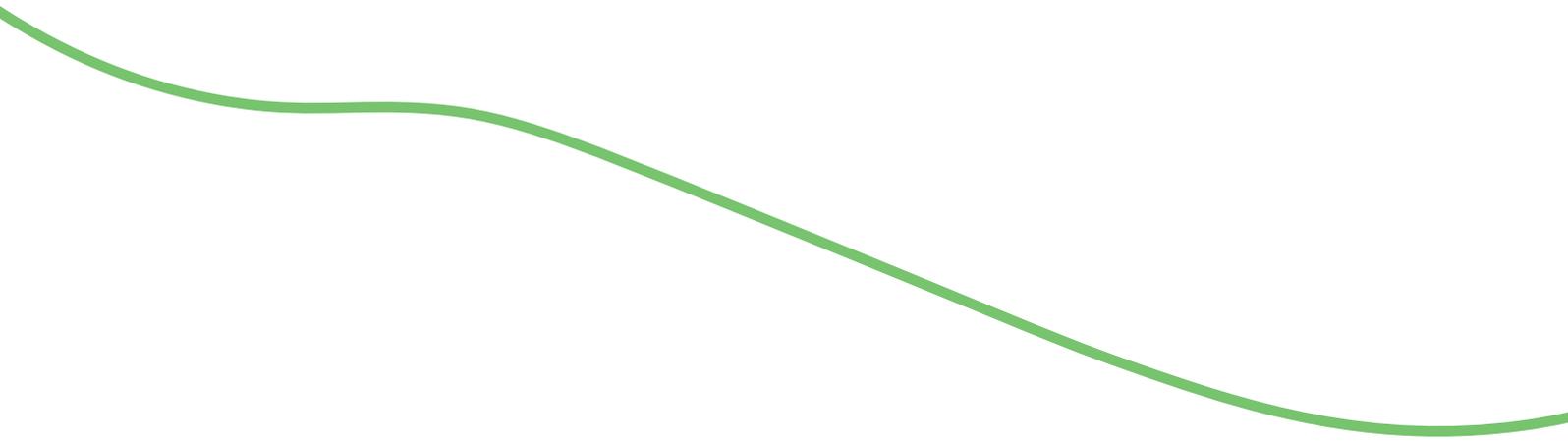
After being evacuated from an erupting volcano, I know what it is like to experience a natural hazard and face its impact. Forecasting damage can reduce disasters, yet people affected may have their own culture and traditions and will respond to the hazard and impact in their own ways. If it wasn't for the pandemic, I would have been able to include this social aspect in my thesis as well. As my thesis topic focused more on programming and using machine learning (ML), I can say I learned a lot and I was still able to work on natural hazards! I really hope that my thesis can contribute to both the landslide studies in Bicol and the typhoon impact-based forecasting model of 510!

My interests in disaster risk resilience, natural hazards and environmental engineering brought me to Thom half a year before the start of this thesis. You brought me together with Marc from 510 and Faraz of Deltares and gave me the opportunity to shape and define the topic of my thesis with them. I want to take this chance to express my gratitude to you, Thom, for setting up the committee and for talking to me weekly! You understood from the first meeting onward that my many interests and enthusiasm would result in the ever widening research scope. Thank you, Faraz, for guiding me while defining and finding the scope of my interest and thank you Frederiek for supporting me well during the last months of the thesis. Being able to go to Deltares weekly, even with the restrictions, was fantastic and I've had many nice talks and walks there. I would like to thank Marc for the brainstorm meetings, suggestions and constructive feedback, which really improved the quality of my work. Thank you also for getting me in touch with local humanitarian and Philippine organizations, such as the START network, professor Lagmay, the PRC, GRC and 510. I would like to thank Aki for being patient and mister Melillo for the help with CTRL-T.

On a more personal note, I want to thank the many sweet friends who cooked with/for me and walked with me during these past 10 months. Thank you Aussissies for the coffees and the cover of this thesis. I would like to thank Loek for helping in hacking. Thank you Gritclub, for making the lockdown life a little less endless. I want to thank the Environmental Maidon (and bois of course) and the Valencia girls for the coffees at the TU and Roffa and who made the master and thesis life really a lot of fun and cloudy in a good way :). Thanks to my roomies for the fun times at home. And of course, Lizzy, thanks for always sleeping in my window and being there every step of the way! Thank you, papa and mama, for shaping my thesis and helping me find the important parts that were missing. Thank you, Eefje, for returning my thesis in yellow! And finally thank you, Jos, for being with me while being abroad, for reading every version and for listening to me in the very excited (a lot) and less exciting moments. Who knows, we might visit the Bicol region soon together!

I have enjoyed writing this thesis and I now know much more about disaster risk resilience. Thanks to everyone who helped me to make this happen! Enjoy reading :)

*Renske Free*



## ABSTRACT

---

Anticipatory action requires models that can accurately and reliably predict the impact of natural hazards. However, impact forecasts are often underestimated when consecutive hazards are not considered. In the Bicol region in the Philippines, typhoons trigger 90% of landslides, causing a lot of fatalities and damage to infrastructure and agriculture. The lack of information on past landslide events has hampered the construction of landslide forecasting models. Currently, a machine learning (ML) impact-based forecasting (IBF) model for typhoons is operational in the Philippines. The model was developed by 510, an initiative of the Netherlands Red Cross. The model predicts impact due to the high wind speeds associated with typhoons and includes the possible impact due to landslides only via a static landslide susceptibility map. Hence, this study focused on extending the 510 typhoon model via hybrid modeling into a multi-hazard forecasting model for both typhoons and landslides to improve the forecast by considering impact from typhoon-induced landslides. The implementation of the hybrid multi-hazard impact-based forecasting model was tested on two typhoon events in the Bicol region.

A hydrometeorological landslide IBF model was successfully created, even with the limited data on landslide occurrences and rainfall available. The newly established regional event duration threshold for Bicol was applied on the case study events with an increased impact boundary of 300 km compared to the typhoon impact boundary of 100 km. The results of the hybrid multi-hazard model showed an improved impact forecast -compared to the model considering solely static input of landslides, which underestimated impact- in both location extent of the impact forecast and in accuracy: the True Positives doubled, whereas the False Negatives reduced by half. The separate landslide IBF model as an extension of the existing ML typhoon model provided additional benefits as these models can be decoupled to optimize the performance and reliability of both. This study resulted in the prototype of an impact-based multi-hazard model for typhoons and landslides for the Philippines and demonstrated the importance of considering impact from consecutive hazards.

**Keywords:** Landslide, typhoon, consecutive hazards, impact-based forecasting, rainfall, machine learning, Philippines

## DEFINITIONS

---

The following definition apply to this thesis and may differ from definitions used in other publications.

**Consecutive hazards:** Two or more hazards taking place in sequence within the same region and of which the impact overlaps. The first hazard alters the vulnerability of the region, increasing the susceptibility of the people or property for the secondary hazard, resulting in increased impact. This thesis also considers consecutive hazards with a triggering relation; the initial hazard causes the consecutive hazard(s) to occur (De Ruiter et al., 2020; Gill and Malamud, 2014; King, 2021).

**Coping capacity:** The ability to react to and reduce impact of hazards, such as institutions and infrastructure, the number of evacuation centres, health facilities and travel times to those institutions (Teklesadik and Riquet, 2021; UKMO, 2018).

**Data collection:** The collection of data for both the hazard forecast, i.e. sensors or satellites imagery, geographical and weather information, and impact forecast, i.e. socio-demographic information, and exposure, coping capacity and vulnerability data. The first component in an impact-based forecasting chain. Adapted from Zhang et al. (2019).

**Elements:** In this research; factors that influence the reliability and success of an individual component within the impact-based forecasting chain. A distinction is made between those which are essential for a component to succeed and those affecting the reliability of a component, the influential elements.

**Event duration threshold:** A type of rainfall threshold obtained from the accumulated rainfall and duration of specific events. Event duration (ED) thresholds are one of the most common methods to characterise a rainfall threshold for heavy rainfall events, such as typhoons. An ED threshold is the base line (lower bound) of rainfall conditions that will likely initiate a landslide event (Melillo et al., 2016, 2018; Zamudio and Orogo, 2021).

**Exposure:** The situation of people, infrastructure, buildings and other concrete human assets located in hazard-prone areas. Exposure refers to who and what might be affected in an area when a hazard occurs. If population and economic resources were not exposed to a hazard, no impact would exist. Exposure is time and space dependent (CMA and GFDDR, 2016; IFRC, 2021; WMO, 2015; Wilkinson et al., 2018).

**Hazard:** A hazard is defined as a hydrometeorological-based, geophysical or human-induced element that poses a level of threat to life, property or the environment (WMO, 2015). In this thesis, the hazard of typhoons and landslides are considered.

**Hazard forecast:** The possibility of a hazard occurring in the near future (IFRC, 2021). By the aggregation of input sources, modelling and programming techniques a dynamic, spatially and temporally, hazard forecast can be created. A threshold is needed to identify those areas prone to the hazard (Teklesadik and Riquet, 2021; Wilkinson et al., 2018). The second component of the impact-based forecasting chain.

**Hybrid multi-hazard impact-based forecasting model:** In this research, this model combines a machine learning impact-based forecasting model for one hazard (typhoon) with a separate hydrometeorological impact forecasting model for a second/consecutive hazard (landslides). Hybrid refers to the extension of the existing machine learning model with a separate model for consecutive hazard events. The hybrid model creates a total impact forecast of both the primary and secondary hazards. In this research, shortened to *hybrid model*

**Impact:** The sum of the hazard, the vulnerability, coping capacity and the exposure (CMA and GFDDR, 2016; Teklesadik et al., 2022; Teklesadik and Riquet, 2021; WMO, 2015).

$$Impact = (Vulnerability \times Coping\ capacity) + Exposure + Hazard$$

**Impact forecast:** The aim of impact-based forecasting. The impact assessment combines the impact data from the data collection with the hazard forecast to predict the probability of impact occurring at a location (IFRC, 2021). The impact is calculated by summing hazard with vulnerability, coping capacity and exposure (Teklesadik and Riquet, 2021; Wilkinson et al., 2018). The third component in an impact-based forecasting chain.

**Impact inventory:** Data set containing information of the actual impact caused by a historical hazard. This information helps to determining the impact magnitude, location and trigger (IFRC, 2021). Impact inventories are essential to verify the methods used for the impact forecast (UKMO, 2018; WMO, 2015).

**Impact-based forecasting (IBF):** A forecasting method which considers the impact a disaster will make instead of only forecasting a hazard. It combines temporal and spatial assessments of the hazard probability of occurrence with vulnerability, coping capacity and exposure of people and property, to create an impact forecast. Forecasting impact instead of hazards results in a better translation of forecasts into anticipatory action, such that disaster risk resilience is enhanced (CMA and GFDDR, 2016; Gill and Malamud, 2014; IFRC, 2021; UKMO, 2018; WMO, 2015; Wilkinson et al., 2018)

**Impact-based forecasting chain:** A chain of 5 components, which forms the impact-based forecasting method and leads to the decision on the type and place of early actions. Component 1 starts with collecting forecast and impact data. Component 2 creates a hazard forecast and in component 3 impact is forecasted. Component 4 sends a warning if an impact threshold is exceeded. Component 5 ends the chain with response of the warning into early actions. Modified from Zhang et al. (2019).

**Impact-based forecasting model:** A model which applies the impact-based forecasting method to create an impact forecast. It combines the first three components of the impact-based forecasting chain: data collection and hazard forecast components to form an impact forecast. Modified from Zhang et al. (2019).

**Lag time:** Time required to complete the data collection of an impact-based forecasting model.

**Landslide:** Defined by Guzzetti et al. (2012) as *The movement of a mass of rock, debris, or earth down a slope, under the influence of gravity*. In this research, landslides are confined to landslides caused by rainfall from typhoon events.

**Landslide cause:** The underlying, often long term, change, event or consecutive events that occurred which prepares the slope to become unstable and fail. The cause can relate to longer timescales and larger spacial scales (Bogaard and Greco, 2016).

**Landslide trigger:** The last push (or drop of water) for a slope to become unstable. The trigger has a smaller and shorter spatial and temporal scale than the cause of a landslide. It can be difficult to distinguish cause from trigger (Bogaard and Greco, 2016; Bogaard and Greco, 2018). This thesis focuses on landslides triggered by the excess rainfall from typhoons.

**Landslide inventory:** Information on the landslide events in a particular area showing the locations and runout of landslides that occurred. A landslide inventory is a data set that may represent single or multiple events. The landslide inventory is essential in performing any assessment on the hazard or impact (Chacón et al., 2006; Van Westen et al., 2008). A reliable landslide inventory includes information on the location, timing, cause, trigger, type, runout and impact of a landslide (Corominas et al., 2014; Van Westen et al., 2006). A landslide inventory is essential for hazard or impact forecasting (Corominas et al., 2014)

**Landslide susceptibility map:** Map indicating where landslides may occur. It includes a spatial assessment of factors relating to the instability processes in order to determine zones of landslide-prone areas without any temporal inclusion (Chacón et al., 2006).

**Landslide hazard map:** A map showing the probability of landslides occurrences within a specified period of time within a given area. It includes both a spatial and temporal event. It should ideally show the chance that a landslide forms at a particular place and also the chance that a landslide from further up slope strikes that place (runout) (Chacón et al., 2006).

**Landslide impact map:** Map showing the expected landslide impact in the affected area. It combines the probability information from a landslide hazard map with vulnerability, exposure and coping capacity data (Chacón et al., 2006; Teklesadik and Riquet, 2021).

**Lead time:** The time between the start of an impact forecast until the occurrence of the event that is forecasted to happen (IFRC, 2021).

**Multi-hazard:** Defined by WMO; (2017) as *'The selection of multiple major hazards that the country faces, and the specific contexts where hazards may occur simultaneously, cascadingly or cumulatively over time, and taking into account the potential interrelated effects'*. In this research, typhoons and landslides are considered as the major hazards.

**Multi-hazard impact-based forecasting model:** An impact-based forecasting model applied to multiple hazards. It captures the inter-hazard relations to form a total impact forecast for multi-hazards (WMO; 2017; Zhang et al., 2019).

**Rainfall threshold:** In this research an the event duration threshold is applied.

**Reliability:** In this thesis, reliability implies that the flow of the components within the impact-based forecasting chain is sustained and is defined by *'The combined reliability of each individual component of the impact-based forecasting chain'* Thirugnanam et al. (2020). Reliability is confined to a spatial assessment and temporal reliability is not considered.

**Reliable forecast:** In this research, defined by the resemblance of a forecast with the actual impact regarding its accuracy and geographical location. The accuracy is quantified as a ratio of the number of times a forecast is true or false (by the F1 score). The reliability concerning the location indicates the spatial correctness and geographical extent of the forecast. The temporal component of reliability is not considered in this research.

**Typhoon:** A (tropical) cyclone is a general term referring to a revolving weather disturbance that develops in the tropics. It can be a Tropical Depression, a tropical storm, a severe tropical storm, a typhoon or a super typhoon depending on the maximum sustained winds near the center. Typhoons are those weather system where the sustained winds exceed 200 km/h for 1 minute sustained wind speed (PRC, 2019; Teklesadik and Riquet, 2021).

**Typhoon-induced landslides:** Landslides triggered by the excess rainfall from a typhoon and occurring during and and shortly after typhoon events (Acosta et al., 2016; Zamudio and Orogo, 2021). Also referred to as typhoon-triggered landslides or rainfall-induced landslides during typhoon events.

**Vulnerability:** The increase in susceptibility of exposed elements by physical, social, economic and environmental factors or processes, i. e. human beings, their livelihoods and property, to suffer adverse effects when affected by a hazard. Vulnerability is situation specific and interacts with the hazard to form risk. Therefore, vulnerability is dynamic, relative and time- and space dependent (CMA and GFDDR, 2016; IFRC, 2021; WMO, 2015; Wilkinson et al., 2018).

**510 typhoon model:** In this research, the machine learning impact-based forecasting model for typhoons as constructed by 510, an initiative of the Netherlands Red Cross.

# CONTENTS

---

|       |  |    |
|-------|--|----|
| 1     | INTRODUCTION   | 1  |
| 1.1   | Impact-based forecasting (IBF)                                 | 1  |
| 1.2   | Reliability of the IBF chain and model                         | 1  |
| 1.3   | Case study area: The Bicol region in the Philippines           | 2  |
| 1.4   | Research question  | 3  |
| 1.5   | Thesis outline   | 4  |
| 2     | RELIABLE FORECASTS FOR LANDSLIDES & TYPHOONS                   | 5  |
| 2.1   | The increased impact of landslides and typhoons                | 5  |
| 2.2   | Reliable impact forecasting for landslides                     | 6  |
| 2.2.1 | Data collection for landslide IBF                              | 7  |
| 2.2.2 | Landslide hazard forecast                                      | 8  |
| 2.2.3 | Landslide impact forecast                                      | 9  |
| 3     | METHOD: CREATION OF THE HYBRID MODEL                           | 10 |
| 3.1   | The 510 typhoon model  | 10 |
| 3.2   | Creating a hybrid model  | 11 |
| 3.2.1 | Step 1: Regional event duration threshold                      | 12 |
| 3.2.2 | Step 2: Hazard and impact forecast                             | 14 |
| 3.2.3 | Step 3: Reliability assessment                                 | 16 |
| 4     | RESULTS: THE HYBRID MODEL IN THE BICOL REGION                  | 17 |
| 4.1   | A regional event duration threshold for the Bicol region       | 17 |
| 4.2   | Reliability assessment of the hybrid model in the Bicol region | 18 |
| 4.2.1 | Actual impact of typhoons Durian and Sarika                    | 18 |
| 4.2.2 | The reliability of the landslide hazard forecast               | 19 |
| 4.2.3 | The reliability of the landslide impact forecast               | 20 |
| 4.2.4 | The reliability of the landslide IBF model                     | 22 |
| 4.2.5 | The reliability of the hybrid model                            | 23 |
| 5     | DISCUSSION: THE RELIABILITY OF THE HYBRID MODEL                | 25 |
| 5.1   | Data collection  | 25 |
| 5.1.1 | The 510 typhoon model and its data sources                     | 25 |
| 5.1.2 | The steep and lowered event duration threshold                 | 26 |
| 5.1.3 | The four values of accumulated rainfall                        | 27 |
| 5.2   | The reliability of the landslide IBF model                     | 27 |
| 5.2.1 | The landslide hazard forecast                                  | 27 |
| 5.2.2 | The landslide impact forecast                                  | 28 |
| 5.3   | The hybrid model   | 29 |
| 5.4   | The reliability assessment                                     | 30 |
| 5.4.1 | Temporal reliability   | 31 |
| 6     | CONCLUSION   | 32 |
| 7     | RECOMMENDATIONS & NEXT STEPS                                   | 34 |
| 7.1   | Improving the landslide IBF model                              | 34 |
| 7.2   | Improving the prototype of the hybrid multi-hazard IBF model   | 35 |
| 7.3   | Next steps and future research                                 | 36 |

|  |    |
|--|----|
| BIBLIOGRAPHY   | 37 |
| A APPENDIX A: RELIABILITY ELEMENTS                                 | 43 |
| B APPENDIX B: ADDITIONAL FILES FOR THE METHOD                      | 45 |
| B.1 Elaboration on the 510 typhoon model                           | 45 |
| B.1.1 The ML model   | 45 |
| B.1.2 The 510 typhoon forecasting model                            | 45 |
| B.2 Full size figures  | 46 |
| B.3 Landslide susceptibility map                                   | 50 |
| B.4 Impact boundary analysis                                       | 51 |
| B.5 Typhoons Durian and Sarika                                     | 52 |
| C APPENDIX C: ADDITIONAL FILES FOR THE RESULTS                     | 53 |
| C.1 Event duration threshold                                       | 53 |
| C.2 Rainfall patterns for typhoons Sarika and Durian               | 54 |
| C.3 Accuracy statistics for hazard and impact forecasts            | 56 |
| D CREATING AN IBF MODEL USING EXPERT-BASED JUDGEMENT               | 59 |
| D.1 Phase 1: Data preparation                                      | 60 |
| D.2 Phase 2: Monte Carlo - Analytical Hierarchy Process (MC - AHP) | 60 |
| D.2.1 The Analytical hierarchy Process (AHP)                       | 60 |
| D.2.2 Phase 3: Global Sensitivity Analysis                         | 62 |
| D.3 Summary of method  | 64 |

# 1

## INTRODUCTION

---

### 1.1 IMPACT-BASED FORECASTING (IBF)

Worldwide, the occurrence and impact of natural hazards is becoming more frequent and increasingly intense (De Ruiter et al., 2020; Wilkinson et al., 2018). Forecasting impact from natural hazards reduces damage, destruction and death. Anticipatory action and early warning became possible when forecasting information became available. Traditional early warning systems provide information on when and where a hazard will occur. This type of hazard forecasting does not include the impact a hazard will make, whereas impact-based forecasting (IBF) does (UKMO, 2018; WMO, 2015; Wilkinson et al., 2018). IBF combines temporal and spatial assessments of the hazard probability of occurrence with vulnerability, coping capacity and exposure of people and property, to create an impact forecast. This is desirable as it helps to predict the magnitude of the impact and the location of the area affected by the hazard, instead of predicting where a hazard may occur. Disasters will only occur if a hazard coincides with exposure and vulnerability, resulting in loss and damage. Therefore forecasting impact provides for a better translation of forecasts into anticipatory action, such that disaster risk resilience is enhanced (CMA and GFDDR, 2016; Gill and Malamud, 2014; UKMO, 2018; WMO, 2015; Wilkinson et al., 2018).

The combination of the physical hazard's likelihood, severity and impact forecast with socio-political factors and stakeholder involvement together provides all the information to make a decision on the anticipatory action strategy. The forecast and decision includes an uncertainty as it is a prediction of both a natural hazard and its impact. In the end, this decision remains a human action on which end-users must rely (Bierens et al., 2020; Guzzetti et al., 2012; UKMO, 2018; WMO, 2015; Wilkinson et al., 2018). Studies by Bierens et al. (2020), Van den Homberg et al. (2020), and WMO (2015) have shown that stakeholder acceptance of the decision can be improved if the reliability of an impact forecast is improved.

### 1.2 RELIABILITY OF THE IBF CHAIN AND MODEL

IBF consists out of a chain of components, which together lead to the decision of where and what kind of early action should be taken (Zhang et al., 2019). The IBF chain in figure 1.1 is modified from Zhang et al. (2019) and starts with collecting forecast, vulnerability and exposure data in component 1; to create the hazard and impact forecasts in component 2 and 3, respectively; a warning is send (component 4) when a certain impact threshold is exceeded; which results in response of the warning (component 5).

The IBF chain functions in a sequence or continued flow in which each component to the right depends on the components left from it. The reliability of each component should be maximized, as subsequent components are influenced by their predecessors. If one component does not work or is unreliable, the sequence cannot be sustained, resulting in an overall decrease in reliability and trustworthiness of the impact forecast and warning (Guzzetti et al., 2020; Thirugnanam et al., 2020; WMO; 2017; Zhang et al., 2019). The *reliability* in this research uses the definition of Thirugnanam et al. (2020) as ‘The combined reliability of each individual component of the chain’.

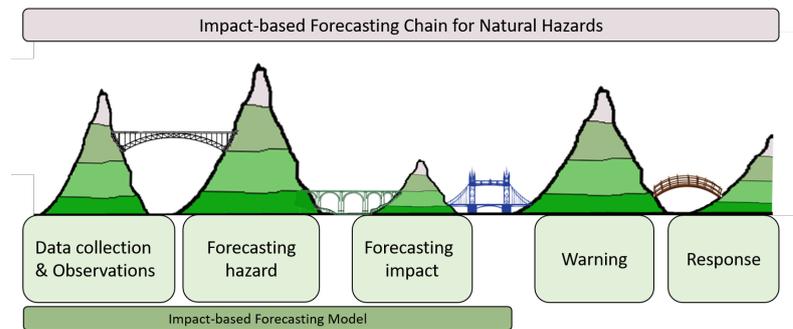


Figure 1.1: A schematic representation of an impact-based forecasting chain for natural hazards indicating the successive components. An impact-based forecasting model consists of the first three components of the chain. Modified from Guzzetti et al. (2020), WMO; (2017), and Zhang et al. (2019)

IBF often contains a data-driven model (Teklesadik et al., 2022; Thirugnanam et al., 2020), represented by the first three components, which results in an impact forecast. The aim of this model is to forecast the location and magnitude of impact as reliably as possible. In this research, an IBF model for typhoons is assessed with the goal of improving the reliability of the impact forecast.

### 1.3 CASE STUDY AREA: THE BICOL REGION IN THE PHILIPPINES

The Philippines is a disaster-prone country and natural hazards occur frequently. Typhoons are the most recurring hazard; around 20 typhoons come within reach of the country and 10 make landfall each year. Typhoons, also referred to as tropical cyclones which exceed a wind speed of 200 km/hr for 1 minute, are most common at the end of the year during the wet season. They bring heavy amounts of rainfall, even when the typhoon does not make landfall. Such rain events can cause or trigger landslides (Acosta et al., 2016; PRC, 2019; Teklesadik and Riquet, 2021; Zamudio and Orogo, 2021).

In the Bicol region in the Philippines (figure B.6) typhoons occur mostly in November and have serious impact due to the high number of inhabitants and the importance of the agriculture sector (PRC, 2019). Landslides in Bicol happen due its proneness for landslide hazard by the mountainous terrain, presence of active faults and its geographical location in an area where heavy rains, storms and typhoons are frequent (indicated in red in figure B.6). Landslides are closely related to typhoons due to their heavy rainfall and in the Bicol region, 90% of landslides are caused by typhoons (Acosta

et al., 2016; Start Network, 2021; Zamudio and Orogo, 2021). This research focuses on those landslides induced by the rainfall from typhoons, either during or shortly after landfall of the typhoon.

In March 2021, a landslide study was launched in the Bicol region by the Start Network in collaboration with the Philippines Red Cross (PRC), the IBF department of PRC, Humanitarian Inclusion, PLAN, ADRA and the German Red Cross. Their aim is to develop a model which assesses typhoon-induced landslides in the Bicol region. The Bicol region was chosen due to its susceptibility to landslides and the presence of various cities located in landslides-prone areas (indicated in figure B.6) (Start Network, 2021; Zamudio and Orogo, 2021).

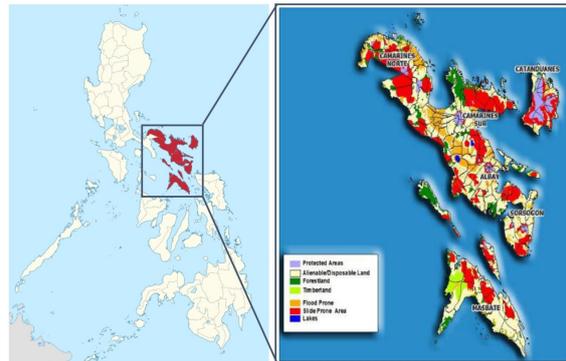


Figure 1.2: The Bicol region in the Philippines (left) and its susceptibility to hazards (right), with landslide prone areas in red. Modified from Zamudio and Orogo (2021).

510, an initiative of the Netherlands Red Cross, has developed and implemented a machine learning (ML) IBF model to forecast the impact of typhoons in the Bicol region. The model, from now on referred to as the 510 typhoon model, sends a warning to PRC when a threshold of percentage of damaged buildings within a municipality is exceeded (PRC, 2019). As most landslides in Bicol are triggered by typhoons (Zamudio and Orogo, 2021), the latter two projects may be combined. Hence, the aim of this research is to assess whether the reliability of the impact forecast of the 510 typhoon model improves if typhoon-induced landslides are taken into account.

#### 1.4 RESEARCH QUESTION

This thesis focuses on improving the reliability of a multi-hazard IBF model for natural hazards by assessing and expanding an existing ML model for typhoons in the Philippines with a hydrometeorological forecasting model for typhoon-induced landslides. The main research question of the thesis is defined as:

*Can the reliability of an existing typhoon impact-based forecasting model be improved by including the impact of typhoon-induced landslides?*

The following **sub-questions** (SQ) are defined to answer the main research question:

1. What is the relevance of consecutive hazards?
2. What influences the reliability of a landslide hazard and impact forecast?
3. How can a landslide impact forecast be included in an existing machine learning impact-based forecasting model for typhoons?
4. Does the reliability of an impact-based forecasting model improve when both typhoons and typhoon-induced landslides are considered?

