Evaluating the explainability and performance of an elementary versus a statistical impact-based forecasting model

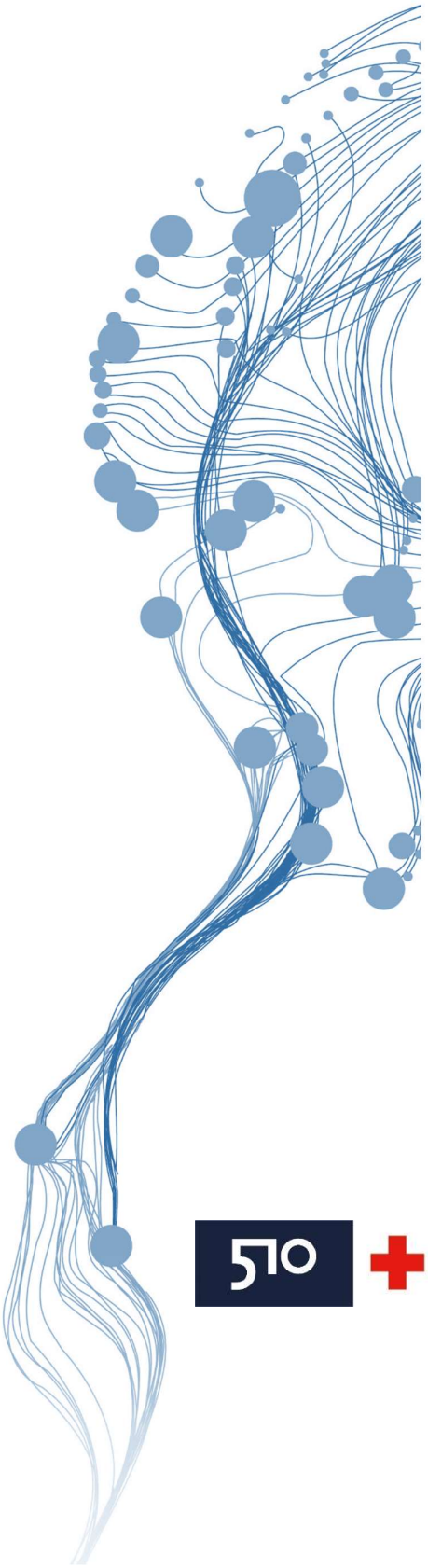
A case study of tropical cyclone early action in the Philippines

SAHARA SEDHAIN

June 2022

SUPERVISORS:

Prof. Dr. Norman Kerle
Prof. Dr. Maarten van Aalst
Dr. Marc van den Homberg (510 Global)



Evaluating the explainability and performance of an elementary versus a statistical impact-based forecasting model

A case study of tropical cyclone early action in the Philippines

SAHARA SEDHAIN

Enschede, The Netherlands, June 2022

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

Specialization: Natural Hazard and Disaster Risk Reduction

SUPERVISORS:

Prof. Dr. Norman Kerle
Prof. Dr. Maarten van Aalst
Dr. Marc van den Homberg

THESIS ASSESSMENT BOARD:

Prof. Dr. Mark. van der Meijde (Chair)
Dr. Saman Ghaffarian (External Examiner, University College London)

510

An initiative of the Netherlands Red Cross

ITC UNIVERSITY OF TWENTE.

Abstract

Climate vulnerable countries can expect an increase in the number of disasters, so investing in preparedness needs to be scaled up. Recently, there has been a remarkable shift in the focus of disaster risk practitioners, from traditional response mechanisms to proactive approaches of acting early based on impact-based forecasts (IbF). An effective implementation of these activities can only happen when the right information reaches the right people at the right time. For that, automatic trigger mechanisms are being developed, where pre-designed models are used to assess the impact and inform decisions with minimal human judgement. As the complexities in modelling algorithms increase, the interpretability of results from such models becomes more difficult, especially for users outside the domain. Therefore, benchmarking different approaches to IbF along with an interpretable evaluation mechanism is a top priority for humanitarian decision-makers, and is relatively unexplored.

This study attempted to evaluate two different models: (1) an existing statistical trigger model, operationalized for informing decisions for typhoon early actions in the Philippines, which uses a machine learning algorithm with several predictor variables, and (2) an elementary trigger model used for informing cyclone early action in Bangladesh, that combines damage curves and composite index overlay. For an objective comparison, the elementary model was adapted to the Philippines, placing both the statistical and elementary model in the same spatial context. The models were evaluated based on (1) their performance for damage prediction and their sensitivity to different risk indicators in hindsight for Typhoon Kammuri (2019) in the Philippines, and (2) their interpretability/explainability based on the architecture and parameters. To support this further, an interactive decision support tool was built for post-hoc evaluation.