This thesis results into four **deliverables**:

1. Assessment of the importance of consecutive hazards in impact-based forecasting models.
2. A regional event duration rainfall threshold for landslide initiation in the Bicol region of the Philippines.
3. A prototype of a hybrid multi-hazard impact-based forecasting model for typhoons and landslides.
4. Recommendations for improvements of the hybrid multi-hazard model.

## 1.5 THESIS OUTLINE

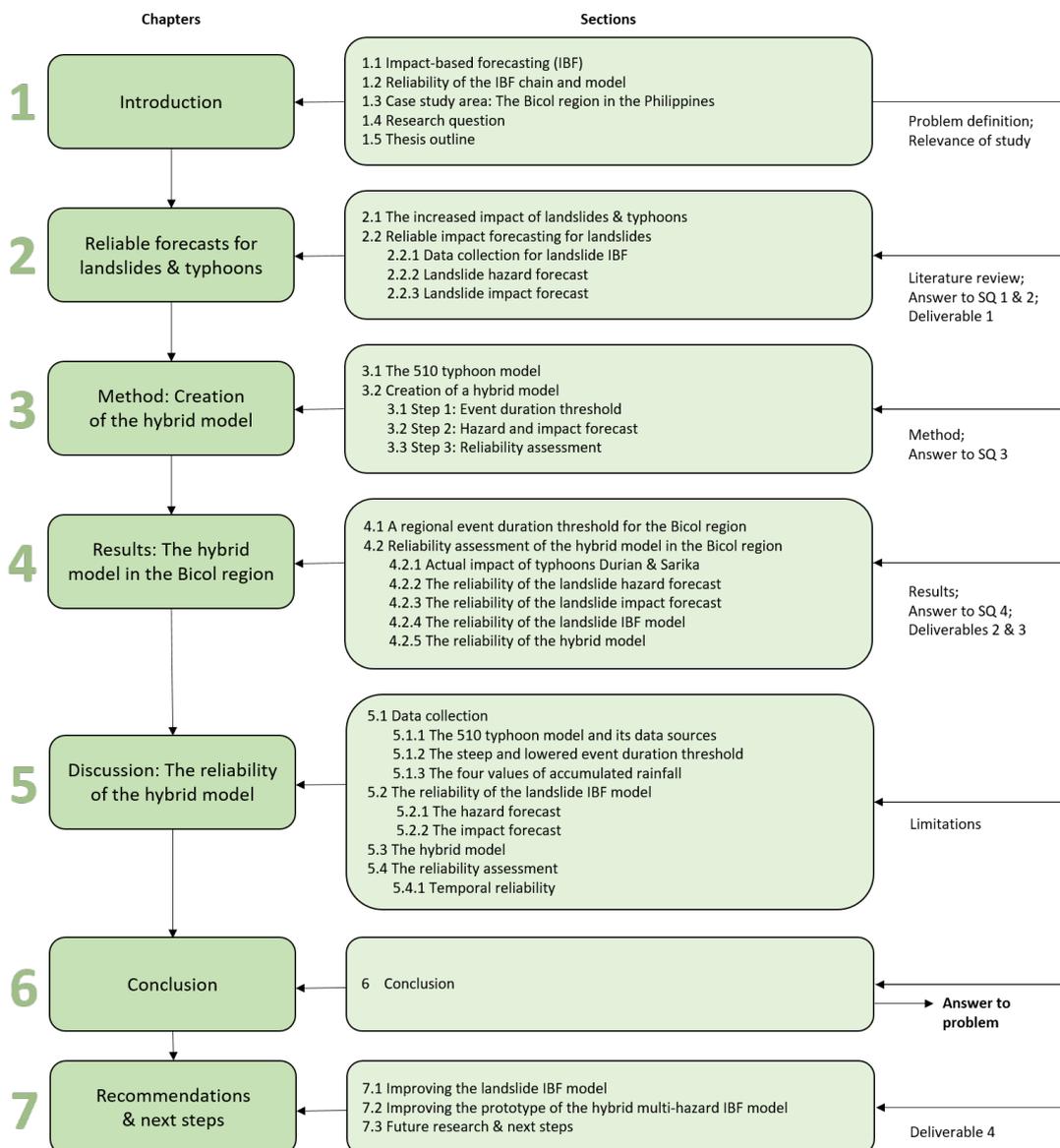


Figure 1.3: Thesis outline indicating the chapters, sections, sub-questions (SQ) and deliverables of the research.

## 2 RELIABLE FORECASTS FOR LANDSLIDES & TYPHOONS

The relevance of impact of consecutive hazards, with a focus on typhoons and typhoon-induced landslides, is discussed. Factors of influence on the reliability of a landslide impact forecast are assessed for the design of a landslide IBF model.

### 2.1 THE INCREASED IMPACT OF LANDSLIDES AND TYPHOONS

Natural hazards can have inter-hazard relations when they occur within the same area and time. A triggering relation is defined by a primary hazard causing or triggering one or multiple secondary hazards, resulting in a cascade of hazards (Cabrera et al., 2021; De Ruiter et al., 2020; Gill and Malamud, 2014; King, 2021). *Consecutive hazards* are two or more hazards taking place in sequence within the same region and of which the impact overlaps. The first hazard alters the vulnerability of the region, increasing the susceptibility of people and property for the secondary hazard, resulting in increased total impact. Primary and consecutive hazards should not be assessed independently as impact may be underestimated (Cabrera et al., 2021; De Ruiter et al., 2020; Gill and Malamud, 2014; King, 2021).

In this research, the two interacting hazards are typhoons and landslides. Landslides triggered by the rainfall of a typhoon have a wider area of range than the winds from the typhoon (IFRC, 2016a). Therefore the impact of landslides and typhoons combined occurs on a larger spatial extent compared to the impact of solely a typhoon (De Ruiter et al., 2020; Gill and Malamud, 2014; Segoni et al., 2018; Van Westen et al., 2006).

A typhoon alters the initiation conditions of the environment for landslides and increases the vulnerability of the region (Cabrera et al., 2021; De Ruiter et al., 2020). Too little time between subsequent hazards results in insufficient vulnerability recovery, schematically shown in figure 2.1 (adapted from De Ruiter et al. (2020)). Consecutive hazards change the vulnerability of a region from static to dynamic when measures, e.g. relocation and evacuation, are taken during the course of the hazard forecast and event (De Ruiter et al., 2020; Gill and Malamud, 2014; King, 2021). Increased vulnerability and exposure from hazard-interactions results in increased impact and possibly exceeds impact when hazards are assessed independently (De Ruiter et al., 2020).

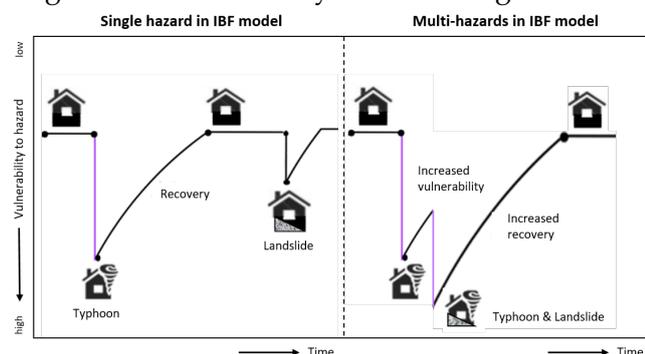


Figure 2.1: The increased vulnerability when considering consecutive hazards compared to single hazards, adapted from (De Ruiter et al., 2020)

Long term measures to reduce landslide hazard and impact include e. g. adaptation of land use planning strategies. This requires continued monitoring of landslides over the course of multiple years and involvement of various stakeholders for decision-making (Segoni et al., 2018). Short term measures to protect people and property, e. g. building local construction to slow down landslides, take time to be constructed. Within the limited time available prior to the typhoon, effective early actions for landslides other than evacuation and relocation are therefore difficult to implement. Landslide impact forecasts provide for a suitable and applicable measure to mitigate and reduce landslide impact (Segoni et al., 2018; TSR, 2007). The measures for landslides, thus, have another character from the early actions appropriate for typhoons (Cabrera et al., 2021; Segoni et al., 2018).

Due to the increased and dynamic form of vulnerability, exposure and impact; enlarged spatial extent; and difference in early action measures possible within the lead time, the severity of impact can be much more if a landslide occurs in combination with a typhoon. Hence, forecasting the impact of landslides and typhoons combined can result in a more reliable impact forecast (Cabrera et al., 2021; De Ruiter et al., 2020; Gill and Malamud, 2014; King, 2021).

As explained above, there is a need to include consecutive hazards into existing forecasting models (CMA and GFDDR, 2016; Cabrera et al., 2021; Georisk, 2021; Monteverde et al., 2020; WMO, 2017; Zamudio and Orogo, 2021). In the Philippines, where multi-hazards occur frequently, such models are required (Monteverde et al., 2020). The construction of a reliable landslide or multi-hazard model where landslides are included has been hampered by the lack of information on past landslide events, the hazard and impact (Cabrera et al., 2021; Zamudio and Orogo, 2021). Expanding forecasting models with secondary hazards can be beneficial as part of the data collections has been done for the primary hazard (Gill and Malamud, 2014). This can reduce lag time compared to creating two separate models. This study aims to create a multi-hazard typhoon and landslide IBF model for the Philippines by expanding the existing 510 typhoon model with a separate landslide model.

## 2.2 RELIABLE IMPACT FORECASTING FOR LANDSLIDES

A *reliable forecast* in this research is defined by the resemblance of a forecast regarding its accuracy and location with the actual impact. The accuracy is quantified in a ratio of the number of times a forecast is true or false. The location indicates the spatial correctness and geographical extent of the forecast. A reliable impact forecast of an IBF forecasting model is influenced by the reliability of each component (figure 1.1) and of influential elements of these components. Understanding what elements influence the reliability of an impact forecast helps to create reliable forecasting models, e. g. a typhoon-induced landslide IBF model.

A literature study on the subjects of early warning chains for natural hazards and rainfall-induced landslides, landslide hazard and impact assessments, and IBF was performed. It resulted in the selection of *reliability elements* of influence on the components of an IBF landslide model. These elements thus affect the reliability of an

impact forecast for landslides (Cabrera et al., 2021; Guzzetti et al., 2020; Joshi et al., 2020; Nadim and Intrieri, 2011; Segoni et al., 2018; Thirugnanam et al., 2020; UKMO, 2018; Van Westen et al., 2006; WMO, 2017; Waidyanatha, 2010; Wilkinson et al., 2018; Zhang et al., 2019). Appendix A includes the definition of the reliability elements per component. Essential elements are required for a component to be successful and influential elements affect the reliability of a component. The latter are considered for the expansion of the 510 typhoon model with a reliable landslide IBF model.

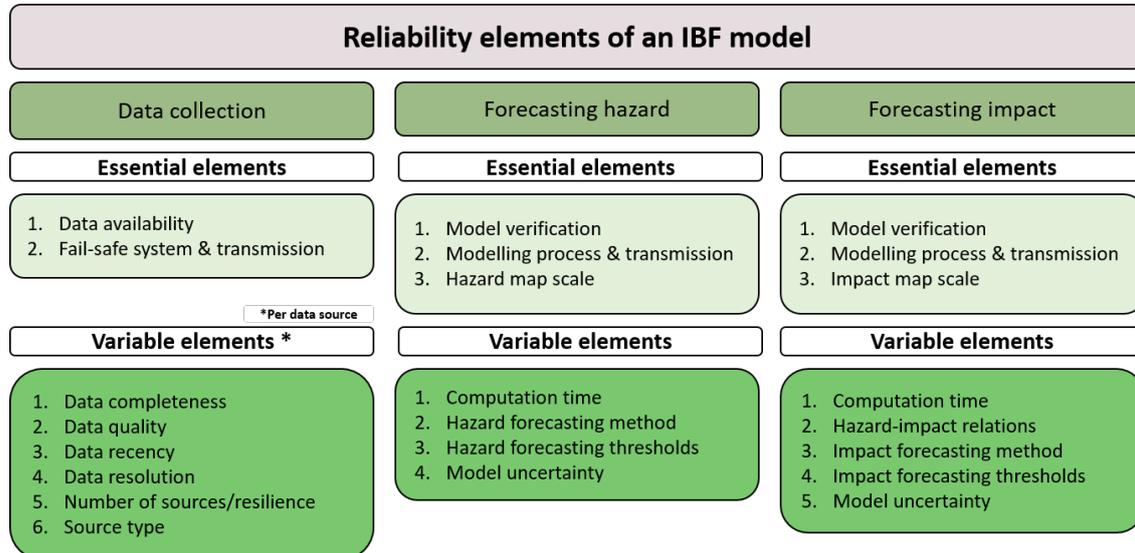


Figure 2.2: Overview of the reliability elements of an impact forecast indicating the essential and influential elements for the components in an IBF model for landslides. Appendix A includes the definition of the elements (Cabrera et al., 2021; Guzzetti et al., 2020; Joshi et al., 2020; Nadim and Intrieri, 2011; Segoni et al., 2018; Thirugnanam et al., 2020; UKMO, 2018; Van Westen et al., 2006; WMO, 2017; Waidyanatha, 2010; Wilkinson et al., 2018; Zhang et al., 2019).

Previous studies in landslide hazard and impact assessments were assessed to identify what is required for a reliable data collection, hazard and impact forecast for landslides. Requirements of the components and possible challenges which can be encountered link to the influential reliability elements (printed in *italic*) and are explained per component.

### 2.2.1 Data collection for landslide IBF

A landslide hazard and impact forecast requires the input of many data sources as described by Corominas et al. (2014), Van Westen et al. (2006), and Van Westen et al. (2008). An essential input is a landslide inventory used for validation procedures of the hazard and impact forecast (Cabrera et al., 2021; Van Westen et al., 2008). A reliable landslide inventory includes information on the location, timing, cause, trigger, type, runout and impact of a landslide (Corominas et al., 2014; Van Westen et al., 2006). Inventories often lack *data completeness* as many information sources only record those landslide events that caused substantial impact (Cabrera et al., 2021; Van Westen et al., 2006), resulting in an underestimation of past landslide events. It can take years of proper monitoring and data collection to create or complete inventories (Van Westen et al., 2006). Reliable techniques to increase the *quality* of landslide inventories have

been described by Guzzetti et al. (2012) and Van Westen et al. (2008). Especially the impact caused by landslides is frequently of poor quality due to the triggering relation of landslides with other hazards and their overlapping impact. Impact maps indicating solely landslide damage are scarce (CMA and GFDDR, 2016; Cabrera et al., 2021; De Ruiter et al., 2020). For the creation and validation of landslide impact forecasting models, an inventory that is as complete as can be must be used, as stated by Corominas et al. (2014).

Distinguishing between the cause and trigger of a landslide requires data sets relating to triggering mechanisms, environmental factors, antecedent conditions, previous landslides, inter-hazard relationships, the type of landslide and their difference in (hydrometeorological) trigger (Acosta et al., 2016; Bogaard and Greco, 2018; CMA and GFDDR, 2016; Cabrera et al., 2021; Sidle and Bogaard, 2016; Van Westen et al., 2006). Such information often lacks in *quality or completeness*, but is necessary for a reliable forecast of landslide probability and source location (Cabrera et al., 2021; Van Westen et al., 2006).

For impact forecasts, data on vulnerability, coping capacity and exposure of elements at risk is required (Cabrera et al., 2021; UKMO, 2018; Van Westen et al., 2006; WMO, 2015). As landslides are commonly classified as secondary hazard (Cabrera et al., 2021; Corominas et al., 2014), the elements at risk suitable for landslides can differ from those of the primary hazard. Appropriate indicators for the vulnerability and exposure to landslide hazard and impact can be *incomplete* or missing (Corominas et al., 2014; De Ruiter et al., 2020; Gill and Malamud, 2014; Van Westen et al., 2006).

### 2.2.2 Landslide hazard forecast

The *hazard forecasting method* used in landslide IBF models should be applicable for the type of impact forecast (Van Westen et al., 2006), but largely depends on the available information. Various susceptibility and hazard mapping techniques are described by Corominas et al. (2014).

A reliable landslide hazard map should include temporal and spatial factors, i. e. a value of the magnitude-frequency or spatially-distributed intensity characteristics. Current *hazard forecasting methods* do not fully and reliably include the intensity, volume or runout as output (Bogaard and Greco, 2018; Cabrera et al., 2021; Corominas et al., 2014; Guzzetti et al., 2012; Sidle and Bogaard, 2016; Van Westen et al., 2014).

Obtaining a reliable value for a *hazard forecasting threshold* is challenging as it requires the calibration of data to intrinsic values and antecedent conditions which are difficult to measure and model (Bogaard and Greco, 2018; Corominas et al., 2014) and depend on the landslide trigger (Van Westen et al., 2006; Van Westen et al., 2008).

### 2.2.3 Landslide impact forecast

The *impact forecasting method* depends on the output of the hazard map and end-users of the forecast (Cabrera et al., 2021; Corominas et al., 2014). Quantitative forecasts require values of mass, volume and runout extent, and for qualitative impact forecasts, the hazard map should contain spatially-distributed intensities (Van Westen et al., 2006; Van Westen et al., 2008).

Understanding *hazard-impact relations* in multi-hazard models is important, as explained in section 2.1 (Corominas et al., 2014; De Ruiter et al., 2020; Gill and Malamud, 2014; King, 2021). Reliable impact assessments should link the type of hazard with suitable indicators for the vulnerability and exposure of the elements at risk (De Ruiter et al., 2020; Van Westen et al., 2006; Van Westen et al., 2008).

The total *uncertainty* of the impact forecast is the combined uncertainty of the input data, hazard forecast and impact forecast, as the uncertainty propagates through the components of the IBF model (Gill and Malamud, 2014). A reliable impact forecast is therefore influenced by the combined reliability of each component (Thirugnanam et al., 2020).

Thus, creating an IBF model for landslides does not come without challenges. These are mainly due to data scarcity; the difference between cause and trigger; inter-hazard relationships of landslides; the output of hazard maps; and the type of vulnerability and impact assessment possible depending on the available information.

# 3

## METHOD: CREATION OF THE HYBRID MODEL

This chapter describes the machine learning IBF typhoon model and its static inclusion of landslides. The method for the creation of a prototype of a hybrid multi-hazard IBF model consists of three steps: establishing a regional event duration threshold, hazard and impact forecasting, and the reliability assessment.

### 3.1 THE 510 TYPHOON MODEL

The machine learning (ML) 510 typhoon model is trained on historical typhoon data to form a relationship between the typhoon hazard and its impact. The 510 typhoon model contains a ML model (grey outlined boxes in figure 3.1) and the 510 typhoon model (black outlined boxes in figure 3.1). The output of the ML model in the form of a hazard-impact curve is a source of input for the 510 typhoon model.

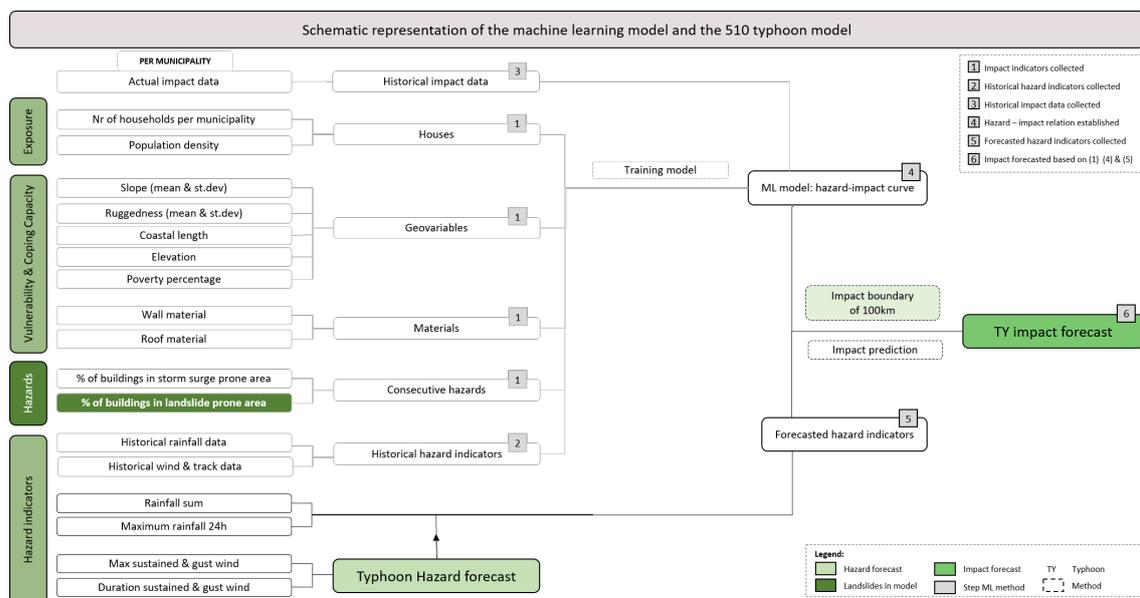
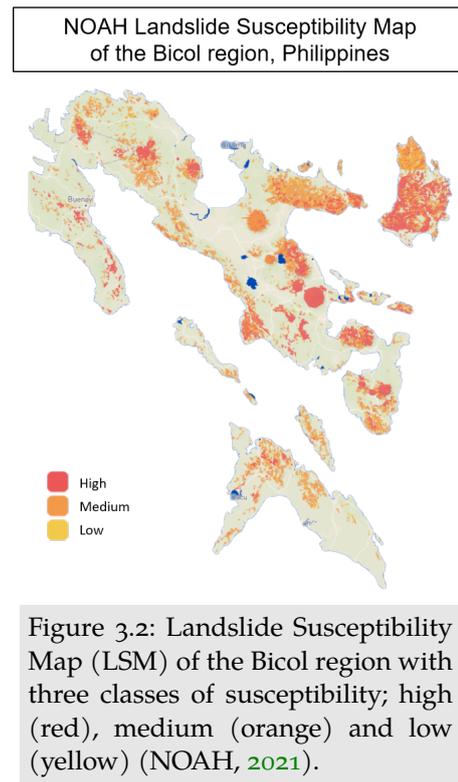


Figure 3.1: Overview of the 510 typhoon model. Grey boxes indicate 4 steps to create a machine learning (ML) model. The ML model is combined with the 510 typhoon model (black outlined boxes) to create an impact forecast by linking hazard indicators with exposure, vulnerability and coping capacity (adjusted from Teklesadik and Riquet (2021)). The dark green box shows the landslide component of the 510 typhoon model. A full size image can be found in Appendix B.2.

Numbers 1 to 4 (in grey boxes in figure 3.1) indicate four steps which result in the creation of the *ML model*. Data on vulnerability, coping capacity and exposure are collected in step 1. These impact indicators are combined with historical hazard indicators (rainfall and wind) and actual impact data in step 2 and 3, respectively. A hazard-impact curve is created in step 4 by relating the collected hazard and impact indicators with the actual impact. When a typhoon enters the Philippines area of reach,

hazard forecasting indicators are obtained in step 5. The hazard indicators and the ML model are provided as input for the 510 forecasting typhoon model - a code in R - (black outlined boxes in figure 3.1). An impact prediction is made in step 6 by comparing the hazard characteristics of the approaching typhoon with the hazard predictors of previous typhoons in the ML model, such that comparable impact is determined from the hazard-impact curve. This results in an impact forecast for the approaching typhoon in the format of a percentage of houses likely to get completely damaged (Teklesadik et al., 2022; Teklesadik and Riquet, 2021). A detailed explanation of the ML model and the 510 typhoon model is included in Appendix B.1.

Data on consecutive hazards of typhoons, i.e. landslides and storm surges, is integrated in the ML model, indicated by the dark green box in figure 3.1. The landslide input is static and consists of a percentage of buildings per municipality located in landslide-prone areas. A Landslide Susceptibility Map (LSM) (figure 3.2) constructed by the Philippine Nationwide Operational Assessment of Hazard (NOAH) indicates three areas susceptible to landslides: high (red), medium (orange) and low (yellow) (Eco et al., 2015; NOAH, 2021). Appendix B.3 provides background information on the creation of the map and a figure of the LSM of the Philippines. The LSM is overlain with housing data per municipality, resulting in a percentage of buildings per municipality located in the three susceptibility zones. This percentage per zone is one of the predictors in the ML model, as indicated in figure 3.1. The 510 typhoon model can be referred to a multi-hazard model as it considers this static input for landslides. However, research indicated little to no influence of the landslide predictor on the hazard-impact curve in the ML model (Teklesadik et al., 2022). No additional or dynamic data for landslides is included in the 510 typhoon model and therefore the impact forecast lacks the inclusion of impact from landslides.



### 3.2 CREATING A HYBRID MODEL

Landslide hazard and impact can be integrated either in the 510 typhoon model in the ML model or the 510 typhoon model. To create suitable landslide hazard-impact relations in the ML model, reliable landslide inventories providing information on e.g. previous landslide occurrences location and timing, landslide types and landslide impact, are required. Such inventories do not exist for the Bicol region. The hazard-impact relations created by the ML model are based on the available data and cannot be assessed or validated due to the complexity and of machine learning (Tunkiel et al., 2020; Wagenaar et al., 2020). Therefore, the multi-hazard IBF model is created by extending the 510 typhoon model with a separate and simplistic IBF model for

landslides. Thereby, the inclusion of landslides is altered from static in the ML model to dynamic in the 510 typhoon model.

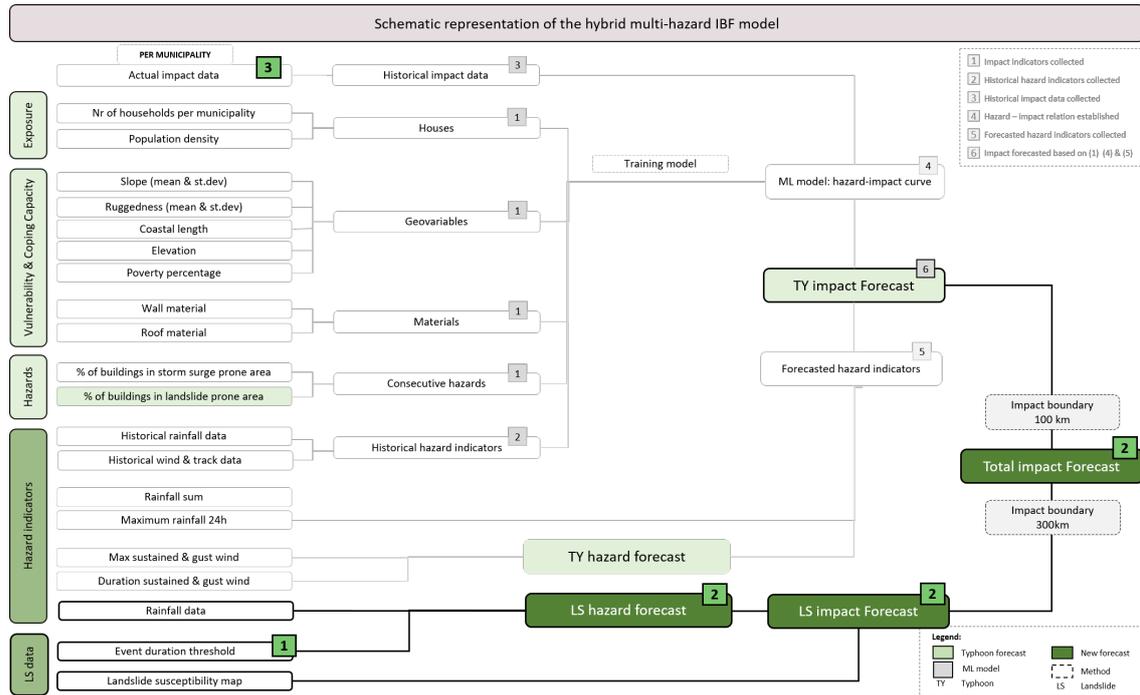


Figure 3.3: Overview of the hybrid multi-hazard IBF model in which the dark green boxes indicate landslide components. Light grey and grey outlined boxes are not altered or used. Black outlined variables are applied or created in the design of the hybrid model. The bright green boxes show the step of the method and link to figure 3.4. Appendix B.2 has a full size image.

The prototype of the combined models for landslides and typhoons is referred to as the hybrid multi-hazard (MH) IBF model, or hybrid model (figure 3.3). The 510 typhoon model, ML model and typhoon impact forecast (indicated in light grey) are not altered during the creation of the hybrid model. A separate IBF model for landslides is constructed (black outlined boxes in black figure 3.3) and integrated as extension of the typhoon impact wind forecast to create a total impact forecast.

The landslide IBF model combines rainfall data and the LSM (available from the 510 typhoon model) with a regional event duration (ED) threshold. The ED threshold is created in step 1 of the method (green box in figure 3.3). In step 2, a landslide hazard and impact forecast is made and combined with the typhoon impact forecast. In the reliability assessment in step 3, the total impact forecast of the hybrid model is compared to the typhoon impact forecast of the 510 typhoon model and the actual impact. The three steps of the method for the creation of the hybrid model in figure 3.4 link to those indicated in figure 3.3.

### 3.2.1 Step 1: Regional event duration threshold

The exceedance of a rainfall threshold by the actual rainfall indicates the onset of landslides occurrences. In this research, a regional event duration (ED) curve is used to determine the rainfall threshold. It is created with the automatic "Calculation of Rainfall

Thresholds for Landslide" occurrence (CTRL-T) (Melillo et al., 2016, 2018). The ED approach considers the duration of a rainfall event (D) and the associated accumulated rainfall (E) in mm. It is a suitable approach in case of landslides triggered by typhoons, instead of other threshold methods, as typhoon events cause large amount of rainfall in a short amount of time. The ED method has been applied in the landslide study project of the START network in the Bicol region and in the LANDSLIP project in India, which makes use of the CTRL-T model (Brunetti et al., 2021; Zamudio and Orogo, 2021).

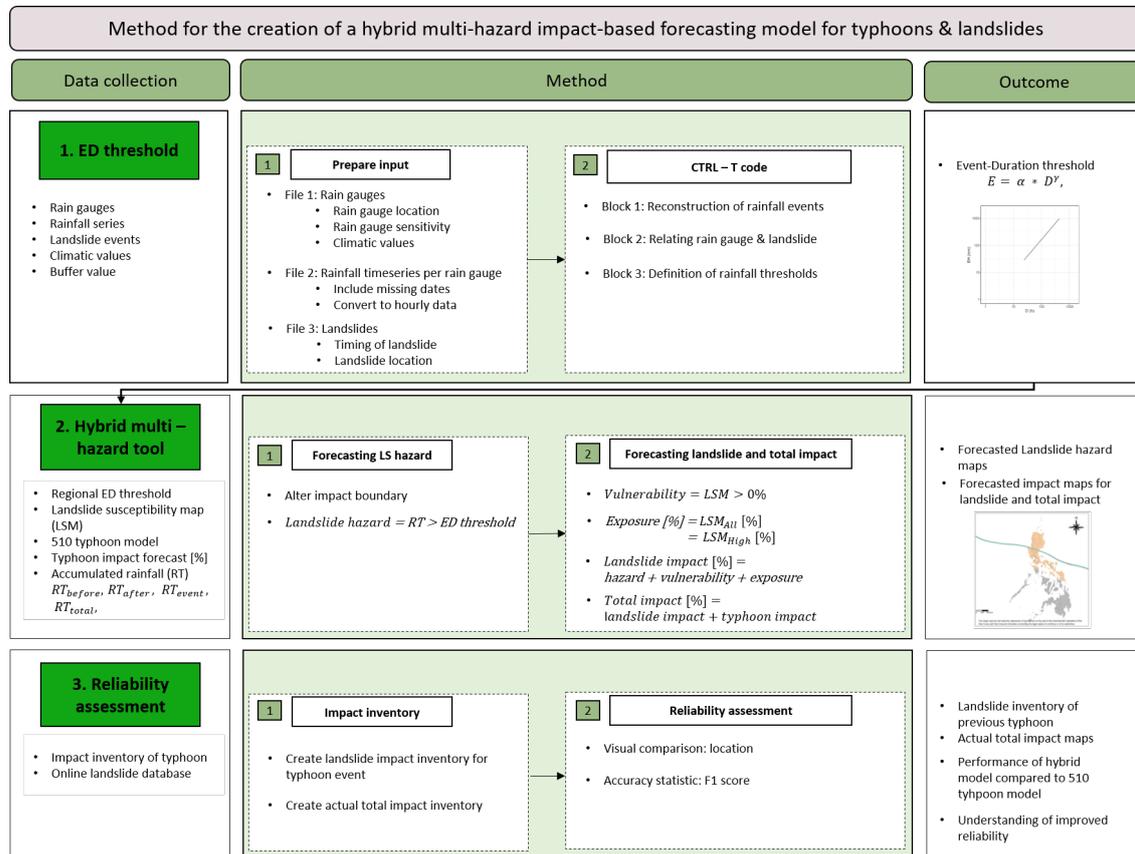


Figure 3.4: Method overview showing the three steps in the creation of a hybrid multi-hazard IBF model. Per step, the data collection, method and outcome is indicated. The steps correspond to the numbers indicated in figure 3.3. A full size image can be found in Appendix B.2.

The first step of the method of CTRL-T (visualised in step 1, the first row of figure 3.4) is the collection of data and preparation of input files. Input values are shown in table C.2 for three file types: rain gauge locations, rainfall time series and landslide events (Melillo et al., 2016, 2018). Per rain gauge, rainfall time series containing dates and hourly precipitation are collected. Missing dates are included by inserting precipitation values of 0 or 200 mm, depending on the occurrence of a typhoon at the missing date. Precipitation of 200 mm per day is applied to account for the extreme precipitation associated with typhoon events and validated with accumulated rainfall from previous typhoon events, shown in Appendix B.5 (JAXA, 2021). Daily rainfall is converted to hourly precipitation by equal distribution. This can cause inaccuracies for the selection of the duration of a rainfall event and affects the uncertainty parameters of the ED curve, but is applicable as the ED threshold has a daily temporal scale (Brunetti et al., 2021; Gariano et al., 2020; Melillo et al., 2016). Landslide timing is set to 23:59 of the day of

occurrence to include the rainfall occurring at the day of landslide onset. The spatial range of which landslides are linked to which rain gauge, the buffer, is adjusted in the code of CTRL-T. An increased buffer value is used if more reconstructed rainfall events are required or in mountainous regions to improve the ED threshold accuracy.

After the input preparation, three blocks are considered in CTRL-T (step 1 of figure 3.4) (Melillo et al., 2016, 2018). Block 1 reconstructs rainfall events out of the rainfall series by filtering the irrelevant rain events using climate parameters. In Block 2, the landslide and rainfall events are coupled by selecting the rain gauge that measured the specific rainfall event responsible for a landslide occurrence. This is firstly done by selecting rainfall prior to the landslide event (antecedent rainfall). Then, rainfall conditions responsible for a landslide and their representative rain gauges are linked. Finally, the most likely rainfall condition for a landslide event is selected. Block 2 returns an image of the landslide event, the rain gauges related to that landslide and information on the duration and accumulated rainfall of the rain event associated with the landslide initiation. In the final block, event duration curves are created from the various rainfall conditions.

The result is a set of ED curves for different probabilities in the form of the power law in formula 3.1 and figure 3.4.

$$E = (\alpha + \Delta\alpha) * D^{(\gamma + \Delta\gamma)} \quad (3.1)$$

where  $E$  is the accumulated rainfall (mm),  $\alpha$  is the mean intercept and its uncertainty ( $\Delta\alpha$ ),  $D$  is the duration of the rainfall event (in h) and  $\gamma$  is the mean slope and its uncertainty ( $\Delta\gamma$ ).

The power law indicates the minimum amount of accumulated rainfall ( $E$ ) required for the initiation of landslides for a certain period of time ( $D$ ). The event duration curve at 5% exceedance probability is selected as regional daily ED threshold. The ED curve is only valid for durations exceeding 24h hours (Brunetti et al., 2021; Melillo et al., 2016, 2018) and is required as input for the hybrid model.

### 3.2.2 Step 2: Hazard and impact forecast

In step 2, a landslide and total impact forecast are created by expanding the 510 typhoon model with a separate landslide IBF model (figure 3.3). In the first phase of step 2 (figure 3.4), the 510 typhoon model, LSM, ED threshold and rainfall data are combined.

The accumulated rainfall, *Rainfall Total* ( $RT$ ) for four different time ranges in relation to typhoon landfall (figure 3.5) are defined, of which  $RT_{before}$  and  $RT_{after}$  can be used in forecasting applications.  $RT_{after}$  and  $RT_{total}$  are used as quality checks of the forecasts.

- $RT_{before}$  : accumulated rainfall before landfall with duration  $D = 72h$
- $RT_{after}$  : accumulated rainfall after landfall with duration  $D = 72h$
- $RT_{event}$  : accumulated rainfall before and after landfall with duration  $D = 96h$
- $RT_{total}$  : accumulated rainfall over the total time range with duration  $D = 144h$

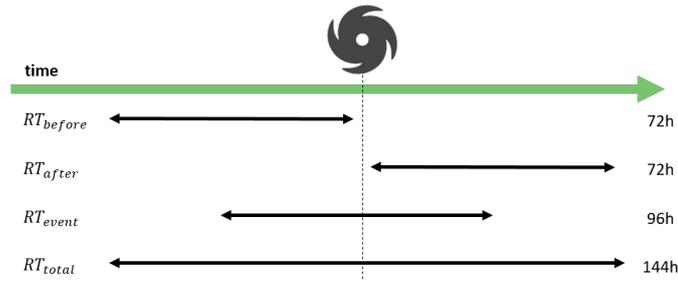


Figure 3.5: Accumulated rainfall (RT) for four different time ranges in relation to typhoon landfall.

After the data collection, landslide hazard (equation 3.2) is defined by the exceedance of the ED threshold by the accumulated rainfall (RT):

$$\text{Landslide hazard} [-] = RT > ED \text{ threshold} \quad (3.2)$$

Per municipality, a landslide hazard forecast is created for each  $RT$  within a suitable hazard and impact boundary. The typhoon hazard and impact boundary is set to 100 km in the 510 typhoon model (Teklesadik et al., 2022). The hazard and impact boundary for landslide and total impact was assessed as rainfall associated with typhoons reaches further from the typhoon track compared to extreme winds. Based on rainfall patterns using the JAXA Global Rainfall Watch and actual landslide events (IFRC, 2016a; JAXA, 2021), the impact boundary was increased to 300 km. The impact boundary analysis can be found in Appendix B.4.

Landslide impact (as defined by CMA and GFDDR (2016) and WMO (2015)) is expressed in percentage of buildings per municipality likely to get damaged:

$$\text{Landslide impact} [\%] = \text{hazard} + \text{vulnerability} [\%] + \text{exposure} [\%] \quad (3.3)$$

where the *vulnerability* and *exposure* are to linked the LSM, using the available percentage of buildings in hazard-prone areas per municipality. Two classes of exposure are distinguished, based on the susceptibility zones of the LSM: *All* for all classes and *High* for medium and high vulnerable zones (figure 3.2). Landslide impact in percentage of buildings is predicted for both exposure classes and each  $RT$ :

$$\text{Vulnerability} = \text{LSM} > 0\% \quad (3.4)$$

$$\text{Exposure}_{all} = \text{LSM} [\%]_{high} + \text{LSM} [\%]_{medium} + \text{LSM} [\%]_{low} \quad (3.5)$$

$$\text{Exposure}_{high} = \text{LSM} [\%]_{high} + \text{LSM} [\%]_{medium} \quad (3.6)$$

The total impact forecast of the hybrid multi-hazard IBF model combines the landslide impact forecast of the landslide IBF model with the typhoon impact forecast as predicted by the 510 typhoon model (figure 3.3):

$$\text{Total impact} [\%] = \text{typhoon impact} [\%] + \text{landslide impact} [\%] \quad (3.7)$$

### 3.2.3 Step 3: Reliability assessment

In step 3, the reliability of the hazard and impact forecast is assessed by comparing the predicted impact with the actual impact for case study typhoons (figure 3.4).

Impact inventories are created by combining the actual typhoon impact, available from the ML model of the 510 typhoon model (box 3 in figure 3.1), with a typhoon-specific landslide inventory. The 'Cooperative Open Online Landslide Repository' (COOLR) project is used as a basis to find landslide occurrences from previous typhoons used in the ML model of the 510 typhoon model (NASA, 2021; Teklesadik et al., 2022). The COOLR database combined with additional research defined the case study typhoons; those with at least 3 landslides for one typhoon event in the Bicol region: Typhoon Durian (2006) and Typhoon Sarika (2016). Appendix B.5 provides information on the typhoon tracks and landslide impact of the typhoons.

The landslide hazard and impact forecasts are compared with validated landslide impact. Total impact forecasts are compared with both the actual impact and the forecast of the 510 typhoon model. The latter determines whether the reliability of an impact forecast is improved by the hybrid model compared to the impact forecast of the 510 typhoon model. The reliability of a forecast is assessed regarding the similarity of the forecast with the actual impact in twofold:

- The geographical *location* of the forecast: by visually comparing the location and extent of the forecast with the actual impact.
- The *accuracy* of the forecast: by statistically comparing the forecasted impact with the actual impact using the F1 score (Chicco and Jurman, 2020; Sisters, 2020):

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (3.8)$$

where *TP* are True Positives and account for correctly forecasted impact; *FP* are False Positive forecasts where impact occurred, but is missed and *FN* are False Negatives and impact is incorrectly predicted. True Negatives (*TN*) correctly predict no impact.

The F1-score is an accuracy measure, which does not consider *TN* forecasts. The True Scale Statistics (TSS) and the Matthew's Correlation Coefficient (MCC) are measures which do include all types of forecasts (*TP*, *TN*, *FP* and *FN*) and give objective performance of the hazard and impact predictions (formulas are shown in Appendix C in section C.3). The MCC also functions well in unbalanced data sets such as those with scarce events (Chicco and Jurman, 2020; Rawat, 2019). But, MCC and TSS do not distinguish between the preference of stakeholders on the type of forecast, i. e. positives or negatives. The application of the typhoon and landslide forecast in a humanitarian context should be considered, where *TP*s are more important compared to *TN*s and acting in vain (*FP*s) is preferred over not acting at all (*FN*s) (Chicco and Jurman, 2020; IFRC, 2021; Rawat, 2019; Sisters, 2020; Wilkinson et al., 2018). Therefore, the F1-score is applied as accuracy statistic in this research instead of TSS or MCC. In case of a forecast with no *TP*s and of which the F1 score is zero, the MCC and TSS measures can be assessed.

# 4

## RESULTS: THE HYBRID MODEL IN THE BICOL REGION

The hybrid multi-hazard IBF model (hybrid model) was created for the Bicol region and tested on two typhoon events. The reliability of the forecasts of the landslide IBF model and hybrid model were assessed and compared to the reliability of the 510 typhoon model.

### 4.1 A REGIONAL EVENT DURATION THRESHOLD FOR THE BICOL REGION

#### Rain gauges & landslide locations in the Bicol region, Philippines

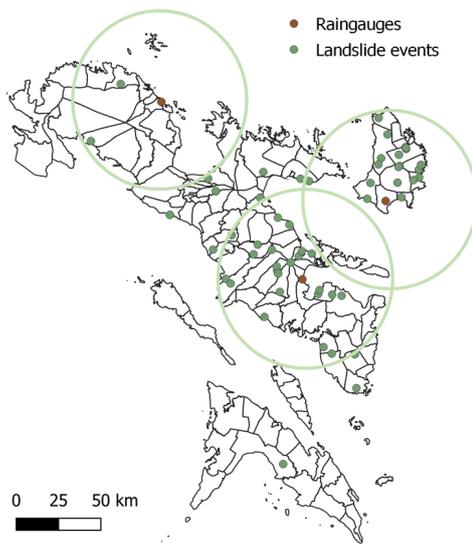


Figure 4.1: Rain gauge (red) and landslide (green) locations used as input for CTRL-T. Circles show the buffer around rain gauges.

In the Bicol region three rain gauges (Daet, Legaspi and Virac station) with a sufficient number of years of rainfall measurements were available (red locations in figure 4.1). Daily precipitation values were obtained for the years 2006 till 2018, as these coincided with the 59 available landslide occurrences selected from the COOLR project from NASA (2021) (green locations in figure 4.1). Any alterations of the default settings of CTRL-T for the case study area of Bicol can be found in table C.2 in Appendix C.1. The sensitivity value for the rain gauges and climate parameters of this research were based on the study of LANDSLIP in India (Brunetti et al., 2021). The buffer value around the rain gauges was increased from 16km (default) to 60km to account for the low rain gauge density (light green circles).

The event duration (ED) threshold computed for the Bicol region for the 5% non-exceedance probability (black graph in figure 4.2) was:

$$E = (0.6 \pm 1.0) \times D^{(1.2 \pm 0.2)} \quad (4.1)$$

The threshold of Bicol (in black) was compared to the thresholds of the LANDSLIP project (in grey) in figure 4.2 and showed a decrease in scaling parameter  $\alpha$  (lowered curve), but had a similar slope parameter,  $\gamma$ , (dashed line) (Brunetti et al., 2021). The decrease in  $\alpha$  is a result of the limited number of

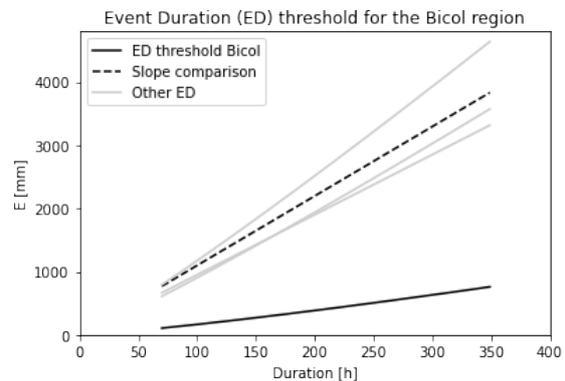


Figure 4.2: ED threshold for Bicol in black, compared to thresholds of Brunetti et al. (2021).

landslide occurrences and rainfall conditions, visible in Appendix C.1. The daily resolution of rainfall data resulted in a clustering of the rainfall duration into multiples of 24h. This decreased the number of reconstructed rainfall conditions and lowers the threshold of Bicol compared to the study in India (Brunetti et al., 2021; Gariano et al., 2020; Melillo et al., 2016, 2018). The ED threshold has a validity interval equal to the duration interval of the rainfall conditions (Melillo et al., 2016). Therefore, the ED threshold is only valid for the prediction of landslides using accumulated rainfall with durations of multiples of 24 hours.

In Appendix C.1, ED thresholds for other non-exceedance probabilities, the reconstructed rainfall conditions and landslide events and formulas for ED thresholds of the LANDSLIP study can be found.

## 4.2 RELIABILITY ASSESSMENT OF THE HYBRID MODEL IN THE BICOL REGION

### 4.2.1 Actual impact of typhoons Durian and Sarika

The municipalities in which impact occurred for typhoon Durian and Sarika were created using the typhoon impact data from the 510 typhoon model and landslide inventories for typhoons Durian and Sarika (figure 4.3). The color shade of the municipalities for typhoon Durian indicates the magnitude of the impact in % of buildings completely damaged. Impact for typhoon Durian was concentrated around the typhoon track passing over the Bicol region (figure B.8 in Appendix B) and included a limited number of municipalities at distances further away from the track.

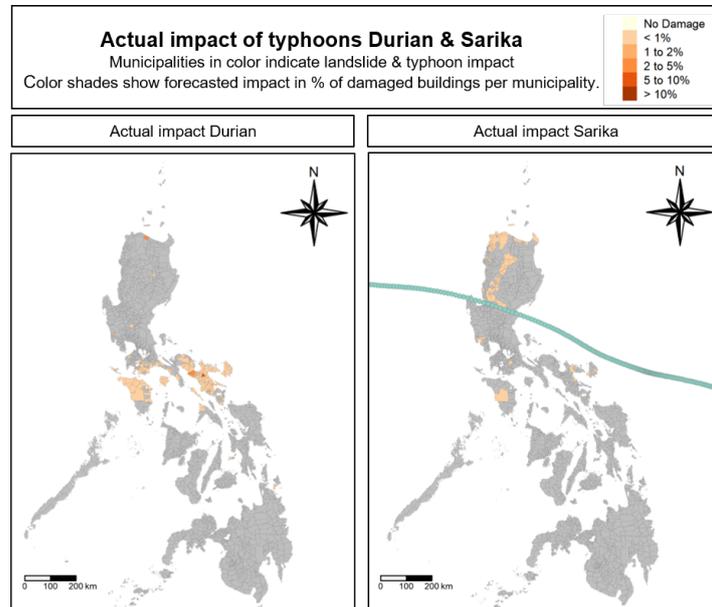


Figure 4.3: Municipalities in which typhoon and landslide impact occurred for typhoons Durian and Sarika. For typhoon Durian, color shades show the % of buildings damaged (Fano et al., 2007; NASA, 2021). For typhoon Sarika impact was not quantified (DSWD, 2016a,b; IFRC, 2016a,b; NASA, 2021; PH government, 2008).

For typhoon Sarika, landslide impact cannot be quantified as it considered impact to buildings and additional elements at risk. The actual impact maps of Sarika indicated only the forecasted landslide impact occurrences (referred to as FLIO), not the magnitude of the impact. Typhoon and landslide impact combined in figure 4.3 showed impact occurred mostly north of the track (figure B.8). In the south and in the Bicol region, municipalities in which impact occurred were scattered and impact was mainly caused by landslides.

It should be noted that information on landslide impact in the Philippines for specific typhoon events is scarce (Zamudio and Orogo, 2021). The actual impact maps were used to validate the reliability of the forecasts of the hybrid model.

#### 4.2.2 The reliability of the landslide hazard forecast

##### Typhoon Durian

Four landslide events were found in the Bicol region for typhoon Durian, concentrated around the slopes of a volcano (left image in figure 4.4). Landslide hazard forecasts (top images) showed an overestimation in locations for hazard maps  $RT_{before}$ ,  $RT_{event}$  and  $RT_{total}$  compared to the actual hazard locations.  $RT_{after}$  showed a hazard forecasts only in the west of Bicol. Rainfall decreased with the passing of the typhoon at 72 hours after landfall, causing an exceedance of the ED threshold solely in the west. The forecast of  $RT_{after}$  was therefore less reliable as it considers accumulated rainfall at a time range too long after typhoon landfall. Rainfall patterns can be found in Appendix C.2.

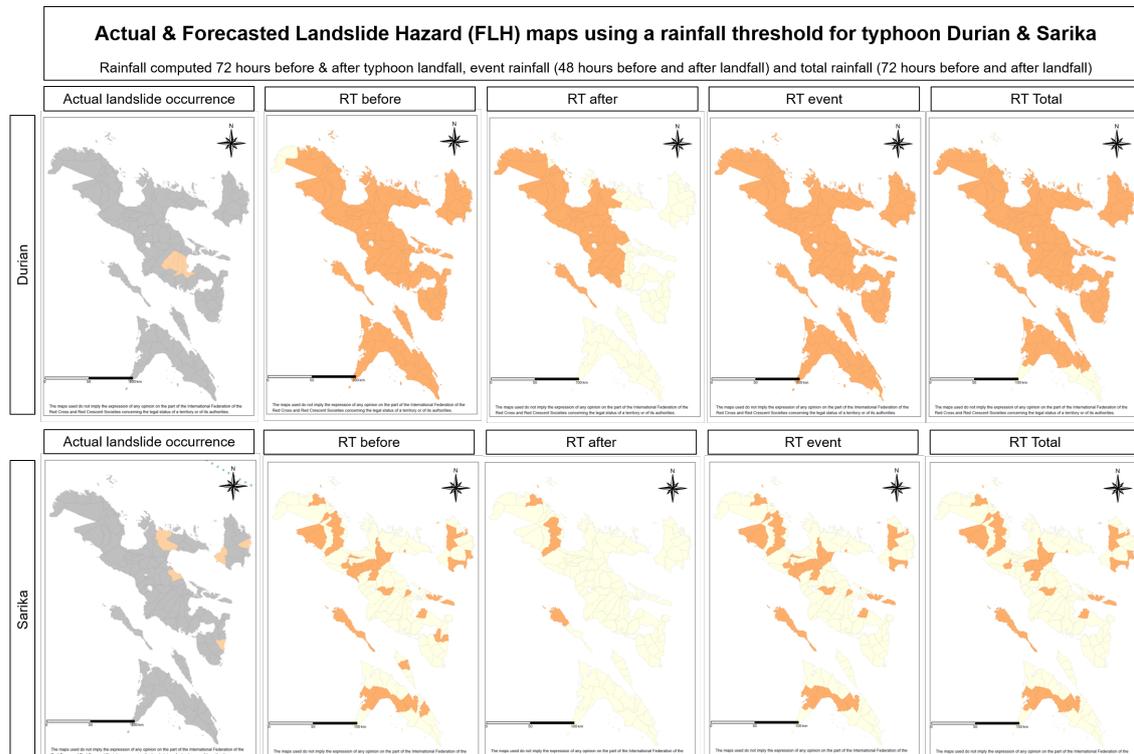


Figure 4.4: Actual and forecasted landslide hazard (FLH) maps for typhoon Durian (top) and Sarika (below) for four types of accumulated rainfall (RT); 72h before typhoon landfall (before) and after (after); 48h before and after landfall (event); the total rainfall 72h before and after landfall (total).

All hazard forecasts for typhoon Durian had a low F1 score (table 4.1) due to the large number of False Positives (FP) and scarcity of True Positives (TPs). For  $RT_{before}$ ,  $RT_{event}$  and  $RT_{total}$ , all actual TP values were found, unlike  $RT_{after}$  which missed 2 TPs, reducing the F1 score and reliability of the hazard forecast.

Table 4.1: Accuracy statistics for hazard forecasts of typhoon Durian and Sarika for four types of accumulated rainfall (RT); 72h before typhoon landfall (before) and after (after); 48h before and after landfall (event); the total rainfall 72h before and after landfall (total).

|    | Durian: 4 Landslides |              |              |              | Sarika: 5 Landslides |              |              |              |
|----|----------------------|--------------|--------------|--------------|----------------------|--------------|--------------|--------------|
|    | $RT_{before}$        | $RT_{event}$ | $RT_{total}$ | $RT_{after}$ | $RT_{before}$        | $RT_{event}$ | $RT_{total}$ | $RT_{after}$ |
| TP | 4                    | 4            | 4            | 2            | 0                    | 0            | 0            | 0            |
| F1 | 0.069                | 0.0683       | 0.072        | 0.048        | 0                    | 0            | 0            | 0            |

### Typhoon Sarika

The hazard forecast of  $RT_{after}$  for typhoon Sarika (lower images of figure 4.4) was unreliable in geographical location, for the same reason as  $RT_{after}$  for typhoon Durian. Location extent was overestimated for the residual forecast maps. No correct locations were predicted, but locations in which landslide hazard was forecasted were near actual hazard-prone municipalities.

No TPs of the hazard forecast (table 4.1) resulted in an F1 score of zero. Accuracy measurements of TSS and MCC (Appendix C.3), whose numerators do not depend solely on the value of TP, were negative, indicating low reliability of the hazard forecast.  $RT_{after}$  had less negative results due to the increase in True Negatives (TNs) and decrease in FP.

#### 4.2.3 The reliability of the landslide impact forecast

### Typhoon Durian

Landslide impact forecasts for *All* susceptibility classes of the LSM, referred to as *All*, in the top images of figure 4.5 showed an overestimation of forecasts in location extent. Similar and less reliable landslide impact forecast (FLI) using  $RT_{after}$  for *All* and high susceptibility classes, *High*, were obtained due to the less reliable hazard forecast. The location extent of impact forecasts for *High* classes using  $RT_{before}$ ,  $RT_{event}$  and  $RT_{total}$  deviated least from the validated impact map (left image).