Our findings suggest that, in retrospect, both models would have triggered with a minimum lead time of 72 hours, which is considered adequate for carrying out the pre-defined early actions. However, the performance of both models at the trigger time is not satisfactory, with a F1 score of 0.05 and 0.26 for the statistical and elementary models, respectively. The performance did not show an improvement over lead time, which can be attributed to the characteristics of this typhoon with considerable deviation from its forecasted track. However, in relative terms the elementary model performed better, and would have been able to maximize the impact reduced through early action, suggesting that, for this particular case, complex was not necessarily a better choice. At the same time, the overall results show that both models' performances are inconsistent in terms of lead time, and the elementary model does not show improvement in performance, even with observed typhoon data. Out of the two models, the elementary model was able to correctly predict higher damage percentages, while the statistical model was more conservative in its predictions. The statistical model better captures the characteristics of damage associated with the typhoon track, which is not considered in the elementary model.

In conclusion, the results are evidence that a more statistical analysis of events of different characteristics is needed to examine the overall suitability of these models for the implementation goal. A common evaluation framework needs to be built, not only to benchmark IbF models against each other, but also to communicate the uncertainties and considerations to relevant stakeholders. The interactive dashboard built in this research has the potential to be further expanded to fit that purpose.

ACKNOWLEDGEMENTS

Approaching the end of this journey at ITC, a lot of people have been a part of it, that I am extremely grateful for. I would like to start by thanking my supervisor Prof. Dr. Norman Kerle, for being ever so supportive and encouraging, but at the same time teaching me how to be critical and very sharp with my work. He has been not only a supervisor but also a constant mentor, helping me realize my potential and driving me in the right direction. I would also like to thank Prof. Dr. Maarten van Aalst for his very useful insights on the practical dimensions of the topic. My sincere gratitude also to my advisor Dr. Marc van den Homberg for his energy throughout, and for not only sharing his tremendous knowledge, but also helping me build my connection. I also must mention a big thanks to all the faculty of ITC for being so resourceful and approachable. A special thanks to Aklilu Teklesadik, Dr. Ahmadul Hassan, Olaf Neussner and colleagues from Build Change for their valuable insights in the research process. I feel extremely fortunate to have connected with so many experts in this domain who have helped me position my thesis in the sight of real-world applications, and for that, I am deeply indebted to the Netherlands Red Cross – 510 for letting me be a part of the team while developing this research.

There are many others in my ITC-family without whom I could not have survived these 2 years. Firstly, Ashok pasa for being a mentor and always having one answer to all my problems - 'of course you can do it'. I owe a big thanks to Enzo, for not only being the local perspective of my study area but letting me borrow his brilliant brain anytime and being the bouncing board of my ideas. The covid restrictions ended right on time, allowing me to spend my second year outside my room. That is when we formed this fantastic work group in room 5-036 of ITC- Sachita, Binita, Biman, Diana, Lynette, Nithesh, who made the thesis journey filled with less anxiety and more laughter. Along with that Sadi, Ranju, Om 'dai', Mahnoor, and all the names that I haven't mentioned but were a part of it, words cannot describe how much I have learnt from each of you.

And to wrap it up, all this has been possible because of my family – Aama, Abua, Aashu and Bishen, my biggest support system. Thank you for being so patient with me in these times.

Contents

LIST OF FIGURES

LIST OF TABLES

GLOSSARY OF TERMS

Impact-based Forecasting: A forecast of a hydrometeorological event that provides the weather's information along with its potential impact. This is done by considering the hazard along with vulnerability and exposure. They are intended to make the forecasts actionable.

Early Warning Early Action, Forecast-based Action or Anticipatory Action: Terms used by different organizations to commonly describe pre-defined actions that are taken before a foreseeable crisis, based on scientific forecasts and evidence.

Trigger: A pre-determined threshold for probability and magnitude of impact that activates an early action when reached.

Forecast-based Financing: It is one of the implementation techniques for early action where the necessary humanitarian funds are released automatically once the pre-defined trigger threshold has been reached.

Lead time: The time duration between the release of the forecast for an event and its occurrence.

Early Action Protocol: These are the step-by-step guidelines for the implementation of early action in an area. It provides information on how and when the triggers are activated, who are responsible and how the funds will be released. It also contains a detailed description and justification of the chosen trigger model.

Intervention Maps: An output map of the impact-based forecast, which visualizes the areas with the highest impact and is used for prioritizing the implementation of early action.

Elementary Trigger Model: An impact forecasting model where the relationship between variables is constructed based on mathematical equations.

Statistical Trigger Model: An impact forecasting model which trains a machine learning algorithm to find patterns in the data and does not rely on rule settings.

The definitions in this section are based on the Forecast-based Financing manual [1]

[1] https://manual.forecast-based-financing.org/en/chapter/glossary/

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
AWS	Automatic Weather Stations
BMD	Bangladesh Meteorological Department
C-EAP	Cyclone Early Action Protocol of Bangladesh
DREF	Disaster Relief Emergency Fund
EAP	Early Action Protocol
ECMWF	European Centre for Medium-Range Weather Forecasts
EWEA	Early Warning Early Action
FAO	Food and Agriculture Organization
FbA	Forecast-based Action
FbF	Forecast-based Financing
FN	False Negative
FP	False Positive
GloFAS	Global Flood Awareness System
HDX	Humanitarian Data Exchange
IbF	Impact-based Forecast
IBTrACKS	International Best Track Archive for Climate Stewardship
IFRC	International Federation of Red Cross Red Crescent Societies
JTWC	Joint Typhoon Warning Center
NCEP	National Centers for Environmental Predictions
NDRRMC	National Disaster Risk Reduction and Management Council
NMHS	National Meteorological and Hydrological Services
PAGASA	Philippine Atmospheric Geophysical and Astronomical Services Administration
PAR	Philippines Area of Responsibility
PRC	Philippines Red Cross
RI	Rapidly Intensification
TCP	Tropical Cyclone Program
T-EAP	Typhoon Early Action Protocol of the Philippines
TN	True Negative
TP	True Positive
UKMET	United Kingdom's Meteorological Office
WFP	World Food Program
WMO	World Meteorological Organization