Accuracy statistics for typhoon Durian (table 4.2) were close to zero and identical for  $RT_{before}$ ,  $RT_{event}$  and  $RT_{total}$  using *All* classes, due to the scarcity of actual TPs and overestimation of the landslide hazard. No TPs were found for the residual FLI maps. Therefore MCC and TSS values were assessed (Appendix C.3) and were found to have slightly negative values, indicating low reliability of the forecasts.  $RT_{before}$  and  $RT_{event}$  showed most reliable values for MCC, TSS and F1 and can be applied for forecasting.

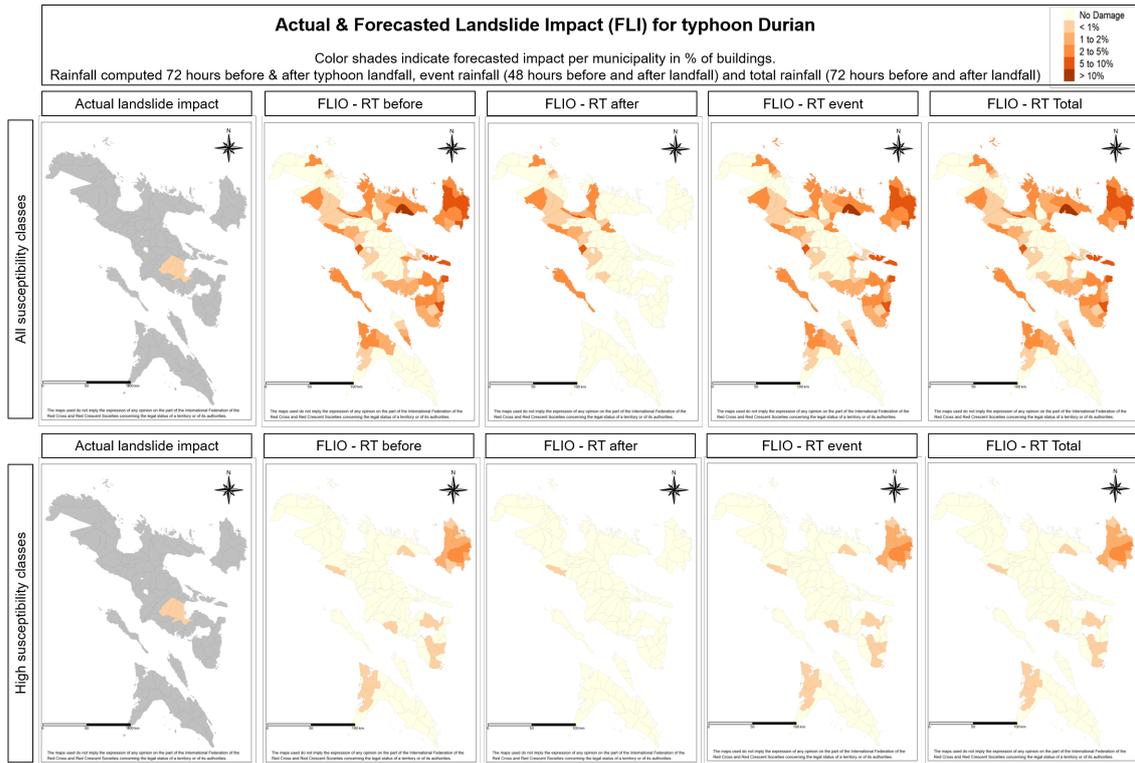


Figure 4.5: Actual and forecasted landslide impact (FLI) maps for typhoon Durian for four types of accumulated rainfall (RT); 72h before typhoon landfall (before) and after (after); 48h before and after landfall (event); the total rainfall 72h before and after landfall (total) and for two susceptibility classes: All (above) and High (below).

Table 4.2: Accuracy statistics for landslide impact forecasts of typhoon Durian & Sarika for four types of accumulated rainfall (RT); 72h before typhoon landfall (before) and after (after); 48h before and after landfall (event); the total rainfall 72h before and after landfall (total) and for two classes of susceptibility (All and High).

|    |  | All classes          |              |              |              |                      |              |              |              |
|----|--|----------------------|--------------|--------------|--------------|----------------------|--------------|--------------|--------------|
|    |  | Durian: 4 Landslides |              |              |              | Sarika: 5 Landslides |              |              |              |
|    |  | $RT_{before}$        | $RT_{event}$ | $RT_{total}$ | $RT_{after}$ | $RT_{before}$        | $RT_{event}$ | $RT_{total}$ | $RT_{after}$ |
| TP |  | 1                    | 1            | 1            | 0            | 0                    | 0            | 0            | 0            |
| F1 |  | 0.028                | 0.028        | 0.028        | 0            | 0                    | 0            | 0            | 0            |
|    |  | High classes         |              |              |              |                      |              |              |              |
|    |  | $RT_{before}$        | $RT_{event}$ | $RT_{total}$ | $RT_{after}$ | $RT_{before}$        | $RT_{event}$ | $RT_{total}$ | $RT_{after}$ |
| TP |  | 0                    | 0            | 0            | 0            | 0                    | 0            | 0            | 0            |
| F1 |  | 0                    | 0            | 0            | 0            | 0                    | 0            | 0            | 0            |

### Typhoon Sarika

Forecasts of typhoon Sarika are referred to as 'Forecasted Landslide Impact Occurrence' (FLIO) (figure 4.6), as the actual landslide impact map indicated locations of the municipalities in which impact occurred, not the magnitude of the impact .

FLIOs for *All* classes using  $RT_{before}$ ,  $RT_{event}$  and  $RT_{total}$  showed a scatter of the forecasts regarding the location, mostly towards the outer reaches of the Bicol region (figure 4.6). No correct locations were found, yet forecasted locations were in the vicinity of actual landslide impact locations. All other FLIO maps showed an underestimation of the impact and were therefore least reliable regarding location of the forecasts.

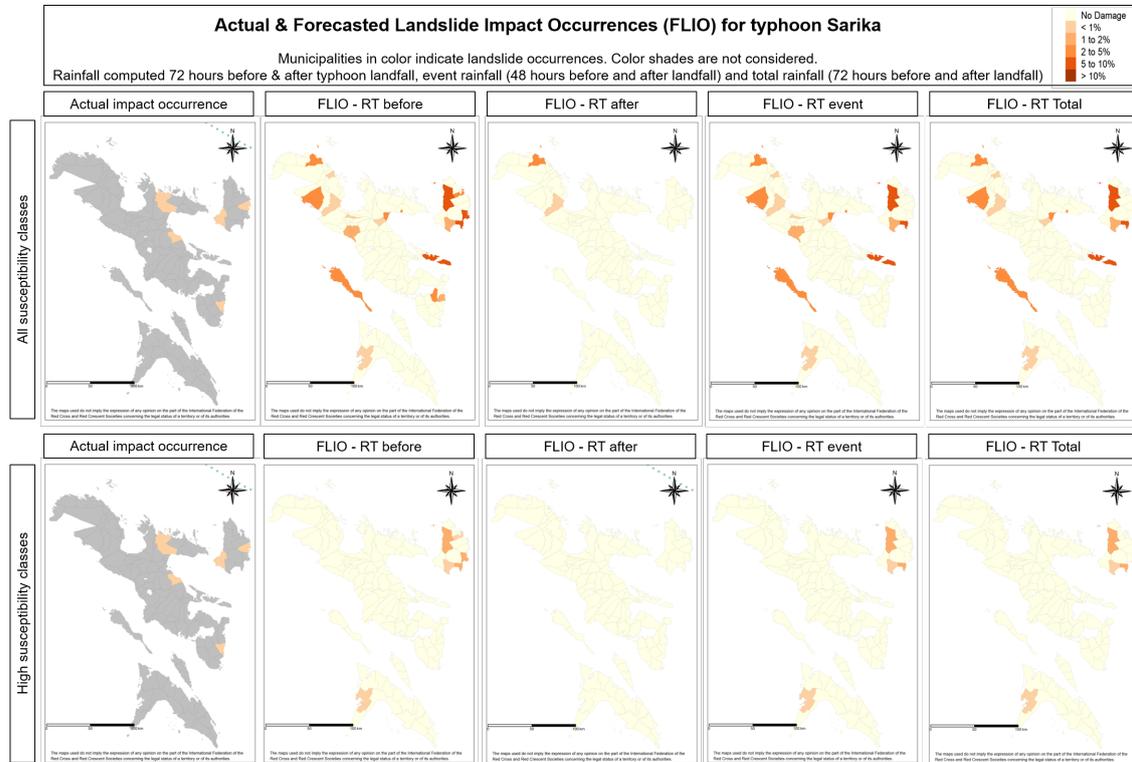


Figure 4.6: Actual and forecasted landslide impact occurrence (FLIO) maps for typhoon Sarika for four types of accumulated rainfall (RT); 72h before typhoon landfall (before) and after (after); 48h before and after landfall (event); the total rainfall 72h before and after landfall (total) and for two susceptibility classes: All (above) and High (below).

As no TPs the hazard forecast of typhoon Sarika included no TPs, the impact forecast either was alike and had an F1 score of zero. Accuracy statistics of MCC and TSS (Appendix C.3) were negative for all FLIO maps. For forecasting applications, landslide impact forecasts using  $RT_{before}$  or  $RT_{event}$  were the most reliable compared to  $RT_{total}$  and  $RT_{after}$  with regard to the actual impact of typhoon Sarika.

#### 4.2.4 The reliability of the landslide IBF model

The most reliable landslide hazard forecasts for both typhoon events in terms of location were obtained by  $RT_{before}$ ,  $RT_{event}$  and  $RT_{total}$ , when considering all RTs. Correct location forecasts were found only for typhoon Durian. The ED threshold was exceeded in less municipalities for typhoon Sarika compared to typhoon Durian, due to a decrease in accumulated rainfall at larger distance from the typhoon track.

The reliability in location of the impact forecasts was best for  $RT_{before}$ ,  $RT_{event}$  and  $RT_{total}$  of *High* classes for typhoon Durian. The same RT maps but for *All* classes for typhoon Sarika displayed the most reliable forecasts in location. TP values were obtained solely for typhoon Durian using  $RT_{before}$ ,  $RT_{event}$  and  $RT_{total}$  of *High* classes.

$RT_{before}$  and  $RT_{event}$  can be used for forecasting purposes, as both indicated most reliable hazard and impact forecasts of the landslide IBF model.

#### 4.2.5 The reliability of the hybrid model

The impact boundary for typhoons is exceeded in the Bicol region for typhoon Sarika. Therefore, no total impact can be calculated.

The location extent of the forecasted total impact (FTI) of typhoon Durian (figure 4.7) was underestimated by the hybrid model for  $RT_{after}$  using both susceptibility classes. Locations of these forecasts were constrained to municipalities located in the vicinity of the typhoon track and were similar to the FTI of the 510 typhoon model (second image). FTI using  $RT_{before}$ ,  $RT_{event}$  and  $RT_{total}$  determined with *All* classes (bottom images) displayed more reliable forecasts in municipalities located further away from the track. The latter FTI were comparable to the actual total impact (left image).

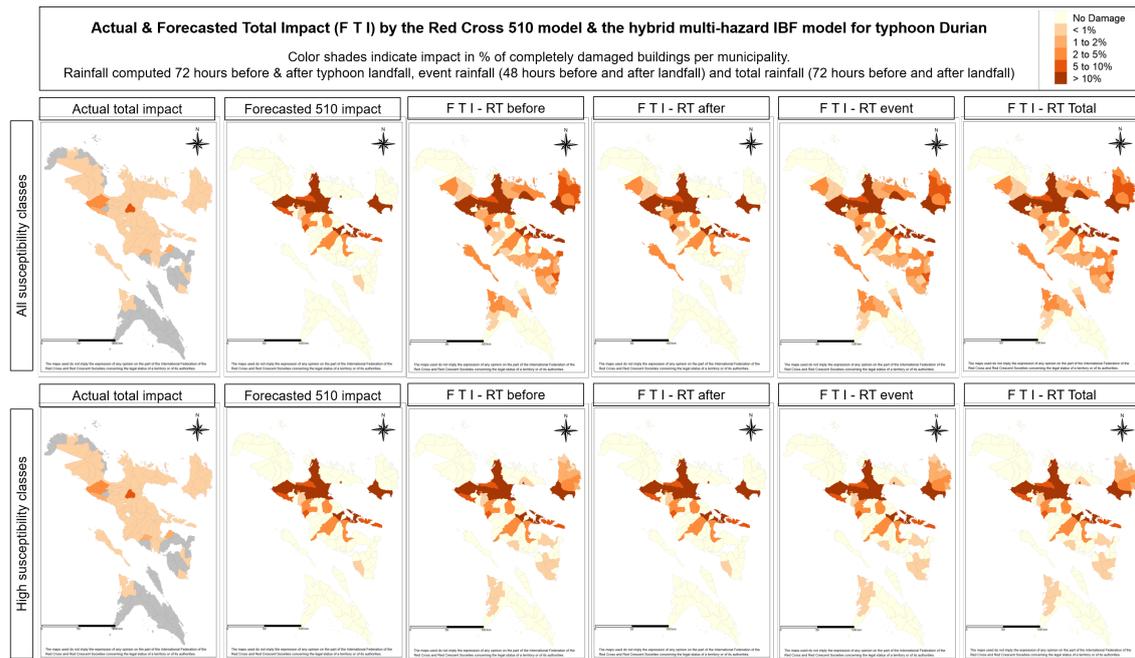


Figure 4.7: Total impact maps for typhoon Durian of the validated impact; typhoon impact; and combined landslide and typhoon impact. Two susceptibility classes (All and High) and four types of accumulated rainfall (RT) were applied to the hybrid model; 72h before typhoon landfall (before) and after (after); 48h before and after landfall (event); the total rainfall 72h before and after landfall (total).

Values for TP, FP, TN, FN and all accuracy statistics (F1, TSS and MCC) for the total impact forecasts for the hybrid and 510 typhoon models can be found in Appendix C.3.

For each FTI of both the hybrid and the 510 typhoon model TPs were underestimated compared to the actual TPs (table 4.3). Most reliable results regarding the number of TPs were obtained for RTs using *All* classes, of which highest F1 scores were obtained by  $RT_{before}$ ,  $RT_{event}$  and  $RT_{total}$  due to a reduction in FNs. Total impact as forecasted by the 510 typhoon model was identical to the impact predicted by the hybrid model using  $RT_{after}$  for *High* classes. The number of TPs and F1 score indicated least reliable results for this FTI.

Table 4.3: Accuracy statistics for the total impact maps of the typhoon impact using a machine learning method (510 typhoon model - shortened to 510 model) and the hybrid model for two susceptibility classes (*All* and *High*) and four types of accumulated rainfall (RT); 72h before typhoon landfall (before) and after (after); 48h before and after landfall (event); the total rainfall 72h before and after landfall (total) and two susceptibility classes.

| Durian | 76 impacted municipalities |                                    |              |              |              |
|--------|----------------------------|------------------------------------|--------------|--------------|--------------|
|        | 510 model                  | Hybrid model - <i>All</i> classes  |              |              |              |
|        | Total                      | $RT_{before}$                      | $RT_{event}$ | $RT_{total}$ | $RT_{after}$ |
| TP     | 31                         | 59                                 | 59           | 61           | 59           |
| F1     | 0.559                      | 0.771                              | 0.771        | 0.763        | 0.690        |
|        | 510 model                  | Hybrid model - <i>High</i> classes |              |              |              |
|        | Total                      | $RT_{before}$                      | $RT_{event}$ | $RT_{total}$ | $RT_{after}$ |
| TP     | 31                         | 38                                 | 38           | 38           | 31           |
| F1     | 0.559                      | 0.628                              | 0.628        | 0.628        | 0.559        |

The total impact as forecasted by the hybrid model for *All* classes and all RTs showed improved reliability compared to the 510 typhoon model impact forecast as a result of the doubling of TPs and the reduction of FNs by half compared to the predicted impact of the 510 typhoon model. Improved and most reliable F1 scores were found for  $RT_{before}$ ,  $RT_{event}$  and  $RT_{total}$  for the hybrid model. The hybrid model can be used in forecasting applications for the accumulated rainfall of  $RT_{before}$  and  $RT_{event}$  for *All* classes. These impact forecasts indicated improved reliability regarding both the accuracy statistics and geographical location compared to the impact as predicted by the 510 typhoon model, which was underestimated.

# 5

## DISCUSSION: THE RELIABILITY OF THE HYBRID MODEL

---

Assumptions and limitations apply to the approach used for the creation of the landslide IBF model, the hybrid model and the reliability assessment and are linked to the *reliability elements* of figure 2.2 and Appendix A. Recommendations to resolve or improve the limitations are defined in chapter 7.

### 5.1 DATA COLLECTION

#### 5.1.1 *The 510 typhoon model and its data sources*

The propagated uncertainty of all components within the 510 typhoon model for results in a combined uncertainty of the total impact forecast in the hybrid model (De Ruiter et al., 2020; Thirugnanam et al., 2020). Typhoon impact forecasts of  $<0.5\%$  are set to a value of  $0.5\%$  by default by the machine learning model (ML model) of the 510 typhoon model (Teklesadik et al., 2022). It was assumed that removing values  $<0.5\%$  from the typhoon impact forecast increases its *data quality*.

The reliability of the typhoon impact forecast decreases if impact is  $<3\%$  of damaged buildings or if impact exceeds a 100 km range from the typhoon track (Teklesadik et al., 2022). For typhoon Sarika no typhoon impact was forecasted for the entire Philippines due to its low impact values of  $<0.5\%$  (*data quality*) and location of the Bicol region ( $>100$  km distance from track) (*spatial resolution*). By creating a separate landslide IBF model with a larger impact boundary, the limitation of the spatial extent of the landslide and total impact forecast is avoided.

A limitation of the 510 typhoon model is the exclusion of typhoon events that do not make landfall (*data incompleteness*). The extent of rainfall from passing typhoons can be within the boundary of landslide hazard and impact, even if the typhoon does not make landfall. Therefore, the landslide IBF model can include typhoons that do not make landfall and has an increased impact boundary of 300 km compared to the 100 km impact boundary of the 510 typhoon model.

The Landslide Susceptibility Map (LSM), as provided by the 510 typhoon model, was used for in the landslide IBF model as measure for vulnerability. The LSM does not consider intensity-frequency, volume or runout extent of landslides, which are required for a reliable landslide hazard forecast (section 2.2) and decrease the reliability of the LSM concerning *data quality and completeness* (Eco et al., 2015; Jaboyedoff and Labiouse, 2011; Luzon et al., 2016; Rabonza et al., 2016). The percentage of buildings coupled to the susceptibility classes was assumed to be a valid quantitative measure of exposure, yet is based on values from 2015 and should be updated in *data recency*.

### 5.1.2 The steep and lowered event duration threshold

The event duration (ED) threshold  $E = (0.6 \pm 1.0) \times D^{(1.2 \pm 0.2)}$  of formula 4.1 showed a reduction in scaling parameter ( $\alpha$ ) (figure 4.2) and slight increase in uncertainty in the slope parameter ( $\Delta\gamma$ ) parameter compared to the values found in the LANDSLIP project (Appendix C.1). An increased uncertainty and underestimation of the threshold are undesirable for IBF models as these can result in an increase in False Positives (FPs) and reduction in reliability of the hazard and impact forecast (Gariano et al., 2020; Piciullo et al., 2017).

The lowered ED threshold and increased uncertainty of  $\Delta\gamma$  are caused by *data incompleteness* and *reduced data quality* of the landslide inventory and low number of rain gauges in the Bicol region. Limited data (*number of sources*), the daily temporal resolution of available rainfall series (*resolution*) and unsuitable climate variables (*data quality*) hamper the reconstruction of enough rainfall conditions considered for landslide initiation events. Therefore, the rainfall conditions for the Bicol region are not very representative, reducing the reliability of the ED threshold (Brunetti et al., 2021; Gariano et al., 2020; Melillo et al., 2016, 2018; Peres et al., 2018; Piciullo et al., 2017). To account for the low number of reconstructed rainfall conditions, the buffer value around rain gauges was increased to 60 km. On the other hand, Gariano et al. (2020) showed that ED thresholds can also be lowered in case of a higher distance between landslide and rain gauge. The trade-off between the buffer value and reconstructed rainfall conditions is unknown.

The lack of *available data* in the Bicol region affected the reliability of the ED threshold. This region was chosen due to the operational use of the 510 typhoon model and the local project of the START network, which focuses on the creation of an ED threshold for landslide initiation in the Bicol region (Zamudio and Orogo, 2021). The tool and method of CTRL-T can be applied in additional regions of the Philippines for which more landslide and rainfall information is available, e. g. in the province of Luzon.

It was assumed that missing rainfall measurements (*data incompleteness*) in the rainfall series could be set to zero or 200 mm, depending on the occurrence of a typhoon event, rather than interpolating the preceding and succeeding rainfall values. Interpolating the rainfall measurements of nearby rain gauges was impossible, firstly because missing dates often included various consecutive days within the rainfall series. Secondly, rainfall series of other rain gauges lacked measurements on identical dates, possibly due to the sudden and large amount of rainfall affecting the sensors (*source resilience* and *-location*). Therefore, one rainfall amount was chosen and applied to all missing measurements. The value of 200 mm of rainfall was assumed to be a large enough amount to relate typhoon rainfall to landslide events and its exceedance was validated by the accumulated daily rainfall of the case studies in this research (figure C.2).

The *data resolution* of the ED rainfall threshold of CTRL-T is at regional scale, yet is applied for hazard forecasting at municipal level. The amount of rainfall required for landslide initiation can differ per municipality due to antecedent and local climate conditions (Cabrera et al., 2021; Van Westen et al., 2006; Van Westen et al., 2008). A local

threshold ED threshold could be more appropriate considering its application scale. However, local rainfall thresholds focus on more specific landslide events (Chung et al., 2017), while the impact forecasts of the landslide IBF model solely indicates whether a landslide occurred in an entire municipality. Also, IBF models using an ED threshold as hazard forecasting method at larger scale, e.g. the entire Philippines, require the creation of multiple regional ED thresholds. This requires performing step 1 of the developed method (section 4.1) for each region.

### 5.1.3 *The four values of accumulated rainfall*

During the creation of the hybrid model, all four time ranges of accumulated rainfall data (RTs) were assessed to provide understanding of the rainfall patterns during typhoon events and the initiation of landslides in relation to typhoon landfall. For forecasting purposes,  $RT_{before}$  and  $RT_{event}$  can be applied and their time range relates to the lead time of the 510 typhoon model (72h). Wind and rainfall data are collected simultaneously for both models three days prior to typhoon landfall (TSR, 2007). Prolonging the download of forecasted rainfall up to 24 hours after typhoon landfall allows for the use of  $RT_{event}$  and the forecasting of landslides initiated after landfall. Reconstructed rainfall events for landslide initiation in the Bicol region indicate a mean duration of 96 hours (figure C.1 in Appendix C.1), which is similar to the duration of  $RT_{before}$  and  $RT_{event}$  combined.

## 5.2 THE RELIABILITY OF THE LANDSLIDE IBF MODEL

### 5.2.1 *The landslide hazard forecast*

The *hazard forecasting method* applied in this research is the ED threshold instead of other rainfall related method as it is used in similar research in India (Brunetti et al., 2021) and by the Start Network in the Bicol region (Zamudio and Orogo, 2021). Other rainfall thresholds were not assessed e.g. thresholds based on intensity duration curves or antecedent conditions. An alternative threshold approach is explained by Monsieurs et al. (2019) and proposes coupling antecedent soil conditions to an existing landslide susceptibility map. This approach might be an effective method in the Philippines because of the existing LSM. Hazard thresholds other than rainfall can be considered, because rainfall only may not be the most optimal landslide hazard indicator. Improved reliability of the hazard forecast might be obtained by the exceedance of *multiple thresholds*, e.g. multiple rainfall thresholds or a combination with slope stability.

The ED threshold includes an *uncertainty* due to the lowered curve, which could be the reason for the overestimation of landslide hazard for both case study typhoons (section 4.2.2). The hazard forecasts showed the suitability of the ED threshold approach as landslide hazard forecasting method, if the ED curve's reliability improves.

The temporal resolution of the ED threshold in the Bicol region and its application for rainfall measures of durations for multiples of 24 hours only (Gariano et al., 2020) should be considered for operational use. For forecasting applications, both  $RT_{before}$  and  $RT_{event}$  are suitable. As landslides in the Bicol region are mainly triggered by the rains

from typhoons, it is advisable to include a short time frame for the landslide hazard forecast after the typhoon has made landfall, i. e.  $RT_{event}$  (Zamudio and Orogo, 2021). This is consistent with the rainfall conditions established in the Bicol region using CTRL-T and the rainfall patterns of the case study typhoons (figures 4.2 and C.2). It does require prolonged collection of rainfall forecasts for the landslide IBF model compared to the 510 typhoon model (PRC, 2019; TSR, 2007; Teklesadik et al., 2022).

The hazard boundary was set to 300 km from the typhoon track, based on actual rainfall patterns and landslide inventories (figure B.7). However, rainfall patterns are not homogeneously distributed and can be scattered within and outside of the chosen range. Multiple indicators, e. g. a combination of both rainfall and presence of steep slopes, and a variable impact boundary can provide for a more reliable landslide hazard boundary compared to using a fixed value of 300 km.

### 5.2.2 The landslide impact forecast

In the *impact forecasting method* for landslides it is assumed that the LSM coupled to the percentage of buildings in hazardous zones can be used as measure for vulnerability and exposure. Even though the LSM was created for the assessment of buildings at risk for landslides, the zones do not consider vulnerability indicators of buildings for landslides. As not the 'impact' is forecasted but rather the 'percentage of houses in exposed zones', the term 'landslide impact forecast' is not completely correct. A proper impact forecast can be created by coupling the current impact forecast with a hazard-impact curve, similar to the curve for typhoons as constructed by the ML model of the 510 typhoon model. Either the existing ML model of 510 can be used for a proper landslide impact forecast, if additional vulnerability indicators of buildings appropriate for landslides are included, or an additional landslide ML model can be created.

To couple the landslide impact with typhoon impact, the elements at risk had to be the same. Therefore, the *hazard-impact relations* of the elements at risk are assumed to be alike for typhoons and landslides. But, forecasting impact for consecutive hazards should consider other or additional elements at risk more suitable for the secondary hazard. Infrastructure, e. g. roads and bridges, or rural impact, e. g. crops and cattle, can provide for a more reliable hazard-impact relation and impact forecast for landslides (Cabrera et al., 2021; De Ruiter et al., 2020; Gill and Malamud, 2014; King, 2021).

Reliable landslide inventories are required for *model validation* of the the landslide impact forecast, but information on past landslide events in the Philippines is scarce possibly resulting in an underestimation of typhoon-specific landslide events. Inventories in this research consist of a combination of the impact provided by the 510 typhoon model and additional landslide events. The impact of the 510 typhoon model includes percentage of buildings destroyed in a municipality, but does not distinguish between the cause of impact, e. g. typhoon, landslide or storm surge. The additional landslide inventories often include landslide occurrences instead of landslide impact or landslide impact for other elements at risk than buildings. The impact inventory of typhoon Durian (figure 4.3) solely contained the impact inventory of the 510 typhoon model, as no data additional landslide locations were available. The impact inventory of

typhoon Sarika cannot be quantified as the additional landslide inventory consisted of landslide locations instead of the number of buildings damaged by landslides.

This research is limited to the application of the landslide IBF model for two typhoon events in the Bicol region. If an ED threshold and landslide inventory is available, the model can be implemented for additional typhoon events within and outside of the Bicol region to gain more understanding of the reliability of the model.

### 5.3 THE HYBRID MODEL

It was assumed that the typhoon impact from the 510 typhoon model and landslide impact from the landslide IBF model were equally important as the two were added to obtain a total impact forecast of the hybrid model. A weighted sum of the two hazards can provide for a more reliable method, but requires additional research for various reasons. As a first example, landslides are often triggered by typhoons and are therefore the secondary hazard. However, if the landslide is triggered without the typhoon making landfall, it is the primary hazard. Secondly, landslide events are scarce and occur in less, more scattered locations further from the track compared to the mainly aggregated locations of the typhoon impact near the track. Thirdly, the land use of the location of the impact should be considered. Landslides often occur in rural areas on slopes and can cause more impact to agriculture and cattle compared to the rural impact caused by typhoons. On the contrary, landslides happen less frequently in urban areas and will therefore have a smaller impact compared to a typhoon. Finally, the impact from rainfall can cause more damage than the winds from the typhoons (TSR, 2007), as it can trigger secondary hazards. Concluding, merely a sum of two weighted components does not comprise an accurate total impact. Therefore, a more elaborate research to establish the weights of the sum is required.

The total impact forecast considers the LSM twice: firstly as indicator for vulnerability and exposure by the landslide IBF model and secondly as input for the ML model of the typhoon forecast (figure 3.1). Even though the contribution of the LSM on the ML model is little (Teklesadik et al., 2022), the reliability of the total impact forecast will be improved if the ML model is trained without the LSM as predictor. Since, the LSM is included in the typhoon impact forecast of the 510 typhoon model, it is in essence a total impact forecast in itself. This does allow for a comparison of the increased reliability of the hybrid model including a separate and dynamic input for landslide impact with the 510 'multi-hazard' model with static landslide input.

The aggregation of the independent impact forecasts does not consider the inter-relations of the two hazards (De Ruiter et al., 2020; Gill and Malamud, 2014; King, 2021). By independently combining the two impact forecasts, *hazard-impact relations* are excluded and total impact can be underestimated (De Ruiter et al., 2020).

The separate landslide IBF model in the hybrid model provides additional benefits compared to an integrated landslide impact forecast within a ML model. The spatial extent or boundary of impact can be adjusted to suit both hazards. Typhoons that do not make landfall can be included without affecting the reliability of the typhoon impact

forecast. The forecasting time frame for landslides can continue after typhoon landfall, i. e. for  $RT_{event}$ , after typhoon forecasting stops.

#### 5.4 THE RELIABILITY ASSESSMENT

The reliability assessment provides a first insight into the improved reliability of the forecast from the hybrid model in location and accuracy compared to the impact forecast of the 510 typhoon model. The influence of landslide impact, and ideally other consecutive hazards of typhoons, should be assessed for additional typhoon events in Bicol and other regions of the Philippines.

The F1 score indicated improved reliability of the hybrid model compared to the 510 impact forecast, contrary to the subjective accuracy statistics True Scale Statistics (TSS) and Mathews Correlation Coefficient (MCC). MCC was higher for the 510 typhoon model (MCC numerator of 960) compared to the hybrid model (MCC numerator of 600), because of the values of the numerator of the MCC formula (Appendix C.3). TSS is obtained by subtracting the False Positive Rate (FPR) from the True Positive Rate (TPR). TSS is larger for the 510 typhoon model (TPR of 0.8, FPR of 0.06) compared to the hybrid model (TPR of 0.4, FPR of 0.56) because of the difference between FPR and TPR. Other accuracy statistics were not included, e. g. those used by Corominas et al. (2014), Gariano et al. (2020), Guzzetti et al. (2012), and IFRC (2021). Yet these can give additional information on the accuracy of the impact forecast. The decision on which accuracy measure is most suitable for the reliability assessment depends on the stakeholders involved (Sisters, 2020; Wilkinson et al., 2018).

In this research, the reliability was defined as the resemblance of the impact forecast with the actual impact in terms of geographical location and accuracy of the forecast. The magnitude of the impact forecast is not assessed because the impact inventories considered not only buildings. Many other interpretations, definitions and methods of reliability can be assessed e. g. in the studies of Guzzetti et al. (2020), Tate (2012), and Thirugnanam et al. (2020) and may provide for additional understanding of the performance and reliability of the hybrid model.

The reliability assessment is a local sensitivity analysis as it was performed separately for the landslide hazard, landslide impact and total impact (Tate, 2012). A global sensitivity analysis (GSA) (explained by Feizizadeh et al. (2015) and Tate (2012)) considers the combined influence of each component on the total impact simultaneously and indicates the influence of the dynamic input of landslides on the total impact forecast. GSA may be a more appropriate method considering that the reliability depends on the combined reliability of the components (Thirugnanam et al., 2020). The influence of the static landslide input in the 510 typhoon model can be assessed using the GSA method described by Tunkiel et al. (2020), yet is more challenging as the 510 typhoon models uses a ML approach.

#### 5.4.1 Temporal reliability

A reliability assessment on elements of influence of an IBF model was performed (section 3.2.3), but the temporal reliability (lead time) was not included. Without enough lead time, communities are less prepared and more vulnerable for impact and disasters. Lead time is influenced by both decision-making and the time-related reliability elements of the IBF model, i. e. lag time and computation time. A shorter lag time, by improving the reliability elements in the data collection phase, increases the lead time and reliability of a forecast. A ML approach as used in the 510 typhoon model is an effective method to reduce lag time as the exposure and vulnerability of a region are defined ahead of time in the ML model (Thirugnanam et al., 2020).

Data collection of the 510 typhoon model and hybrid model starts five days ahead of a typhoon event and lead time is 72 hours (PRC, 2019). The accumulated rainfall of the landslide IBF model was consistent with the lead time of the 510 typhoon model, yet can be altered to four or five days prior to typhoon landfall (TSR, 2007) to increase the lead time and the implementation of early actions. Additional lead time is obtained when using the time frame of  $RT_{event}$  as well, which considers landslide events shortly after landfall. However, measures for landslide impact, e. g. slope stabilization, constructions to slow down landslides and planning strategies (Cabrera et al., 2021), often require a lead time longer than three to five days. Using landslide hazard and impact forecasts, especially in Asia, can be an appropriate and cost-effective mitigation measure to determine evacuation strategies and reduce impact (Segoni et al., 2018; TSR, 2007). An increased lead time for landslide impact forecasts may thus not improve the reliability of the forecast in terms of effective measures and may even decrease the accuracy and reliability of the impact forecast, as longer lead times increase the uncertainty of rainfall forecasts (Wilkinson et al., 2018). The trade-off between a longer and reduced lead time depends amongst others on the hazard forecast, the spatial scale of the impact forecast, stakeholders and their preference of acting options e. g. 'acting in vain' or 'low regret options' (Thirugnanam et al., 2020; Wilkinson et al., 2018).

# 6 CONCLUSION

---

Impact-based forecasting (IBF) models consist of successive components: (1) the data collection, (2) the hazard and (3) impact forecast. These influence the reliability of the impact forecast. IBF models underestimate impact when consecutive hazards, e.g. typhoon-triggered landslides, are not considered. Landslide and typhoon impact combined differs from solely typhoon impact, as the combination has a more dynamic form of vulnerability and a larger affected area. Considering the impact of both landslides and typhoons can thus result in a more reliable impact forecast. Therefore, a hybrid multi-hazard IBF model for typhoons and landslides (hybrid model) was created.

A hydrometeorological IBF model to forecast landslide occurrences (landslide IBF model) was created using a newly established regional event duration threshold for the Bicol region. A prototype of a hybrid model was designed by expanding an existing machine learning (ML) IBF model for typhoons from 510, an initiative of the Netherlands Red Cross (510 typhoon model), with the landslide IBF model. Landslide and total impact inventories were created for two case study typhoons in the Bicol region, i. e. typhoons Durian and Sarika. The reliability of both the landslide IBF and hybrid models was assessed with an increased impact boundary of 300 km compared to the previous typhoon impact boundary of 100 km.

Four different time ranges of accumulated rainfall in relation to typhoon landfall and two measures for exposure were established and examined with the landslide IBF model.  $RT_{before}$  (rainfall of 72 hours before landfall) and  $RT_{event}$  (rainfall of 24 hours before and after the typhoon event) indicated the most reliable landslide hazard and impact forecasts and can be used for forecasting impact. The results of the hybrid model using complete exposure values and  $RT_{before}$  and  $RT_{event}$  showed an improved impact forecast compared to the impact predicted by the 510 typhoon model. The geographical location and extent of impact as forecasted by the hybrid model showed increased resemblance with the validated impact compared to the impact of the 510 typhoon model. The accuracy of the impact forecast from the hybrid model was improved by doubling of the True Positives and reduction of the False Negatives by half compared to the existing typhoon model.

The separate landslide IBF model as an extension of the existing ML typhoon model provided additional benefits as these models can be decoupled to optimize the performance and reliability of both. The spatial extent of impact can be adjusted to suit both hazards. Typhoons that do not make landfall can be included for landslide impact forecasts. The forecasting time frame for landslides can include a short period of time after typhoon landfall when forecasting of typhoon impact stops.

The reliability of the hybrid model was mostly influenced by four elements: data quality, data completeness, hazard and impact forecasting method. The model validation and reliability assessment of both the landslide and hybrid model should be performed on additional typhoon events and for additional regions for which reliable landslide inventories are available.

This study resulted in the prototype of an impact-based multi-hazard IBF model for typhoons and landslides for the Philippines and demonstrated the importance of considering impact from consecutive hazards.

# 7

## RECOMMENDATIONS & NEXT STEPS

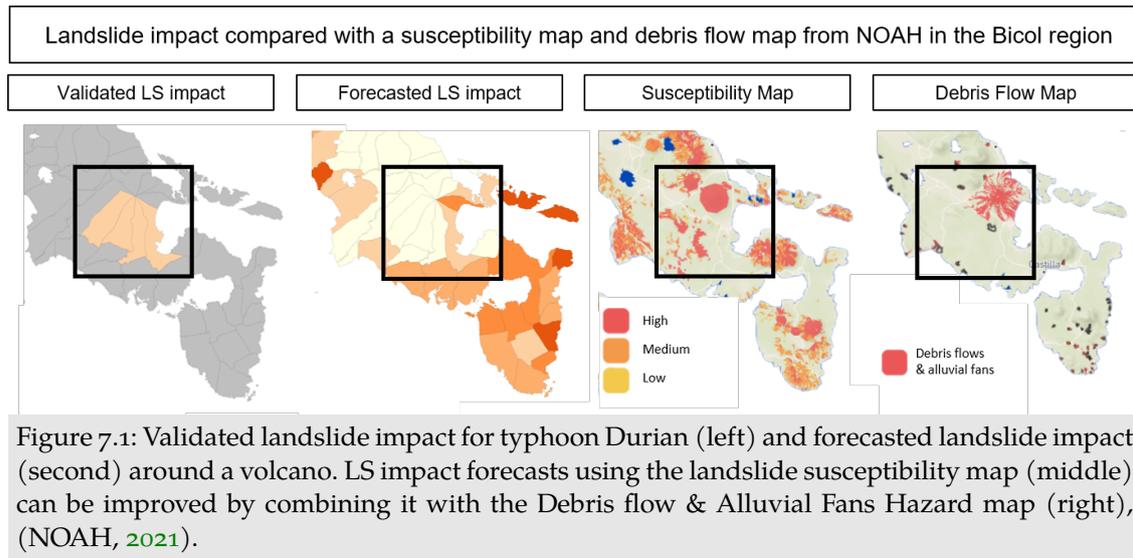
---

Recommendations for improving the landslide IBF model and hybrid multi-hazards IBF model relate to limitations and assumptions discussed in chapter 5.

### 7.1 IMPROVING THE LANDSLIDE IBF MODEL

Firstly, the landslide IBF model can be improved with a revision of the event duration (ED) threshold by increased data availability (sections 5.1 and 5.2). Landslide data in the Bicol region and the Philippines is scarce (Zamudio and Orogo, 2021), therefore it is recommended to focus on upgrading existing landslide inventories. The use of satellite data as source for the rainfall series resolves the issue of the low rain gauge density, the missing measurements, the temporal resolution of the data and the conversion of daily to hourly data (Gariano et al., 2020). The study of Brunetti et al. (2021) demonstrates a method in which satellite data is used in CTRL-T. Also, satellite data collection can be applied on a different spatial scale, both locally and on a wider basis for the entire Philippines. This allows for a reduction of the buffer value, which increases the reliability and eases the expansion of the ED threshold to other regions in the Philippines. It should be noted that obtaining cloud-free satellite imagery in tropical regions is challenging. Besides, satellite imagery does not capture variability at a local scale. Therefore, retrospective forecasts or 'hindcasts' can also be considered as source of rainfall series and are often available from validation data sets of forecasting models (Guzzetti et al., 2020; UKMO, 2018). Reliable climate values can be determined by the method proposed by Melillo et al. (2018), in which a monthly soil water balance (MSWB) model is used to determine the climatic variables required in CTRL-T. The MSWB model requires monthly rainfall, -temperature, maximum field capacity and water storage data to determine the water balance variables of a region, hence requires additional modelling.

The landslide IBF model can be further improved by including runout from landslides in the landslide susceptibility map (LSM). Runout can be calculated using available models e. g. Flow-R (Horton et al., 2013), which is applied by Eco et al. (2015) and Van Westen et al. (2014). A less-time consuming and efficient approach is the addition of the existing 'Debris Flows and Alluvial Fans' hazard map of NOAH (figure B.4) to the LSM, which can be used to simulate runout locations. Part of the LSM and debris flow map in figure 7.1 show the missed landslide impact (False Negatives) for typhoon Durian around the Mayon Volcano. The LSM does not consider the correct impact locations (True Positives), whereas the debris flow map does. Thus, it is recommended to expand the LSM with the debris flow map to simulate runout extent of landslides in a simplistic manner (NOAH, 2021) and update the percentage of buildings coupled the LSM with the most recent data.



An improved reliability of the hazard forecast of the model can be obtained by using multiple values for the accumulated rainfall and multiple sources of rainfall data to reduce uncertainties in rainfall measurements and increase source resilience (Guzzetti et al., 2020; Thirugnanam et al., 2020). The reliability of the hazard forecasting method can be increased when additional and multiple hazard forecasting indicators, e. g. soil moisture or slope stability, are applied to the threshold (Thirugnanam et al., 2020). It is advised to assess other landslide hazard indicators suitable for the Philippines to improve both the hazard forecast and impact boundary.

Finally, the reliability of the landslide IBF model can be increased if vulnerability indicators of appropriate elements for landslides are included in the landslide impact forecast. More appropriate elements at risk for landslides (e. g. crops, cattle, roads and bridges) should be included in the impact forecast and can help to determine suitable early action measures for both landslides and typhoons. A next study can create a prototype of the landslide IBF and hybrid multi-hazard model for the rice crop prediction model for typhoons in the Philippines of 510 of the Netherlands Red Cross and assess landslide impact in rural areas (Van Brussel et al., 2021).

## 7.2 IMPROVING THE PROTOTYPE OF THE HYBRID MULTI-HAZARD IBF MODEL

The prototype of the multi-hazard hybrid IBF model (hybrid model) can be improved by solving and considering the assumptions and limitations of the model, discussed in section 5.3. Firstly, the total impact forecast should be corrected by re-training the variables in the machine learning (ML) model of the 510 typhoon model, which resolves the inclusion of the LSM twice. Also, the total impact forecasting method can be improved by applying a weighted sum of the impact forecasts considered to differentiate between the impact of typhoons, landslides and other consecutive hazards of typhoons, e. g. floods and storm surges. Thirdly, it is recommended to include typhoons that enter the landslide and total impact boundary but do not make landfall for the impact forecast of consecutive hazards.

A next study can include the expansion of this prototype of the hybrid model with other secondary hazards of typhoons, e. g. storm surges which are currently included via static input, similar to that of the landslide (Teklesadik and Riquet, 2021). A similar approach as for the creation of the landslide IBF model can be used to add more consecutive hazards.

An approach to create a 'proper' multi-hazard impact forecasting model (section 5.2.2) can relate the current landslide impact forecast to the ML model of the 510 typhoon model. This requires either training of the variables in the of the existing ML model with additional vulnerability indicators of buildings for landslides (Cabrera et al., 2021; Gill and Malamud, 2014; King, 2021) or the creation of a new ML model for landslides. This approach allows for the inclusion of landslide and typhoon hazard-impact relations and improves the total impact forecast.

### 7.3 NEXT STEPS AND FUTURE RESEARCH

To provide substantial insights in the improved reliability of the impact forecast by the hybrid model, more typhoon events must be assessed. The hybrid model can easily be implemented in other regions or up-scaled to the entire Philippines to include case studies both within and outside of the Bicol region. These reliability assessments can include multiple accuracy statistics (e. g. F1, MCC and TSS as explained in section C.3) to draw reliable conclusions on the performance of the hybrid model.

The assessment of typhoon events requires typhoon-specific landslide inventories, which are scarce. It is recommended to complete existing inventories in the Philippines both for the validation of the hybrid model and for the creation of the ED thresholds or other landslide hazard indicators. Reliable landslide inventories can take years to construct, therefore it is recommended to cooperate and contribute to both local and global initiatives, e. g. LHASA (NASA, 2021). Satellite analyses, e. g. landslide detection using QGIS or Google Earth Engine, can provide for a suitable method to complete inventories as well (Guzzetti et al., 2012; Van Westen et al., 2008).

Lastly, it is advised to collaborate with ongoing local projects, e. g. the START network in the Bicol region, to combine efforts and create reliable landslide and multi-hazard forecasting models. Involving stakeholders and end-users of an IBF model improves its reliability, implementation and response (Bierens et al., 2020; Van den Homberg et al., 2020; WMO, 2015). For a most reliable multi-hazard IBF model, experts can provide advice prior to the design and creation of the model. Expert-based judgement on the reliability of IBF models and chains can provide for a method to determine on which reliability elements modellers can focus to create the most reliable IBF model possible. This method can be found in Appendix D.

## BIBLIOGRAPHY

---

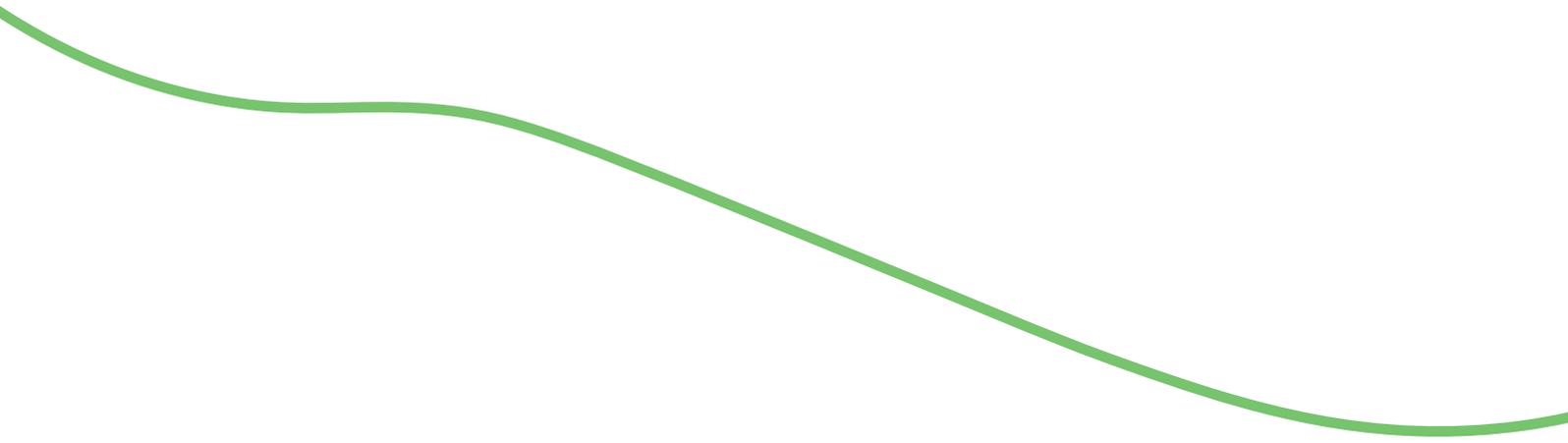
- Acosta, L. A., Eugenio, E. A., Macandog, P. B. M., and Magcale-macandog, D. B. (2016). "Loss and damage from typhoon-induced floods and landslides in the Philippines : community perceptions on climate impacts and adaptation options Elaine Kuan-Hui Lin Edwin R . Abucay and Alfi Lorenz Cura Mary Grace Primavera." In: *International Journal of Global Warming* 9.1.
- Bierens, S., Boersma, K., and van den Homberg, M. J. (2020). "The legitimacy, accountability, and ownership of an impact-based forecasting model in disaster governance." In: *Social Inclusion* 8.4, pp. 445–455. ISSN: 21832803. DOI: [10.17645/pag.v8i4.3161](https://doi.org/10.17645/pag.v8i4.3161).
- Bogaard, T. A. and Greco, R. (2016). "Landslide hydrology: from hydrology to pore pressure." In: *Wiley Interdisciplinary Reviews: Water* 3.3, pp. 439–459. ISSN: 20491948. DOI: [10.1002/wat2.1126](https://doi.org/10.1002/wat2.1126).
- Bogaard, T. and Greco, R. (2018). "Invited perspectives: Hydrological perspectives on precipitation intensity-duration thresholds for landslide initiation: Proposing hydro-meteorological thresholds." In: *Natural Hazards and Earth System Sciences* 18.1, pp. 31–39. ISSN: 16849981. DOI: [10.5194/nhess-18-31-2018](https://doi.org/10.5194/nhess-18-31-2018).
- Brunelli, M. (2015). *Introduction to the Analytical Hierarchy Process*. ISBN: 9783319125015.
- Brunetti, M. T., Melillo, M., Gariano, S. L., Ciabatta, L., Brocca, L., Amarnath, G., and Peruccacci, S. (2021). "Satellite rainfall products outperform ground observations for landslide prediction in India." In: *Hydrology and Earth System Sciences* 25.6, pp. 3267–3279. ISSN: 16077938. DOI: [10.5194/hess-25-3267-2021](https://doi.org/10.5194/hess-25-3267-2021).
- CMA and GFDDR (2016). *Implementing Multi-Hazard Impact-based Forecast and Warning Services*. Tech. rep. Shanghai: *Global Facility for Disaster Reduction and Recovery; China Meteorological Administration*.
- Cabrera, M. A., Prada-sarmiento, L. F., van Westen, C. J., Augusto, F., Arevalo, F., and van den Bout, B. (2021). "Challenges in analyzing landslide risk dynamics for risk reduction planning." In: *International society for soil mechanics* June 2020.
- Chacón, J., Irigaray, C., Fernández, T., and El Hamdouni, R. (2006). "Engineering geology maps: Landslides and geographical information systems." In: *Bulletin of Engineering Geology and the Environment* 65.4, pp. 341–411. ISSN: 14359529. DOI: [10.1007/s10064-006-0064-z](https://doi.org/10.1007/s10064-006-0064-z).
- Chicco, D. and Jurman, G. (2020). "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation." In: *BMC Genomics* 21.1, pp. 1–13. ISSN: 14712164. DOI: [10.1186/s12864-019-6413-7](https://doi.org/10.1186/s12864-019-6413-7).
- Chung, M. C., Tan, C. H., and Chen, C. H. (2017). "Local rainfall thresholds for forecasting landslide occurrence: Taipingshan landslide triggered by Typhoon Saola." In: *Landslides* 14.1, pp. 19–33. ISSN: 16125118. DOI: [10.1007/s10346-016-0698-2](https://doi.org/10.1007/s10346-016-0698-2). URL: <http://dx.doi.org/10.1007/s10346-016-0698-2>.
- Ciavolella, M., Bogaard, T., Gargano, R., and Greco, R. (2016). "Is there Predictive Power in Hydrological Catchment Information for Regional Landslide Hazard Assessment?" In: *Procedia Earth and Planetary Science* 16, pp. 195–203. ISSN: 18785220. DOI: [10.1016/j.proeps.2016.10.021](https://doi.org/10.1016/j.proeps.2016.10.021). URL: <http://dx.doi.org/10.1016/j.proeps.2016.10.021>.

- Corominas, J. et al. (2014). "Recommendations for the quantitative analysis of landslide risk." In: *Bulletin of Engineering Geology and the Environment* 73.2, pp. 209–263. ISSN: 14359529. DOI: [10.1007/s10064-013-0538-8](https://doi.org/10.1007/s10064-013-0538-8).
- DSWD (2016a). *DROMIC Report 13 on Typhoon "KAREN" (SARIKA)*. Tech. rep. October 18. Department of Social Welfare and Development DSWD, pp. 1–10.
- DSWD (2016b). *DROMIC Report 9 on Typhoon "KAREN" (SARIKA)*. Tech. rep. October 16. Department of Social Welfare and Development DSWD, pp. 1–17.
- De Ruiter, M. C., Couasnon, A., van den Homberg, M. J., Daniell, J. E., Gill, J. C., and Ward, P. J. (2020). "Why We Can No Longer Ignore Consecutive Disasters." In: *Earth's Future* 8.3. ISSN: 23284277. DOI: [10.1029/2019EF001425](https://doi.org/10.1029/2019EF001425).
- Eco, R. et al. (2015). "Landslide and debris flow susceptibility mapping of Leyte Province, Philippines using remote sensing, numerical modelling, and GIS." In: *Journal of the Philippine Geoscience and Remote Sensing Society* 1.1. URL: [https://www.researchgate.net/publication/317065288\\_Landslide\\_and\\_debris\\_flow\\_susceptibility\\_mapping\\_of\\_Leyte\\_Province\\_Philippines\\_using\\_remote\\_sensing\\_numerical\\_modelling\\_and\\_GIS](https://www.researchgate.net/publication/317065288_Landslide_and_debris_flow_susceptibility_mapping_of_Leyte_Province_Philippines_using_remote_sensing_numerical_modelling_and_GIS).
- Fano, J. A., Alpasan, M. T., Mitsunaga, T., and Tokunaga, Y. (2007). "The Mayon 2006 Debris Flows: The Destructive Path of Typhoon Reming." In: 3.October.
- Feizizadeh, B., Kienberger, S., and Kamran, K. V. (2015). "Sensitivity and Uncertainty Analysis Approach for GIS-MCDA Based Economic Vulnerability Assessment." In: *GI Forum* 1, pp. 81–89. ISSN: 2308-1708. DOI: [10.1553/giscience2015s81](https://doi.org/10.1553/giscience2015s81).
- Gariano, S. L., Melillo, M., Peruccacci, S., and Brunetti, M. T. (2020). "How much does the rainfall temporal resolution affect rainfall thresholds for landslide triggering?" In: *Natural Hazards* 100.2, pp. 655–670. ISSN: 15730840. DOI: [10.1007/s11069-019-03830-x](https://doi.org/10.1007/s11069-019-03830-x). URL: <https://doi.org/10.1007/s11069-019-03830-x>.
- Georisk (2021). *GeoRiskPH: Innovations for resilience*. URL: <https://www.georisk.gov.ph/>.
- Gill, J. C. and Malamud, B. D. (2014). "Reviewing and visualizing the interactions of natural hazards." In: *Reviews of Geophysics* 52, pp. 680–722. ISSN: 23249250. DOI: [10.1002/2013RG000445](https://doi.org/10.1002/2013RG000445).
- Guzzetti, F., Gariano, S. L., Peruccacci, S., Brunetti, M. T., Marchesini, I., Rossi, M., and Melillo, M. (2020). "Geographical landslide early warning systems." In: *Earth-Science Reviews* 200. September 2019, p. 102973. ISSN: 00128252. DOI: [10.1016/j.earscirev.2019.102973](https://doi.org/10.1016/j.earscirev.2019.102973). URL: <https://doi.org/10.1016/j.earscirev.2019.102973>.
- Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M., and Chang, K. T. (2012). "Landslide inventory maps: New tools for an old problem." In: *Earth-Science Reviews* 112.1-2, pp. 42–66. ISSN: 00128252. DOI: [10.1016/j.earscirev.2012.02.001](https://doi.org/10.1016/j.earscirev.2012.02.001). URL: <http://dx.doi.org/10.1016/j.earscirev.2012.02.001>.
- Horton, P., Jaboyedoff, M., Rudaz, B., and Zimmermann, M. (2013). "Flow-R, a model for susceptibility mapping of debris flows and other gravitational hazards at a regional scale." In: *Natural Hazards and Earth System Sciences* 13.4, pp. 869–885. ISSN: 16849981. DOI: [10.5194/nhess-13-869-2013](https://doi.org/10.5194/nhess-13-869-2013).
- Hsu, T. H. and Pan, F. F. (2009). "Application of Monte Carlo AHP in ranking dental quality attributes." In: *Expert Systems with Applications* 36.2 PART 1, pp. 2310–2316. ISSN: 09574174. DOI: [10.1016/j.eswa.2007.12.023](https://doi.org/10.1016/j.eswa.2007.12.023). URL: <http://dx.doi.org/10.1016/j.eswa.2007.12.023>.
- IFRC (2016a). *DREF operation Philippines: Typhoon Sarika*. Tech. rep. December. International Federation of the Red Cross.

- IFRC (2016b). *Information bulletin Philippines: Typhoon Sarika*. Tech. rep. October. International Federation of the Red Cross, pp. 2–4.
- IFRC (2021). *Glossary of Terms for Forecast-based Financing*. International Federation of the Red Cross, pp. 1–8.
- JAXA (2021). *JAXA Global Rainfall Watch*. Japan Aerospace Exploration Agency. URL: <https://sharaku.eorc.jaxa.jp/GSMaP/>.
- Jaboyedoff, M. and Labiouse, V. (2011). “Preliminary estimation of rockfall runout zones.” In: *Natural Hazards and Earth System Science* 11.3, pp. 819–828. ISSN: 15618633. DOI: [10.5194/nhess-11-819-2011](https://doi.org/10.5194/nhess-11-819-2011).
- Joshi, A., Grover, J., Prasanna Kanungo, D., and Kumar Panigrahi, R. (2020). “Real-time Landslide Monitoring, Detection and Early Warning System for Tangni Landslide.” In: *SSRN Electronic Journal*, pp. 1–6. ISSN: 1556-5068. DOI: [10.2139/ssrn.3511001](https://doi.org/10.2139/ssrn.3511001).
- King, G. (2021). “Multi-Hazard Interrelationships.” In: January.
- Luzon, P. K., Montalbo, K., Galang, J., Sabado, J. M., Escape, C. M., Felix, R., and Lagmay, A. M. F. (2016). “Hazard mapping related to structurally controlled landslides in Southern Leyte, Philippines.” In: *Natural Hazards and Earth System Sciences* 16.3, pp. 875–883. ISSN: 16849981. DOI: [10.5194/nhess-16-875-2016](https://doi.org/10.5194/nhess-16-875-2016).
- Melillo, M., Brunetti, M. T., Peruccacci, S., Gariano, S. L., and Guzzetti, F. (2016). “Rainfall thresholds for the possible landslide occurrence in Sicily (Southern Italy) based on the automatic reconstruction of rainfall events.” In: *Landslides* 13.1, pp. 165–172. ISSN: 16125118. DOI: [10.1007/s10346-015-0630-1](https://doi.org/10.1007/s10346-015-0630-1).
- Melillo, M., Brunetti, M. T., Peruccacci, S., Gariano, S. L., Roccati, A., and Guzzetti, F. (2018). “A tool for the automatic calculation of rainfall thresholds for landslide occurrence.” In: *Environmental Modelling and Software* 105, July, pp. 230–243. ISSN: 13648152. DOI: [10.1016/j.envsoft.2018.03.024](https://doi.org/10.1016/j.envsoft.2018.03.024). URL: <https://doi.org/10.1016/j.envsoft.2018.03.024>.
- Monsieurs, E., Dewitte, O., and Demoulin, A. (2019). “A susceptibility-based rainfall threshold approach for landslide occurrence.” In: *Natural Hazards and Earth System Sciences* 19.4, pp. 775–789. ISSN: 16849981. DOI: [10.5194/nhess-19-775-2019](https://doi.org/10.5194/nhess-19-775-2019).
- Monteverde, M. C. A., Thema, A. C., and Lorenzo, A. M. (2020). “Multi-Hazard Early Warning System ( MHEWS ) in the Philippines.” In: *Workshop on Strengthening Multi-Hazard Early Warning Systems (MHEWS) and Early Actions (EA) in Southeast Asia (SeA)*. February. DOST-PAGASA. URL: [https://www.unescap.org/sites/default/files/2a\\_LorenzoMoron\\_MultiHazardEarlyWarningSysteminthePhilippines.pdf](https://www.unescap.org/sites/default/files/2a_LorenzoMoron_MultiHazardEarlyWarningSysteminthePhilippines.pdf).
- NASA (2021). *Cooperative Open Online Landslide Repository (COOLR) project*. URL: <https://gpm.nasa.gov/landslides/about.html>.
- NOAH (2021). *Nationwide Operational Assessment of Hazards (NOAH)*. URL: <http://noah.up.edu.ph/#/>.
- Nadim, F. and Intrieri, E. (2011). “Early Warning Systems for Landslides : Challenges and New Monitoring Technologies.” In: pp. 1–15.
- PH government (2008). *Typhoon Karen leaves 6 dead, 2 injured and millions worth of damages to crops and properties in Benguet*. URL: <https://reliefweb.int/report/philippines/philippines-typhoon-karen-leaves-6-dead-2-injured-and-millions-worth-damages>.
- PRC (2019). *Forecast-Based Financing Early Action Protocol TYPHOON Philippines*. Philippines Red Cross.

- Peres, D. J., Cancelliere, A., Greco, R., and Bogaard, T. A. (2018). "Influence of uncertain identification of triggering rainfall on the assessment of landslide early warning thresholds." In: *Natural Hazards and Earth System Sciences* 18.2, pp. 633–646. ISSN: 16849981. DOI: [10.5194/nhess-18-633-2018](https://doi.org/10.5194/nhess-18-633-2018).
- Piciullo, L., Gariano, S. L., Melillo, M., Brunetti, M. T., Peruccacci, S., Guzzetti, F., and Calvello, M. (2017). "Definition and performance of a threshold-based regional early warning model for rainfall-induced landslides." In: *Landslides* 14.3, pp. 995–1008. ISSN: 16125118. DOI: [10.1007/s10346-016-0750-2](https://doi.org/10.1007/s10346-016-0750-2).
- Rabonza, M. L., Felix, R. P., Lagmay, A. M. F. A., Eco, R. N. C., Ortiz, I. J. G., and Aquino, D. T. (2016). "Shallow landslide susceptibility mapping using high-resolution topography for areas devastated by super typhoon Haiyan." In: *Landslides* 13.1, pp. 201–210. ISSN: 16125118. DOI: [10.1007/s10346-015-0626-x](https://doi.org/10.1007/s10346-015-0626-x).
- Rawat, S. (2019). *Is accuracy everything?* URL: <https://towardsdatascience.com/is-accuracy-everything-96da9afd540d>.
- Roccati, A., Paliaga, G., Luino, F., Faccini, F., and Turconi, L. (2021). "GIS-Based Landslide Susceptibility Mapping for Land Use Planning and Risk Assessment." In.
- Segoni, S., Piciullo, L., and Gariano, S. L. (2018). "Preface: Landslide early warning systems: Monitoring systems, rainfall thresholds, warning models, performance evaluation and risk perception." In: *Natural Hazards and Earth System Sciences* 18.12, pp. 3179–3186. ISSN: 16849981. DOI: [10.5194/nhess-18-3179-2018](https://doi.org/10.5194/nhess-18-3179-2018).
- Sidle, R. C. and Bogaard, T. A. (2016). "Dynamic earth system and ecological controls of rainfall-initiated landslides." In: *Earth-Science Reviews* 159, pp. 275–291. ISSN: 00128252. DOI: [10.1016/j.earscirev.2016.05.013](https://doi.org/10.1016/j.earscirev.2016.05.013). URL: <http://dx.doi.org/10.1016/j.earscirev.2016.05.013>.
- Sisters, L. (2020). *Matthews Correlation Coefficient: when to use it and when to avoid it*. URL: <https://towardsdatascience.com/matthews-correlation-coefficient-when-to-use-it-and-when-to-avoid-it-310b3c923f7e>.
- Start Network (2021). *Application form : tools to support anticipation programming*. Tech. rep.
- TSR (2007). *TSR Adds Rainfall Forecasts for Tropical Storms Worldwide*. February. *Tropical Storm Risk*, pp. 1–3.
- Tate, E. (2012). "Social vulnerability indices: A comparative assessment using uncertainty and sensitivity analysis." In: *Natural Hazards* 63.2, pp. 325–347. ISSN: 0921030X. DOI: [10.1007/s11069-012-0152-2](https://doi.org/10.1007/s11069-012-0152-2).
- Teklesadik, A, van den Homberg, M, Wagenaar, D, Visser, J, Riquet, D, Margutti, J, and Giodini, S (2022). "Real time forecasting of tropical cyclones damage in a humanitarian context [Unpublished manuscript]."
- Teklesadik, A. and Riquet, D. (2021). *Forecasting the humanitarian impact of typhoons with machine learning; The Philippines. Dutch Red Cross 510; Philippines Red Cross; German Red Cross*.
- Thirugnanam, H., Ramesh, M. V., and Rangan, V. P. (2020). "Enhancing the reliability of landslide early warning systems by machine learning." In: *Landslides* 17.9, pp. 2231–2246. ISSN: 16125118. DOI: [10.1007/s10346-020-01453-z](https://doi.org/10.1007/s10346-020-01453-z).
- Tunkiel, A. T., Sui, D., and Wiktorski, T. (2020). "Data-driven sensitivity analysis of complex machine learning models: A case study of directional drilling." In: *Journal of Petroleum Science and Engineering* 195, July, p. 107630. ISSN: 09204105. DOI: [10.1016/j.petrol.2020.107630](https://doi.org/10.1016/j.petrol.2020.107630). URL: <https://doi.org/10.1016/j.petrol.2020.107630>.

- UKMO (2018). *The Future of Forecasts: Impact-based Forecasting for Early Action*. United Kingdom Met Office. URL: <https://www.510.global/impact-based-forecast/>.
- Van Brussel, M., Teklesadik, A., van den Homberg, M., Margutti, J., and Riquet, D. (2021). *Typhoon prediction model: Rice losses*.
- Van Westen, C. J., Chen, L., and Hussin, H. Y. (2014). "Susceptibility and run-out modeling on a regional scale Analysis and Management of Changing Risks for Natural Hazards." In: November. DOI: [10.13140/2.1.1441.1844](https://doi.org/10.13140/2.1.1441.1844).
- Van Westen, C. J., van Asch, T. W., and Soeters, R. (2006). "Landslide hazard and risk zonation - Why is it still so difficult?" In: *Bulletin of Engineering Geology and the Environment* 65.2, pp. 167–184. ISSN: 14359529. DOI: [10.1007/s10064-005-0023-0](https://doi.org/10.1007/s10064-005-0023-0).
- Van Westen, C. J., Castellanos, E., and Kuriakose, S. L. (2008). "Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview." In: *Engineering Geology* 102.3-4, pp. 112–131. ISSN: 00137952. DOI: [10.1016/j.enggeo.2008.03.010](https://doi.org/10.1016/j.enggeo.2008.03.010). URL: <http://dx.doi.org/10.1016/j.enggeo.2008.03.010>.
- Van den Homberg, M. J., Gevaert, C. M., and Georgiadou, Y. (2020). "The changing face of accountability in humanitarianism: Using artificial intelligence for anticipatory action." In: *Social Inclusion* 8.4, pp. 456–467. ISSN: 21832803. DOI: [10.17645/pag.v8i4.3158](https://doi.org/10.17645/pag.v8i4.3158).
- WMO (2015). *WMO Guidelines on Multi-hazard Impact-based Forecast and Warning Services*. 1150. World Meteorological Organization, p. 23. ISBN: 9781457706691.
- WMO; (2017). "Multi-hazard Early Warning Systems: A Checklist. Outcome of the first Multi-hazard Early Warning Conference 22 to 23 May - Cancún, México." In: May, p. 20. URL: [https://library.wmo.int/doc\\_num.php?explnum\\_id=4463](https://library.wmo.int/doc_num.php?explnum_id=4463).
- Wagenaar, D, Curran, A, Balbi, M, Bhardwaj, A, Soden, R, Hartato, E, Mestav Sarica, G, Ruangpan, L, Molinario, G, and Lallemand, D (2020). "Invited perspectives: How machine learning will change flood risk and impact assessment." In: *Natural Hazards and Earth System Sciences* 20.4, pp. 1149–1161. ISSN: 16849981. DOI: [10.5194/nhess-20-1149-2020](https://doi.org/10.5194/nhess-20-1149-2020).
- Waidyanatha, N. (2010). "Towards a typology of integrated functional Early Warning Systems." In: *International Journal of Critical Infrastructures* 6.1, pp. 31–51. ISSN: 14753219. DOI: [10.1504/IJCIS.2010.029575](https://doi.org/10.1504/IJCIS.2010.029575).
- Wilkinson, E., Weingärtner, L., Choularton, R., Bailey, M., Todd, M., Kniveton, D., and Venton, C. C. (2018). "Forecasting hazards, averting disasters: Implementing forecast-based early action at scale." In: March, p. 40. URL: <https://www.odi.org/sites/odi.org.uk/files/resource-documents/12104.pdf>.
- Xie, Z., Gao, X., He, J., and Feng, C. (2016). "Evaluating rural low-carbon communities: A study of Guangdong Province, China." In: *Energy and Buildings* 133, pp. 777–789. ISSN: 03787788. DOI: [10.1016/j.enbuild.2016.10.042](https://doi.org/10.1016/j.enbuild.2016.10.042). URL: <http://dx.doi.org/10.1016/j.enbuild.2016.10.042>.
- Zamudio, N. C and Orogo, A. B. (2021). *The establishment of Rainfall Threshold for Landslide in Albay and Sorsogon*. Tech. rep. PCR, ACTED, PLAN, Start Network, ADRA, HI, PAGASA.
- Zhang, Q. et al. (2019). "Increasing the value of weather-related warnings." In: *Science Bulletin* 64.10, pp. 647–649. ISSN: 20959281. DOI: [10.1016/j.scib.2019.04.003](https://doi.org/10.1016/j.scib.2019.04.003). URL: <https://doi.org/10.1016/j.scib.2019.04.003>.
- Zhang, X. Y., Trame, M. N., Lesko, L. J., and Schmidt, S. (2015). "Sobol sensitivity analysis: A tool to guide the development and evaluation of systems pharmacology models." In: *CPT: Pharmacometrics and Systems Pharmacology* 4.2, pp. 69–79. ISSN: 21638306. DOI: [10.1002/psp4.6](https://doi.org/10.1002/psp4.6).





## APPENDIX A: RELIABILITY ELEMENTS

| <b>Data collection</b> (assessed per data source) |   | References                      |
|---|---|---------------------------------|
| Data availability                                 | The availability of data, such as rain gauge data, satellite data, vulnerability data, exposure data;   | 3, 4, 5, 6, 7, 10, 11, 12, 13   |
| Data completeness                                 | Measure of the completeness of the data; non-zero values  | 1, 2, 4, 5, 6, 7, 11, 12, 13    |
| Data quality                                      | Low uncertainty and signal-to-noise, high accuracy and precision  | 1, 2, 4, 5, 6, 7, 11, 12, 13    |
| Data recency                                      | Data is of the most recent source; Data is recently updated; Data is updated during the course of the hazard event (e.g. typhoon tracks/hazard forecast changes during the hazard)  | 1, 2, 4, 5, 6, 11, 13           |
| Data resolution                                   | Resolution of the data should be as high as possible and applicable for the aim of the forecast; Resolutions of data should be compatible; Scales should be applicable for the next component/elements  | 1, 2, 5, 6, 11, 13              |
| Fail-safe system & transmission                   | Technical resilience of the system such that data is transmittable also during power/network/communication outage   | 4, 5, 7, 8                      |
| Number of sources / resilience                    | Multiple sources can be used for the same data collection to increase resilience; reduce dependency on one source   | 7, 9, 10, 11                    |
| Source location                                   | Physical locations for sources should be hazard-, vandalism- and animal-prone; Online sources should be freely available and the source should be trusted   | 5, 7, 8, 9, 11                  |
| Source type                                       | The source type should be considered (survey/satellite/sensor/news); the format should be considered and compatible with other data sources; The time for data collection should be short; The number of parameters should be enough and redundant parameters/input should be discarded | 1, 2, 4, 5, 7, 10, 11, 12, 13   |
| <b>Hazard forecast</b>                            |   |                                 |
| Computation time                                  | Lead time should be as large as possible and increases with a low computation time of the hazard modelling  | 5, 11, 13                       |
| Hazard forecasting method                         | The forecasting method for hazards should be considered (probabilistic/statistical; real-time/forecasting); The hazard forecasting method results in a (dynamic) hazard map including features applicable for IBF; Multiple methods can be applied to increase reliability              | 7, 10, 13                       |
| Hazard forecasting thresholds                     | Thresholds should be validated; Multiple thresholds for hazard assessment can be considered; The threshold should be applicable to the region and type of hazard  | 2, 3, 4, 5, 6, 11, 13           |
| Hazard map scale                                  | The zonations of hazard (low-medium-high) should be clear and applicable to the region; The scale of the hazard map should be compatible and applicable for the impact forecast   | 2, 3, 4, 5, 6, 7, 8, 11, 12, 13 |
| Model verification                                | The model should be tested and validated; Test runs should be useable as alternate source of information; Landslide and impact inventories are available  | 5, 9, 11, 13                    |
| Model uncertainty                                 | The uncertainty should be included in the outcome of the hazard forecast; The uncertainty should be as low as possible  | 2, 5, 6, 9, 11, 12, 13          |
| Modelling process & transmission                  | Modelling method should be simple such that changes can be easily incorporated; Models should be understandable for stakeholders; Models should have a high technical resilience  | 5, 7, 9, 10, 11, 13             |

| <b>Impact forecast</b>        |   | Sources                          |
|-------------------------------|---|----------------------------------|
| Computation time              | Lead time should be as large as possible and increases with a low computation time of the impact modelling  | 3, 7, 11, 13                     |
| Hazard-impact relations       | Impact and hazard interactions should be applicable and can change during the cause of the event; Consecutive hazards and impact should/should not be considered  | 2, 3, 5, 12, 13                  |
| Impact forecasting method     | The forecasting method for impact should be considered (Impact-based forecasting; real-time/forecasting; qualitative/quantitative analyses); The impact forecasting method results in a warning/trigger which is applicable to for the region; The IBF method is considered; Multiple methods can be applied to increase accuracy | 3, 4, 5, 6, 10, 11, 12, 13       |
| Impact forecasting thresholds | Multiple thresholds for impact assessment and zonations should be considered (direct/indirect risk; applicable elements at risk); the thresholds should be applicable to the region and type of hazard; vulnerability; coping capacity; exposure  | 3, 7, 10, 11                     |
| Impact map scale/zonations    | The scale/zonation of impact (low-medium-high) should be clear and applicable to the region; The zonations should be applicable for the (multiple) levels of warning  | 2, 3, 5, 6, 7, 13                |
| Model verification            | The model should be tested and validated; Test runs should be usable as alternate source of information; The uncertainty should be included in the outcome of the impact forecast   | 9, 11, 13                        |
| Model uncertainty             | The uncertainty should be included in the outcome of the hazard forecast; The uncertainty should be as low as possible  | 2, 4, 5, 6, 7, 8, 11, 12, 13     |
| Modelling process             | Modelling method should be simple such that changes can be easily incorporated; Models should be understandable for stakeholders; Model should have a high technical resilience; The number of false positives and false negatives should be low  | 3, 4, 5, 7, 8, 9, 10, 11, 12, 13 |

Figure A.1: Terms explained for the first, second and third component in the EW chain, (Waidyanatha, 2010) (1), (Zhang et al., 2019) (2), (Wilkinson et al., 2018) (3), (WMO; 2017) (4), (UKMO, 2018) (5), (Van Westen et al., 2006) (6), (Thirugnanam et al., 2020) (7), (Joshi et al., 2020) (8), (Segoni et al., 2018) (9), (Nadim and Intrieri, 2011) (10), (Guzzetti et al., 2020) (11), (WMO; 2017) (12), (Cabrera et al., 2021) (13)

# B

## APPENDIX B: ADDITIONAL FILES FOR THE METHOD

---

### B.1 ELABORATION ON THE 510 TYPHOON MODEL

The local Philippine weather organisation contacts 510 five days before a typhoon is forecasted to make landfall. The Dutch team starts monitoring the typhoon forecast and runs the machine learning (ML) 510 typhoon model. The 510 typhoon model sends a warning when an impact threshold on damaged houses is exceeded at 72 hours before a typhoon makes landfall. After the warning, early actions in the form of early harvesting of crops; evacuation of livestock and assets; installation of house strengthening kits and basic needs provision through cash for work are set in motion. Over the course of the typhoon event, the impact forecast is updated every 24 hours. After the typhoon event, actual hazard and impact information of the event is collected to be compared with the forecasted impact in order to improve the ML model's performance (PRC, 2019; Teklesadik and Riquet, 2021).

#### B.1.1 *The ML model*

The 510 typhoon model applies a ML approach, in which the model is trained on historical typhoon data to form a relationship between the typhoon hazard, the vulnerability and its impact. The numbers 1 to 4 indicated in the grey boxes in figure 3.1 explain the ML approach. The ML model is created, firstly, by the collection of relevant data, e. g. geographical data (elevation, ruggedness, slope etc); vulnerability, exposure and coping capacity data (population density, number of households, roof types, wall types etc) and consecutive hazard maps (landslides and storm surges). This is indicated by the 1 in the grey box (step 1) and these variables are referred to as the *predictors* of the ML model. Then, for each historical typhoon, hazard related data (wind speed, rainfall, typhoon tracks from previous typhoons) and damage inventories are gathered such that the model can generate a relationship between the impact of a typhoon and the input variables (step 2 and 3). This data is combined by training the predictors and relating the collected hazard and impact indicators with actual impact data. Step 1 to 4 combined takes place within the *ML model* and the outcome is called the *predictant* of the ML model (Teklesadik et al., 2022; Teklesadik and Riquet, 2021).

The predictors and predictants remain the same, but can be updated when new historical data is available. The ML model is created in python and is used as a source of input for the 510 typhoon model (Teklesadik et al., 2022; Teklesadik and Riquet, 2021).

#### B.1.2 *The 510 typhoon forecasting model*

Five days prior to a typhoon event, the hazard forecasting variables for an approaching typhoon are collected (step 5). Wind speed and sustained wind duration as well as rainfall

data from external weather forecasts are obtained. The cumulative rainfall is calculated for each municipality. Wind and typhoon forecasts are downloaded from various different online directories (Teklesadik et al., 2022; Teklesadik and Riquet, 2021).

The hazard indicators and ML model are combined into the 510 forecasting typhoon model. This is a code written in the software Rstudio. After the data collection is completed, the hazard forecast map is created by using a wind speed parameterization model. First, the maximum wind speed and the maximum sustained wind speed for 1 minute are considered. Wind speed parameterization uses the sustained wind speed intensity to distinguish locations where wind speed exceeds 80 km/h. A typhoon track forecast is obtained from weather agencies and included into the wind model. These wind speeds are combined with the sustained wind speed, daily rainfall intensity, accumulated rainfall and the typhoon track forecast to form the hazard forecast. Per municipality, the distance from the forecasted typhoon track is calculated. This is done for different predictions (ensembles) of wind speed forecasts. The parameterization model is run for each ensemble member and the probability of a municipality being within a reach of 50km from the forecasted typhoon tracks is determined. The outcome is a map showing the probability of a municipality located within 50km and within the impact boundary of 100km of the typhoon track. Within a 50km range, the forecasted typhoon impact is more reliable compared to the 100km range. ?? (Teklesadik et al., 2022; Teklesadik and Riquet, 2021).

The impact forecast, step 6, is created by combining the probabilistic hazard map with the hazard-impact curve from the ML model. By comparing the hazard characteristics of the typhoon (wind and rainfall) with the hazard predictants of previous typhoons, the impact is assessed. It is assumed that similar hazard variables can create similar impact when occurring in a similar region. The impact for the forecasted typhoon is calculated and expressed in a percentage of damaged houses per municipality. If over 10% of the houses in at least three municipalities is completely damaged, the impact threshold is exceeded. The probability that the impact threshold is passed is calculated, resulting in a probabilistic impact map (Teklesadik et al., 2022; Teklesadik and Riquet, 2021).

For those municipalities where impact is forecasted, the absolute number of houses completely damaged is calculated. The probability of the hazard occurring and the number of buildings damaged are combined in a rated risk matrix, as applicable in IBF. The 510 typhoon model will issue a warning if high risk is forecasted for both the probability of the hazard and the absolute impact on houses (Teklesadik et al., 2022; Teklesadik and Riquet, 2021).

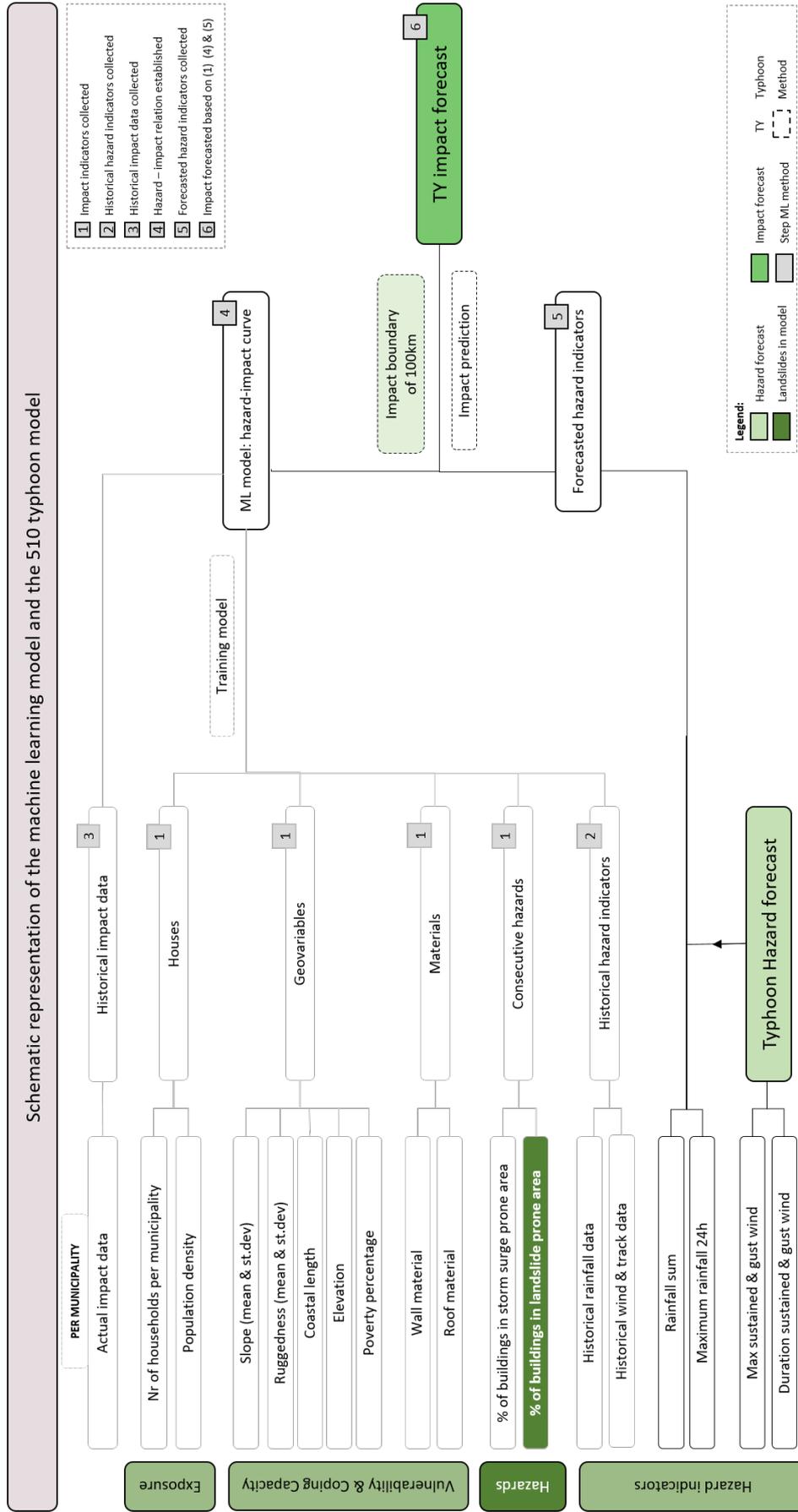


Figure B.1: Overview of the 510 typhoon model. Grey boxes indicate 4 steps to create a machine learning (ML) model. The ML model is combined with the 510 typhoon model (black outlined boxes) to create an impact forecast by combining hazard indicators with exposure, vulnerability and coping capacity (adjusted from Teklesadik and Riquet (2021)). The dark green box indicates the landslide component of the 510 typhoon model.

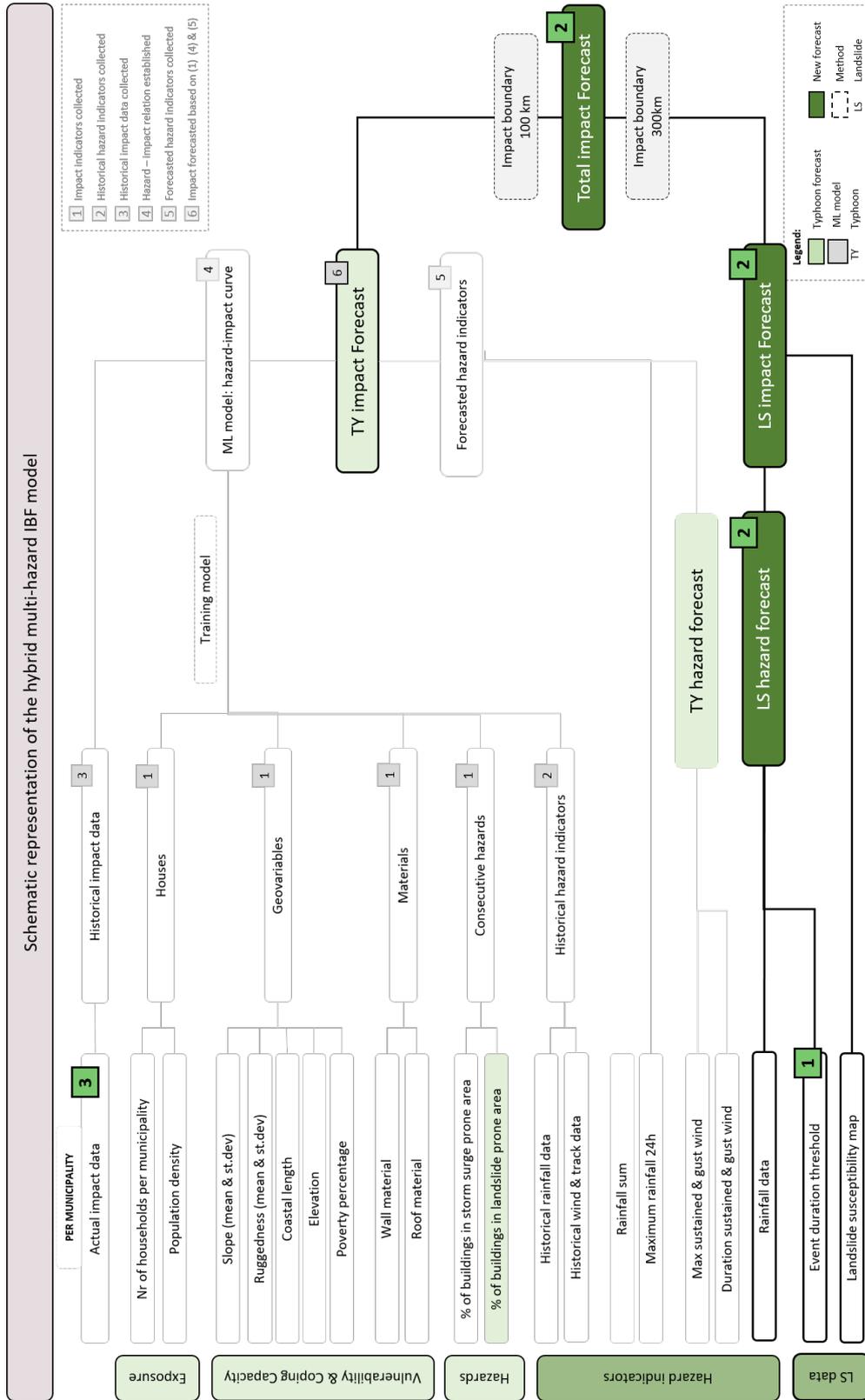


Figure B.2: Overview of the hybrid multi-hazard model in which the dark green box indicates the landslide component. Light grey and grey outlined boxes are not altered or used. Black outlined variables are applied or created in the design of the hybrid model. The bright green boxes indicate the step of the method, which correspond to figure 3.4.

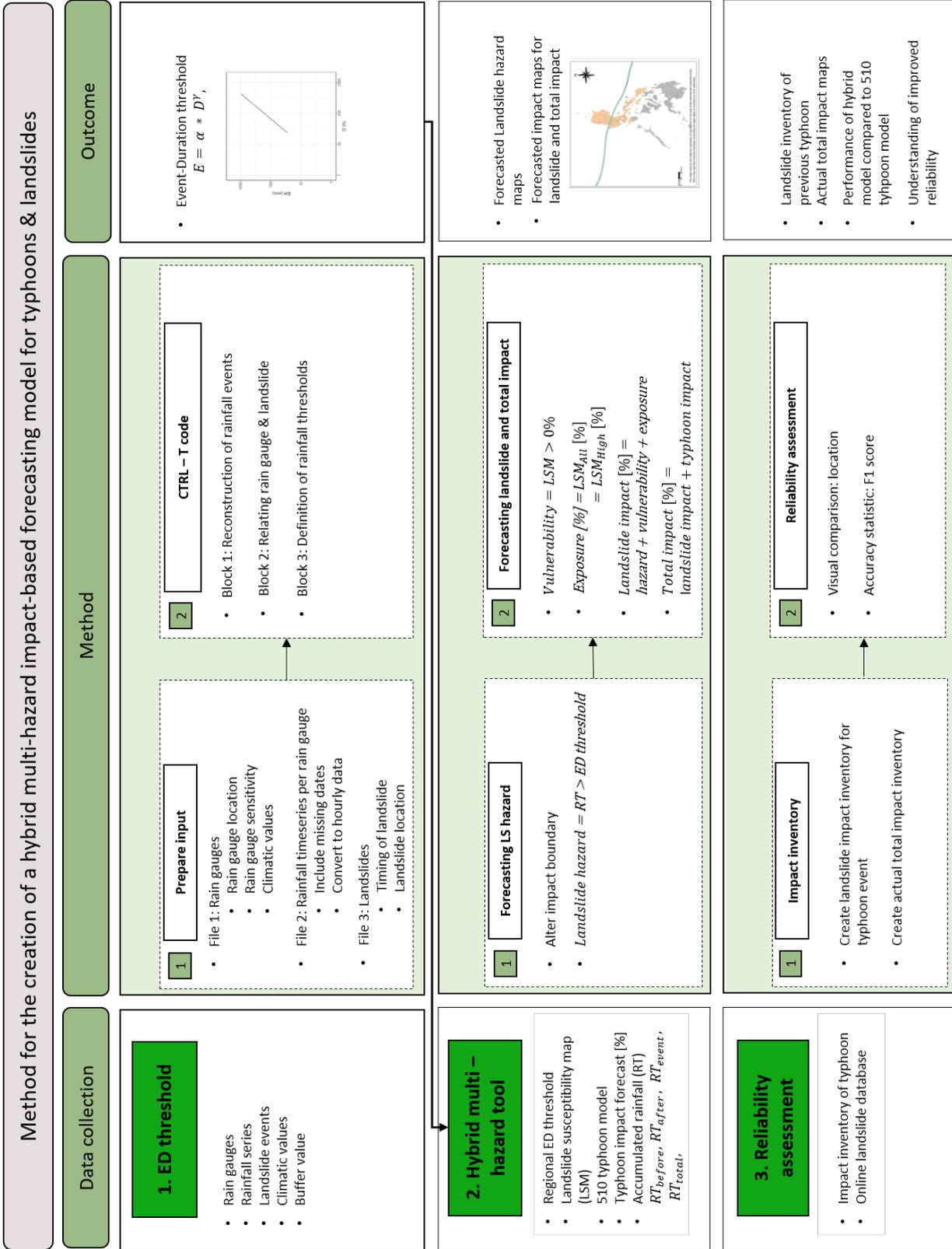
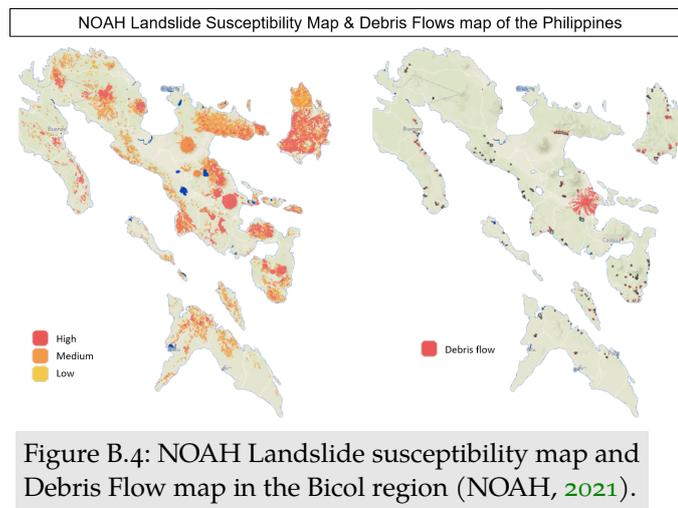


Figure B.3: Method overview showing three steps in the creation of a hybrid multi-hazard IBF model. Per step, the data collection, method and outcome are indicated. The steps correspond to the numbers indicated in figure 3-3.

## B.3 LANDSLIDE SUSCEPTIBILITY MAP

The NOAH project includes three landslide susceptibility maps (NOAH, 2021). The first map is called the '*Landslide Hazard Map*' and identifies the unstable slopes and landslide extent of shallow landslides and is a combination of two maps: the '*Landslide Hazard Map*' and the '*Unstable Slopes Map*'. In this research, the combination of these two maps are referred to as the Landslide susceptibility map (LSM). It is created by defining potential rock slide zones as high susceptibility zones. These zones are overlain with a stability index map resulting in the source areas of landslides presented in the '*Unstable Slopes*' map. This map is overlain with a runout map, of which the runout zones are classified as high susceptible. This results in the creation of the final landslide susceptibility map: the Landslide Hazard Map. The colours on the map indicate whether an area is safe to build infrastructure in (Eco et al., 2015; Jaboyedoff and Labiouse, 2011; Luzon et al., 2016; Rabonza et al., 2016). This map is coupled to buildings per municipality and used as input for the 510 typhoon model to identify areas prone to landslides. The map can be seen as a measure of vulnerability of a location to landslides (left map in figure B.4).

A second map, the '*Alluvial Fans Hazard Map*' shows debris flows and their spatial extent. It was constructed by identifying alluvial fans combined with a susceptibility map for debris flows, which included the runout of these flow (right map in figure B.4). This map is not included in the 510 typhoon model.



It should be noted, firstly, that the hazard maps do not include a time component to determine the probability or likelihood of landslides. Secondly, the '*Landslide Hazard Map*' includes both shallow landslides induced by typhoons and structurally controlled landslides. The thresholds which were used for these maps use a factor of safety of slope stability rather than a rainfall threshold. On top of that, deep-seated landslides and long run-out debris flows are not included. The '*Alluvial Fans Hazard Map*' contains a map of the susceptibility of debris flows only and is not accessible for free. Finally, none of the two maps include magnitude, frequency, volume or intensities of the landslides (Eco et al., 2015).

B.4 IMPACT BOUNDARY ANALYSIS

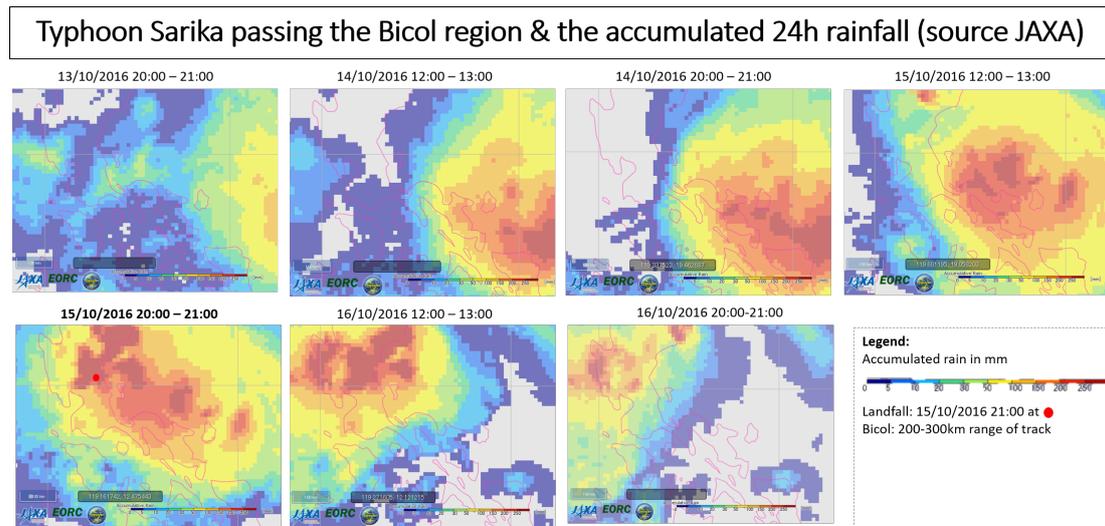


Figure B.5: Rainfall from typhoon Sarika passing over the Bicol region at 48 hours before and 24 hours after typhoon landfall, (JAXA, 2021).

The impact boundary is assessed for typhoon Sarika and is related to the rainfall extent of the typhoon, obtained from JAXA (2021). The rainfall associated with typhoon Sarika as it passes north of the Bicol region is significant and ranges between 200 to 300km distance of the track (figure B.5). Heavy rains occur before typhoon landfall at 15-10-2016 at 21:00 and exceed the current impact range of 100km. The time range of the JAXA rainfall patterns include 48 before typhoon landfall and 24 hours after (similar to the rainfall threshold of  $RT_{event}$ , without the last 24hours).

The impact boundary is evaluated using the actual landslide and typhoon impact. The rainfall extent during landfall considers approximately a range of 300km around the track indicated by the black arrow on the landfall location (red circle) in the left image of figure B.7. For typhoon Sarika, DROMIC information bulletins on typhoon Sarika DSWD (2016a,b) indicated a range of heavy rainfall associated with typhoon Sarika of at least 250km around the track. The typhoon and landslide impact of Sarika (right image) in figure C.4 exceeds the 100km distance as currently set in the 510 typhoon model (left image in figure B.7. The landslide impact is approximately located within a geographical extent of 300km from the track (middle figure).

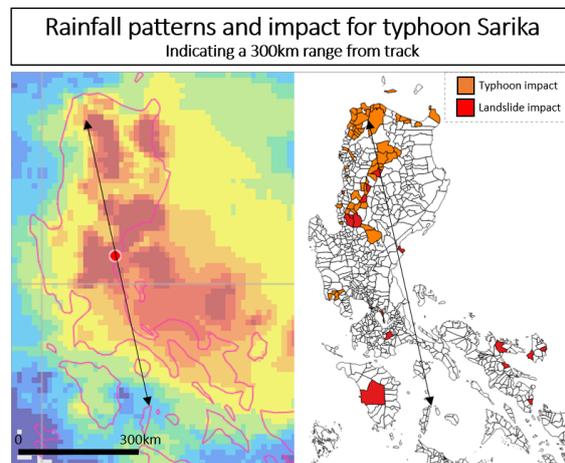


Figure B.6: Rainfall during landfall of typhoon Sarika and the actual landslide and typhoon impact, (DSWD, 2016a,b; IFRC, 2016a,b; JAXA, 2021; NASA, 2021; PH government, 2008).

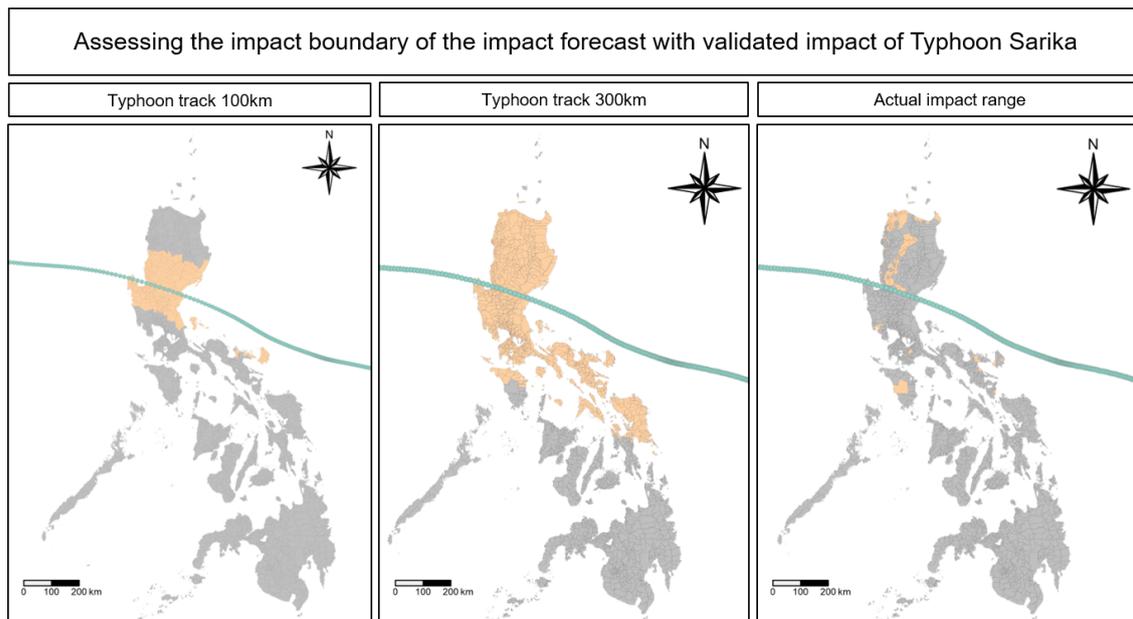


Figure B.7: Hazard and impact boundary of typhoon Sarika showing the original typhoon impact range of 100km range from the typhoon track (left), the impact boundary using a 300km range (middle) and the validated impact (right) (DSWD, 2016a,b; IFRC, 2016a,b; NASA, 2021; PH government, 2008).

## B.5 TYPHOONS DURIAN AND SARIKA

Typhoon Durian made landfall in the Bicol region on November 24 in 2006, after which it destroyed many houses located close to the track. Mudflows from the Mayon Volcano in the south of Bicol were triggered by the combined effect of the rains from the typhoon and the accumulation of dust from a recent eruption. Many mudslides impacted the houses located in the municipalities lowland of the volcano (Fano et al., 2007).

Typhoon Sarika passed the Bicol region a few days prior to its landfall on October 15th 2016. The impact of the typhoon was not comparable in magnitude with typhoon Durian, yet landslides occurred in various regions located outside the impact forecasting boundary, such as the Bicol region (DSWD, 2016a,b; IFRC, 2016a,b; PH government, 2008).

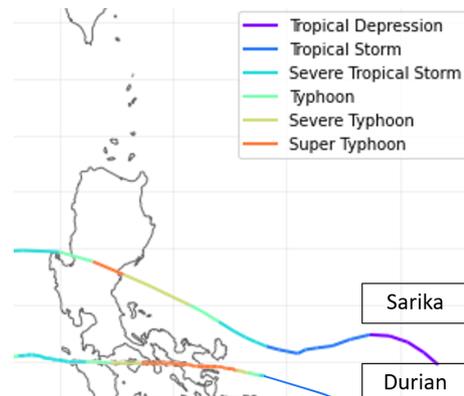


Figure B.8: Tracks of typhoon Durian and Sarika, obtained from the ML model of 510



## C.1 EVENT DURATION THRESHOLD

Three files are used as input for CTRL-T. In the rain gauges file, the rain gauge name, the location of the rain gauge, its sensitivity and climate values are included. The dry and wet parameters of the climate variables indicate the minimum number of hours between consecutive rain events in the wet or the dry months of a region and are based on the LANDSLIP study in India of Brunetti et al. (2021). The rainfall series for each rain gauge in Bicol is converted to hourly resolution. 59 landslide events are included. Any alterations of the default settings are shown in table C.2.

Table C.1: Input values for CTRL-T model altered from default settings for the Bicol region

| Rain gauges                   | Value       |
|-------------------------------|-------------|
| Time range rainfall           | 2006 - 2018 |
| Number of rain gauge stations | 3           |
| Rain gauge sensitivity [mm]   | 0.08        |
| Dry values [h]                | 24          |
| Wet values [h]                | 24          |
| Rainfall series               | Value       |
| Time range rainfall           | 2006 - 2018 |
| Rainfall resolution           | hourly      |
| Landslides                    | Value       |
| Time range landslide events   | 2006 - 2018 |
| Number of landslide events    | 59          |
| Timing of landslide event     | 23:59       |
| Code alteration               | Value       |
| Buffer value [km]             | 60          |

**ED rainfall thresholds of Bicol and the LANDSLIP study in India**

$$E_{Bicol} = (0.6 \pm 1.0) \times D^{(1.2 \pm 0.2)}$$

$$E_{India-rg} = (7.4 \pm 1.2) \times D^{(1.1 \pm 0.06)}$$

$$E_{India-sm} = (5.7 \pm 1.6) \times D^{(1.1 \pm 0.06)}$$

$$E_{India-pm} = (9.5 \pm 2.0) \times D^{(1.0 \pm 0.04)}$$

where *rg* corresponds to rainfall data obtained from rain gauges, *sm* from soil moisture data and *pm* by merging satellite data with soil moisture data. The method and data sets are explained in the study of Brunetti et al. (2021).

Table C.2: Threshold values for various non-exceedance probabilities (NEP) where LS above and below denote the number of landslide events above and below the threshold

| NEP [%] | $\alpha$ | $\Delta\alpha$ | $\gamma$ | $\Delta\gamma$ | LS above | LS below |
|---------|----------|----------------|----------|----------------|----------|----------|
| T1      | 0.32     | $\pm 0.6$      | 1.22     | $\pm 0.2$      | 45       | 4        |
| T2      | 0.41     | $\pm 0.7$      | 1.22     | $\pm 0.2$      | 45       | 4        |
| T5      | 0.59     | $\pm 1.0$      | 1.22     | $\pm 0.2$      | 42       | 7        |
| T10     | 0.83     | $\pm 1.4$      | 1.22     | $\pm 0.2$      | 38       | 11       |
| T20     | 1.3      | $\pm 2.1$      | 1.22     | $\pm 0.2$      | 31       | 18       |
| T50     | 2.8      | $\pm 4.4$      | 1.22     | $\pm 0.2$      | 12       | 37       |

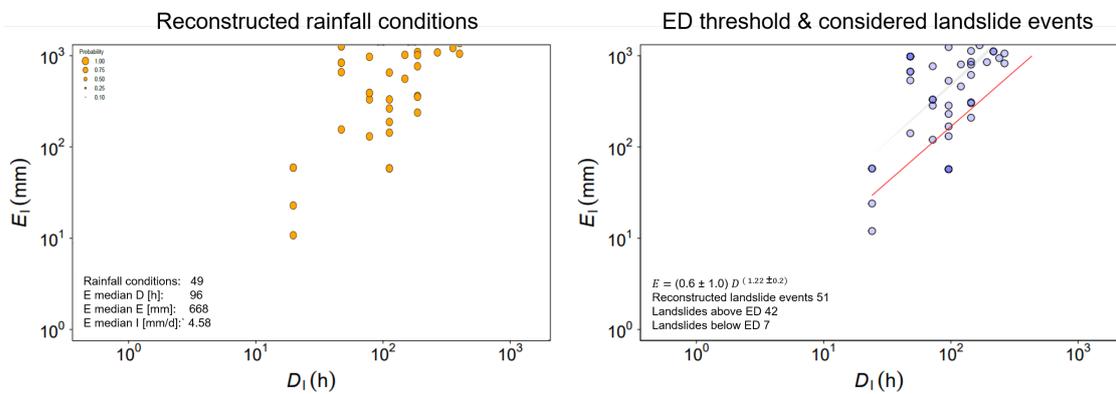


Figure C.1: Reconstructed landslide and rainfall conditions for the Bicol region

C.2 RAINFALL PATTERNS FOR TYPHOONS SARIKA AND DURIAN

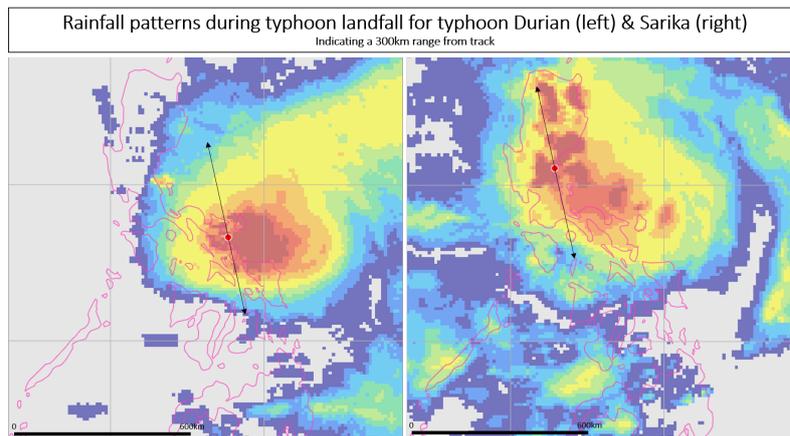


Figure C.2: The 24h accumulated rainfall of typhoon Durian and Sarika during landfall of the typhoon at the red circle, obtained from JAXA (2021). An impact range of 300 km is indicated by the arrow

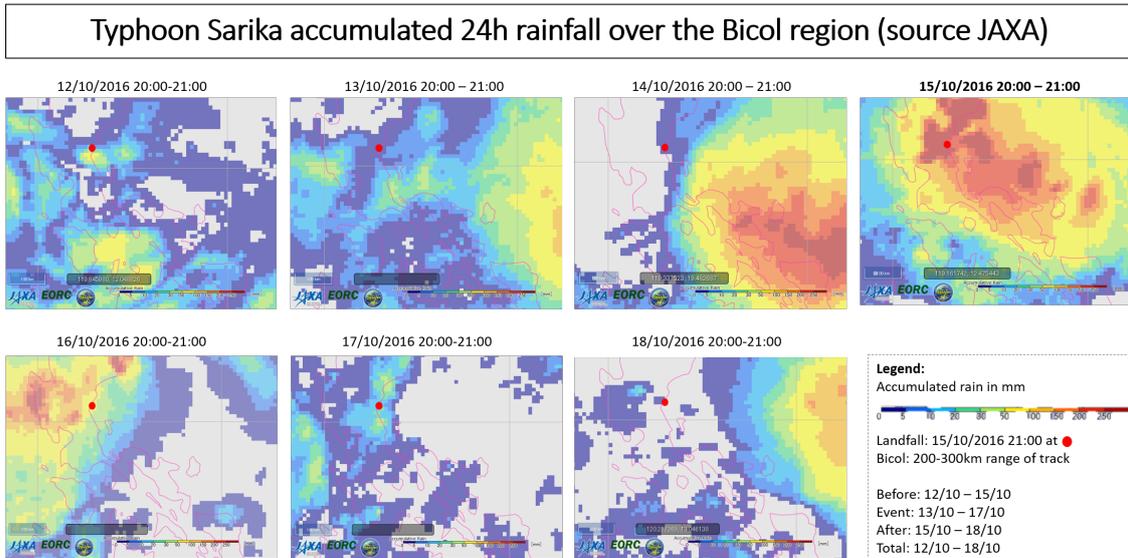


Figure C.3: Rainfall patterns associated with the four rainfall thresholds (before, after, event and total) for typhoon Sarika, obtained from JAXA (2021).

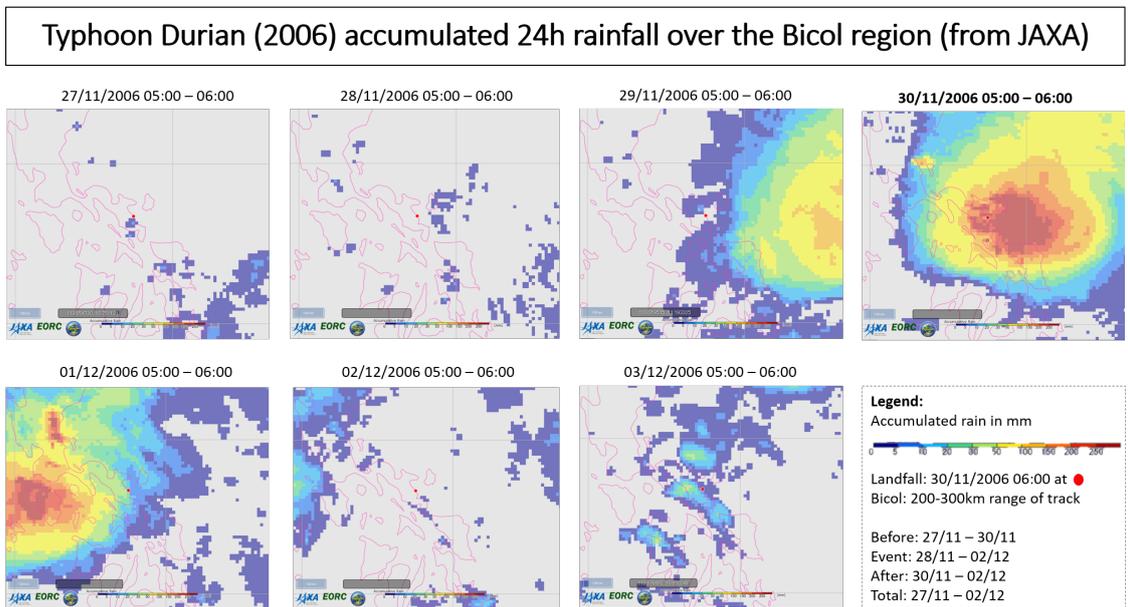


Figure C.4: Rainfall patterns associated with the four rainfall thresholds (before, after, event and total) for typhoon Durian, obtained from JAXA (2021).

## C.3 ACCURACY STATISTICS FOR HAZARD AND IMPACT FORECASTS

The F1-score is an accuracy measure, which does not consider TNs. The True Scale Statistics (TSS) or the Matthew's Correlation Coefficient (MCC) do include all classes of forecasts (TP, TN, FP and FN), give objective performance of the hazard and impact forecasts and MCC also functions well in unbalanced data sets such as those with scarce events (Chicco and Jurman, 2020; Rawat, 2019). In the case of a forecast with no TP values, resulting in an F1 score of zero, the MCC and TSS measures were assessed.

$$F1 = \frac{2TP}{2TP+FP+FN}$$

$$TSS = \frac{TP}{TP+FN} - \frac{FP}{TN+FP} = TPR - FPR$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP \times FP) + (TP \times FN) + (TN \times FP) + (TN \times FN)}}$$

where TPR and FPR are true positive rate and false positive rates, respectively (Chicco and Jurman, 2020; Ciavolella et al., 2016).

**Hazard impact for typhoon Durian and Sarika**

Accuracy statistics for F1 indicated low performance for typhoon Durian. MCC is close to 0, indicating that the forecast was as unreliable as tossing a coin, as stated by Chicco and Jurman (2020). TSS showed larger negative values compared to MCC.  $RT_{event}$  performed best for TSS and  $RT_{total}$  for MCC and F1. For typhoon Sarika, MCC and TSS were negative indicating low performance. The most reliable accuracy statistic was found for forecasts using  $RT_{after}$  due to the increase in TN and decrease in FP.

Landslide hazard forecasts of both typhoons for all three accuracy statistics performed similar and indicated low performance.

Table C.3: Accuracy statistics for hazard forecasts of typhoon Durian and Sarika

| <b>Durian</b> | <b>Actual TP: 4 Landslides</b> |              |              |              |
|---------------|--------------------------------|--------------|--------------|--------------|
|               | $RT_{before}$                  | $RT_{event}$ | $RT_{total}$ | $RT_{after}$ |
| TP            | 4                              | 4            | 4            | 2            |
| F1            | 0.069                          | 0.0683       | 0.072        | 0.048        |
| MCC           | 0.025                          | 0.018        | 0.049        | -0.084       |
| TSS           | -0.298                         | -0.165       | -0.599       | -0.916       |
| <b>Sarika</b> | <b>Actual TP: 5 Landslides</b> |              |              |              |
|               | $RT_{before}$                  | $RT_{event}$ | $RT_{total}$ | $RT_{after}$ |
| TP            | 0                              | 0            | 0            | 0            |
| F1            | 0                              | 0            | 0            | 0            |
| MCC           | -0.146                         | -0.137       | -0.121       | -0.039       |
| TSS           | -0.330                         | -0.303       | -0.255       | -0.035       |

### Landslide impact for typhoon Durian and Sarika

The reliability of landslide impact for typhoon Durian for *All* classes indicated negative values and low performance of the forecasts for  $F_1$ , MCC and TSS for  $RT_{before}$ ,  $RT_{event}$  and  $RT_{total}$ .  $RT_{after}$  provided a less negative value for TSS and MCC compared to the other RTs, because of the decrease in FP. For *high* classes, MCC and TSS had values close to 0 and slightly negative.  $RT_{before}$  and  $RT_{event}$  showed most reliable values for MCC, TSS and  $F_1$  for operation use.

For typhoon Sarika,  $F_1$  is zero, while TSS and MCC indicate both low performance. TSS and MCC give similar results for operational use of  $RT_{before}$  and  $RT_{event}$ . For  $RT_{after}$ , no forecasts were obtained (indicated with an empty value).

Impact forecasts for  $F_1$ , MCC and TSS for both typhoons for all RTs perform alike, except for  $RT_{after}$ , due to a reduction in FP and the dependence of  $F_1$  on TP. For forecasting applications, landslide impact forecasts of either  $RT_{before}$  and  $RT_{event}$  are the most reliable.

Table C.4: Accuracy statistics for landslide impact forecasts of typhoon Durian & Sarika

| All susceptibility classes  |               |              |              |              |                      |              |              |              |
|-----------------------------|---------------|--------------|--------------|--------------|----------------------|--------------|--------------|--------------|
| Durian: 4 Landslides        |               |              |              |              | Sarika: 5 Landslides |              |              |              |
|                             | $RT_{before}$ | $RT_{event}$ | $RT_{total}$ | $RT_{after}$ | $RT_{before}$        | $RT_{event}$ | $RT_{total}$ | $RT_{after}$ |
| TP                          | 1             | 1            | 1            | 0            | 0                    | 0            | 0            | 0            |
| $F_1$                       | 0.028         | 0.028        | 0.028        | 0            | 0                    | 0            | 0            | 0            |
| MCC                         | -0.131        | -0.135       | -0.135       | -0.096       | -0.102               | -0.087       | -0.083       | -0.029       |
| TSS                         | -0.350        | -0.359       | -0.359       | -0.209       | -0.192               | -0.238       | -0.250       | -0.714       |
| High susceptibility classes |               |              |              |              |                      |              |              |              |
|                             | $RT_{before}$ | $RT_{event}$ | $RT_{total}$ | $RT_{after}$ | $RT_{before}$        | $RT_{event}$ | $RT_{total}$ | $RT_{after}$ |
| TP                          | 0             | 0            | 0            | 0            | 0                    | 0            | 0            | -            |
| $F_1$                       | 0             | 0            | 0            | 0            | 0                    | 0            | 0            | -            |
| MCC                         | -0.091        | -0.091       | -0.091       | -0.018       | -0.045               | -0.039       | -0.042       | -            |
| TSS                         | -0.191        | -0.191       | -0.191       | -0.009       | -0.727               | -0.667       | -0.700       | -            |

### Total impact for typhoon Durian

The accuracy measures TSS and MCC indicated a more reliable total impact forecast for the 510 typhoon model compared to the hybrid model. MCC was higher for the 510 typhoon model compared to the hybrid model because of the value of the numerator: 960 for the 510 typhoon tool and 600 for the hybrid tool of  $RT_{before}$ .

TSS was larger for the 510 typhoon model compared to the hybrid model because of the difference between the False and True Positive Rates.

FPR for the 510 typhoon model were much lower compared to the hybrid model: 0.06 and 0.56 respectively, even though TPR was larger for the hybrid model (TPR of 0.8) compared to the 510 typhoon model (TPR of 0.4).

Considering subjective accuracy measures i. e. TSS and MCC, the 510 typhoon model was more reliable compared to the hybrid model. Yet the application of the typhoon and landslide forecast in a humanitarian context should be considered, where TPs are more important compared to TNs and acting in vain (FP) is preferred over not acting at all (FN). Additional subjective assessments of the hybrid model can be done for additional typhoon events both inside and outside of the Bicol region.

Table C.5: Accuracy statistics for the total impact maps using a hybrid multi-hazard model and the typhoon impact using a machine-learning method

| Durian | Actual TP: 76 |  |              |              |              |
|--------|---------------|--|--------------|--------------|--------------|
|        | 510 model     | Hybrid model - All susceptibility classes  |              |              |              |
|        | Total         | $RT_{before}$                              | $RT_{event}$ | $RT_{total}$ | $RT_{after}$ |
| TP     | 31            | 59   | 59           | 61           | 59           |
| TN     | 34            | 15   | 15           | 15           | 30           |
| FP     | 2             | 19   | 19           | 22           | 19           |
| FN     | 47            | 16   | 16           | 16           | 34           |
| F1     | 0.559         | 0.771                                      | 0.771        | 0.763        | 0.690        |
| MCC    | 0.350         | 0.234                                      | 0.234        | 0.208        | 0.236        |
| TSS    | 0.342         | 0.228                                      | 0.228        | 0.198        | 0.247        |
|        | 510 model     | Hybrid model - High susceptibility classes |              |              |              |
|        | Total         | $RT_{before}$                              | $RT_{event}$ | $RT_{total}$ | $RT_{after}$ |
| TP     | 31            | 38   | 38           | 38           | 31           |
| TN     | 34            | 31   | 31           | 31           | 34           |
| FP     | 2             | 5  | 5            | 5            | 2            |
| FN     | 47            | 40   | 40           | 40           | 47           |
| F1     | 0.559         | 0.628                                      | 0.628        | 0.628        | 0.559        |
| MCC    | 0.350         | 0.334                                      | 0.334        | 0.334        | 0.350        |
| TSS    | 0.342         | 0.348                                      | 0.348        | 0.348        | 0.342        |

# D CREATING AN IBF MODEL USING EXPERT-BASED JUDGEMENT

Various challenges within the data collection, hazard forecasting and impact forecasting components of an IBF model for typhoon-induced landslides have been identified in chapter 2.2. To create the most reliable landslide IBF model from scratch (i. e. when no existing model is available), a stakeholder assessment can be done by using a Multi-Criteria Decision-Making Method (MuCDM) to assess the elements of influence on the reliability. Specialists in the fields of landslide hazard and impact mapping; IBF for landslides and other natural hazards (typhoons) can be asked to provide their expert judgement on the influence of the elements on the reliability of each element within each component of the IBF model (2.2). Their judgement can result in a quantitative ranking or score of the reliability elements to determine the most influential ones. The expert-based quantitative ranking can be used to assess the influence of a certain element on the reliability of the components and on the total reliability of the IBF model and chain. The outcome of this method is a list of those elements on which stakeholders could focus during the creation of an IBF chain, such that a most reliable IBF chain for landslides or multi-hazards is created and risk resilience is enhanced.

A method for such an expert-based analysis of the reliability elements is presented in figure D.1 and consists out of 3 phases: data preparation, a Monte Carlo - Analytical Hierarchy Process (MC - AHP) and the Analytical hierarchy Process (AHP). In this Appendix, the method is explained in detail and a summary can be found in section D.3.

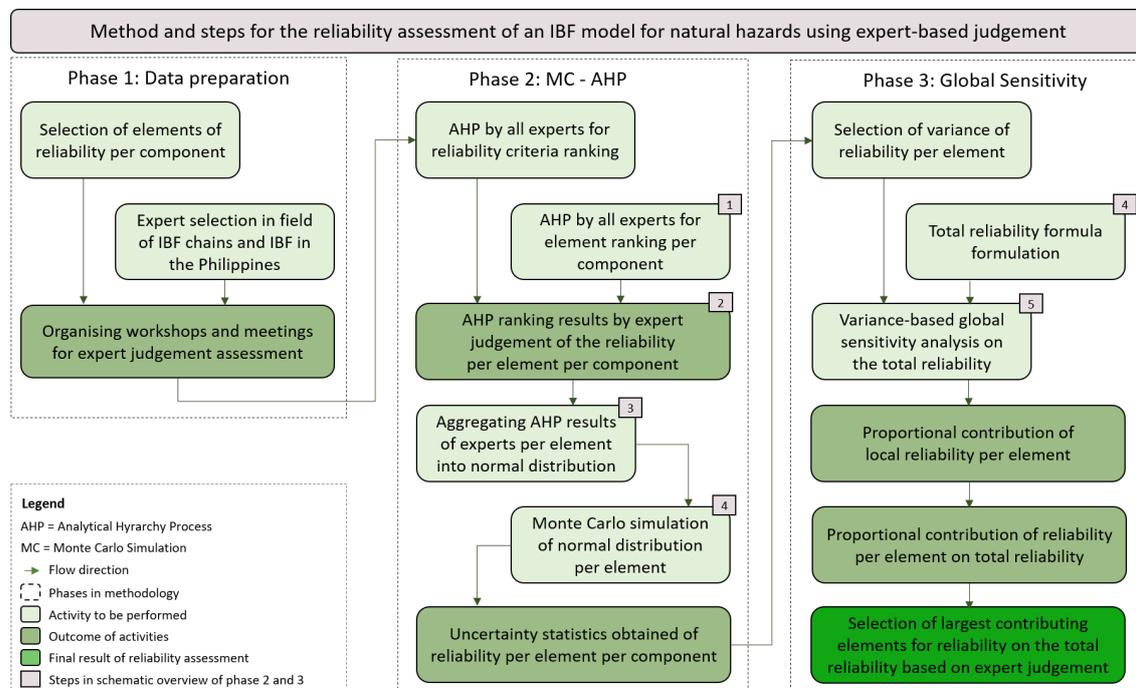


Figure D.1: Method for a reliability assessment on an IBF model using expert-based judgment

## D.1 PHASE 1: DATA PREPARATION

The data preparation phase, as shown in the left box in figure D.1, consists out of the selection of elements which influence the reliability per component. Experts are selected such that their skills can be applied for the reliability assessment of elements for a landslide IBF model. The selection of elements can be performed by brainstorming with the experts or prepared by a literature research on articles relating to landslide hazard mapping, IBF and landslide hazard and impact assessment as visualized in figure 2.2. The first approach is preferred, but requires meetings with experts.

## D.2 PHASE 2: MONTE CARLO - ANALYTICAL HIERARCHY PROCESS (MC - AHP)

In the second phase, a Multi-Criteria Decision-Making Method (MuCDM) is used to rank the selected elements per component by expert judgement shown in figure D.1 and in the first step (indicated by the green box showing a 1) in figure D.2. The Analytical Hierarchy Process (AHP) can provide a quantitative way to perform a qualitative analysis by expert judgement on the influence of the reliability elements (Brunelli, 2015). By combining AHP with a Monte Carlo simulation, an approach referred to as the Monte Carlo - Analytical hierarchy Process (MC - AHP), uncertainty statistics of the AHP outcome of the expert judgement of the rankings are obtained, such that a sensitivity analysis of the rankings by expert judgement can be performed.

### D.2.1 *The Analytical hierarchy Process (AHP)*

The Analytical hierarchy Process (AHP) is a Multi-Criteria Decision-Making Method (MuCDM) which has been widely used in providing a quantitative outcome of a qualitative analysis, such as expert judgement. Also, AHP has been applied in various studies to assist in decision-making for landslide susceptibility mapping (Roccati et al., 2021). The AHP approach is a suitable MuCDM for complex and technical cases where interaction between various criteria and elements is common, as it can order these interconnected elements into a structured ranking (Hsu and Pan, 2009; Xie et al., 2016). Subjective expert judgements are transformed into quantitative outcomes based on mathematical calculations using pairwise comparison. This results in a ranking of the elements based on their influence for each reliability criteria and an overall ranking of the influence of the element on the total goal of the AHP.

The first step in the AHP method is to provide for a goal of the analysis, which is the assessment of the reliability per component. To assess the reliability of a component, the reliability of each element,  $R_E$ , within that component is evaluated, to rank their influence on a criteria. Therefore, criteria of importance should be determined. This hierarchy of goal, criteria and elements is visualized in the first step in figure D.2.

After defining the hierarchy, the AHP assessment ranks the reliability elements for each component separately. The reliability of the elements is analyzed by means of a pairwise comparison. This allows for the experts to make judgements of the reliability of only 2 elements for 1 criteria. The pairwise comparison therefore reduces the complex

judgement of ranking all elements in one go to a simpler and clear comparison. The comparison firstly determines which element (of the two) is more influential regarding a certain criteria and secondly asks for a rating of its influence in numbers. These numbers range from 1, equally important, to 9, extremely important. These ratings determine the ranking of the elements and result in a percentage or quantification of their influence on the reliability of that component. A ranking is obtained for each component and for each element and is presented in a ratio matrix which orders the element from most influential (highest rank) to least influential (lowest rank). The reliability of one element for one component is combined by adding their rated relevance of the criteria of  $R_E$ . (Brunelli, 2015; Feizizadeh et al., 2015; Xie et al., 2016). The outcome of the AHP assessment is a ranking list of the influence on the reliability of elements in one component in percentages, as shown in the second visual in figure D.2.

Comparing the elements by expert judgement is subjective and inconsistencies can occur, i. e. when an expert contradicts him/herself. The AHP method uses a consistency rating (CR) of the comparison matrices, which should be below 0.1 (or 10%) for an acceptable consistency of the ranking by the experts. If the  $CR > 0.1$ , the AHP model used in this research gives the expert various options to alter their ratings of the pairwise comparison, such that the  $CR < 0.1$  (Brunelli, 2015).

For the AHP assessment, an online software can be used (<https://bpmsg.com/>), which can generate a URL link to be send to participants to perform a group AHP or individual assessments can be done. The results of both the comparison matrices and rankings are combined into the total result of all experts. The rankings of each element of all experts can be aggregated into a distribution or range of rankings per element, as shown by step 3 in figure D.2.

#### D.2.1.1 Monte Carlo Simulation

The AHP method is based on expert judgement and is therefore subjective. This subjectivity of the experts judgement of the reliability of elements can be referred to as the uncertainty of its reliability (Brunelli, 2015; Feizizadeh et al., 2015). In order to assess the probability of an expert to rank an element with a certain rating and thus the uncertainty of the expert judgement, a Monte Carlo (MC) analysis can be done.

Firstly, the range of rankings obtained in the previous step (step 3) is used as input samples for the MC simulation. Using AHP ranking outcomes instead of random variables as input for an MC analysis has been performed in other research using the Monte Carlo - Analytical Hierarchy Process approach (MC - AHP) (Feizizadeh et al., 2015; Hsu and Pan, 2009; Xie et al., 2016). For a MC simulation, a list of random variables within a range of a minimum and maximum value is used to generate a probability distribution of a number with a certain uncertainty. The list of rankings from step 3 could be used for the MC instead of a list of random variables. Then, the MC model is run and random variables from the ranking lists are selected and the simulation models a probability distribution of the rankings, or in other words the probability that one element receives a certain ranking by the experts. This process is repeated many times depending on the number of expert judgements performed

(referred to as  $n$ ) and the number of input parameters, i. e. elements ( $k$ ). Each run, other rankings from the input list of rankings are chosen randomly. The number of runs can be calculated using formula D.1 (Tate, 2012):

$$N = 2 * n(k + 1) \quad (D.1)$$

Uncertainty statistics can be obtained from the results of the MC analysis, e.g. minimum rank of the elements reliability; the maximum rank; the mean rank as well as the variance and the standard deviation (Hsu and Pan, 2009; Xie et al., 2016). The outcome of the MC simulation is shown in step 4 of figure D.2 and thus results in an insight into the uncertainty of the expert judgement of the ranking percentage of the elements. This is the final outcome of the second phase of the method, as shown in figure D.1.

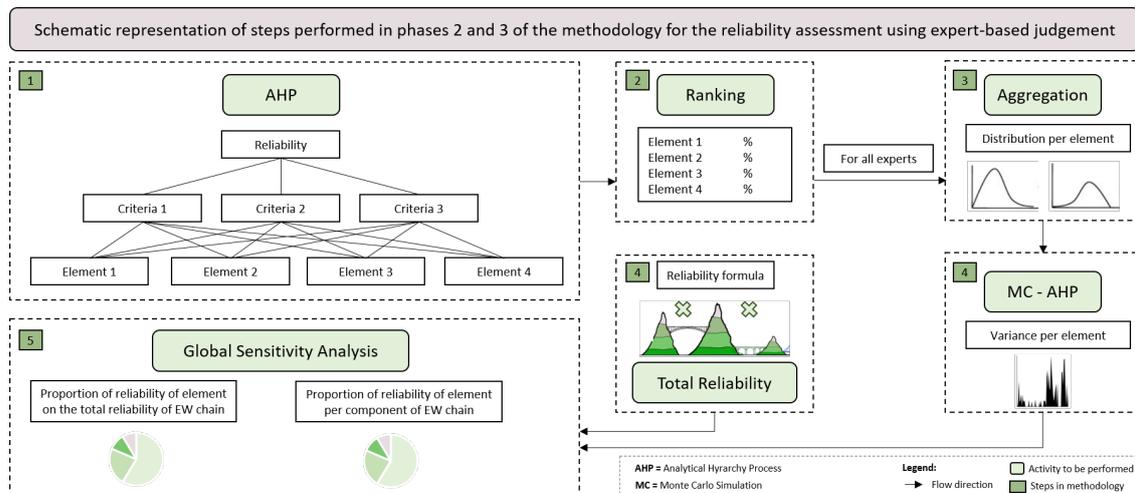


Figure D.2: Schematic representation of the steps performed in phase 2 and 3 of the reliability assessment for an IBF model for natural hazards.

### D.2.2 Phase 3: Global Sensitivity Analysis

In order to understand which elements have the largest influence on the reliability of one component and the total reliability of the IBF model, a Global Sensitivity Analysis (GSA) is performed in the third phase of the method (figure D.1). In a GSA, elements in consecutive components of the IBF model can be assessed simultaneously, such that the interaction of the elements between the components is considered. This is important, especially in IBF models, as components depend on each other and the element's influence on reliability should therefore be assessed relative to the IBF model rather than their individual influence on the total reliability of the model (Guzzetti et al., 2020; Tate, 2012; Thirugnanam et al., 2020; Zhang et al., 2019; Zhang et al., 2015). There are various types of GSA, such as the partial rank correlation coefficient; the multi-parametric sensitivity analysis; Fourier amplitude sensitivity analysis (FAST) and the Sobol's method. The latter two methods are both variance-based sensitivity analyses and can determine the influence of elements as a quantitative output. Variance-based GSA can determine the sensitivity of an elements reliability rank by two means: A first

order (linear) sensitivity index, referred to as  $S$ , gives a measure of the individual contribution of the elements reliability on the its component. Secondly, the Total Effect Sensitivity index (non-linear),  $ST$ , measures the total contribution of one element on the total reliability, thus taking into account the interdependence of the elements between components. The Sobol method is widely used and is said to be more 'powerful' compared to the FAST method if the input sampling size is not too large (Feizizadeh et al., 2015; Zhang et al., 2015).

The Sobol method uses variance as parameter assessment for the GSA. From the uncertainty parameters obtained from the MC-AHP of phase 2, the variance and mean are selected. Histograms showing the ranking and the number of times an element is ranked with a certain score from those, the mean and variance of the expert-based rankings are selected (step 4 in figure D.2). The variance and mean of the rankings can be used to assess the sensitivity of the expert judgement as the mean value of the rankings represents the middle value of the range and the variance represents how much the rankings of the experts deviate from the mean (Feizizadeh et al., 2015; Tate, 2012). The variance thus indicates how much the experts agree with each other on the ranking score and thus on the influence of one element on the reliability.

The Sobol method decomposes the variance of the rankings into the total variance of the components,  $V_C(E)$  and the variance per element per component,  $V_{Ei-C}$ , referred to as the partial variance. This is related to the total reliability (R) of the elements, which is the multiplication of the reliability per component (C) of the IBF model:  $R_{C1} * R_{C2} * R_{C3}$  (Thirugnanam et al., 2020), where  $R_{C1}$  is the sum of all elements within a component. The sensitivity of an element,  $S_E$  is calculated by taking the fraction of the partial variance over the variance of the component, as shown in formula D.2 and D.3, where  $i$  is the number of elements in a component. The sensitivity outcomes of the Sobol methods are called the Sobol indices  $S$  and  $ST$  (Tate, 2012; Xie et al., 2016).  $S$  represents the the individual sensitivity of each element and is a first-order and linear sensitivity calculation, where  $S$  ranges between zero, indicating no influence on the reliability and 1, indicating complete influence. It represents the sensitivity of an element to its component and thus can be seen as the contribution of one element's reliability on the reliability of one component.

$$S_{Ei-C1} = \frac{V_{Ei-C1}}{\sum_{i=1}^n V_{C1}} \quad (D.2)$$

$$S_{C1} = \sum_{i=1}^n S_{Ei-C1} \quad (D.3)$$

Secondly, the total effect sensitivity,  $ST$ , computes the influence of one element on the reliability of the whole IBF model, as shown in formula D.4 (Tate, 2012; Zhang et al., 2015). This is a second- or higher-order and non-linear sensitivity calculation as it takes into account the dependence of the components on each other for the total reliability, which is the sum of all components. The formula is included in the analysis in the second step in figure D.2.  $ST$  thus relates to the influence and contribution of one individual element on the entire IBF model.

$$ST_{Ei-C1} = \frac{V_{Ei-C1}}{\sum_{i=1}^n V_{C1} * \sum_{i=1}^n V_{C2} * \sum_{i=1}^n V_{C3}} \quad (D.4)$$

By performing a GSA on all elements for all components as shown in figure 2.2, the proportions of the influence of the reliability of the elements are obtained, visualized as the final step in figure D.2 and the final outcomes of the method in figure D.1 (Tate, 2012; Zhang et al., 2015).

**In short**, the Sobol method uses variance as parameter assessment. The uncertainty parameters obtained from the MC-AHP of phase 2, the variance and mean are selected. Histograms showing the ranking and the number of times an element is ranked with a certain score from those, the mean and variance of the expert-based rankings are selected, as can be seen in step 4 in figure D.2. The variance and mean of the rankings can be used to assess the sensitivity of the expert judgement as the mean value of the rankings represents the middle value of the range and the variance represents how much the rankings of the experts deviate from the mean (Feizizadeh et al., 2015; Tate, 2012). The variance can thus say something of how much the experts agree with each other on the ranking score and thus on the influence of one element on the reliability.

### D.3 SUMMARY OF METHOD

This method uses expert-based judgement to assess the contribution of the reliability of elements on the reliability of its component and on the IBF chain as a whole. This is done in three phases. Phase 1 represents the preparation of the data by selecting reliability elements for the assessment and selecting experts with skills and understanding of IBF, landslide mapping and IBF chains. The second phase starts with an expert-judgement of ranking the elements per component using the AHP approach. The outcome of this approach is a ranking list of elements with a quantitative score on their reliability. Per element, the quantitative rankings are aggregated to resemble a range of ratings of the elements. A normal distribution of the element's ranking and the probability of an element receiving a certain rank are obtained by performing an MC-AHP analysis. In phase 3, the variance of each element is selected, and the variance for each component and for the entire chain is calculated, using the formula for reliability:  $R_{C1} * R_{C2} * R_{C3}$ . A variance-based Global Sensitivity Analysis (GSA) is performed using the Sobol approach, which assesses the influence and contribution of one element's reliability on its component and the contribution of one element's reliability on the entire IBF chain's reliability.

By assessing the influence of the elements on each component and on the entire IBF chain, an understanding of what experts in the field would indicate as those elements most influential for the reliability of an IBF chain for natural hazards can be obtained. This method can be seen as an approach to use a global sensitivity analysis on expert-judgement to determine which elements have the largest influence on the reliability of such a chain. The outcome represents those elements on which stakeholders could focus such that the entire IBF chain can be improved in reliability in order for a reliable IBF model to be used and risk resilience is to be enhanced.