1. INTRODUCTION

1.1. Background

The effects of natural disasters, especially those concerning hydrometeorological events, have been escalating in recent decades (Alexander, 2018; Masson-Delmotte et al., 2021; WMO, 2021a), raising the challenge of putting more effort into their prediction, preparedness, and mitigation. A significant fraction of these events is known to happen in low and middle income countries, prompting humanitarian agencies to be at the forefront in supporting financially for reducing the impact and quick recovery. Ideally, a long-term risk mitigation strategy must be designed in a community where such risks from hazards are identified. Risk in the context of a natural disaster is defined as the possibility of human, economic or environmental loss caused by the interaction between hazard and a vulnerable condition (ISDR, 2004). However, acting only based on the risk assessment has limited scope in terms of financial capacity and practicality. Records have shown that typically around 80 percent of disaster finance is spent on relief and recovery, and only a small fraction is used for the prevention of risk (Jan and Caravani, 2017), even though the economic loss due to the impact of a disaster is much higher than the cost of its prevention (Mechler, 2005; Shreve and Kelman, 2014). The global disaster risk reduction framework has identified the need to prioritize on making the early warning information more actionable with the help of guidance on necessary measures for each forecast (Hyogo Protocol, 2005-2015). Over the years, the techniques for hazard forecasting have been evolving, while the scientific community is finding newer ways to make the preparedness and preventions well guided. To translate these forecasts into action, it is not only enough to have the location and hazard information, because an event of similar magnitude may have a different impact depending upon the spatially variable risk. Hence, an Impact-based Forecasting (IbF) is an ideal approach, which, unlike the traditional forecast, not only addresses the characteristics of the hazard itself, but also provides information on the space, time, as well as nature of its impact, allowing an objective measurement for preparedness action Figure 1 (Red Cross Red Crescent Center, 2020; WMO, 2015).

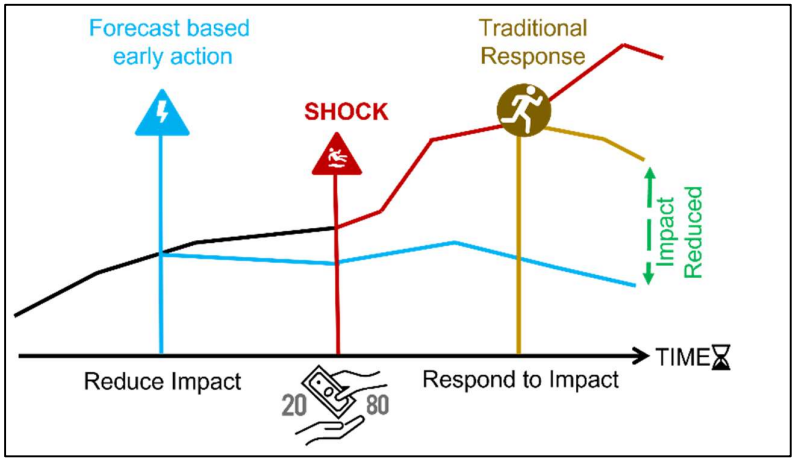


Figure 1: A simplified representation of disaster timeline, highlighting
the significance of early action to reduce impact. Source: Adapted and modified from UN-OCHA

This information from an IbF allows the crucial time between the forecast and occurrence of a disaster to be better utilized in taking the necessary set of actions, commonly termed as 'early action', 'forecast-based action' (FbA) or 'anticipatory actions'. As one of the ways to effectively translate these warnings to action, the humanitarian community has developed a framework of Forecast-based Financing (FbF), where necessary funds for preparedness activities are automatically released early based on the scientific forecast (Coughlan De Perez et al., 2015). In 1985, the International Federation of Red Cross Red Crescent Societies (IFRC) established the Disaster Relief Emergency Fund (DREF) for rapid release of funds post-disaster, and in 2015 had the FbA component added to it (IFRC, 2022a). IFRC defines the formal set of guideline for these anticipatory actions

as Early Action Protocol (EAP) (IFRC, 2020a), which describes the financing mechanism, type of early actions and the method of defining impact level for prioritization in each implementation area. Currently, there are 17 EAPs approved or under review by DREF, and 21 EAPs under development, which takes on average 2 to 3 years (IFRC, 2022b). Although unique to each implementation area, the steps of developing an FbA, can be broadly described in the following categories: (1) understanding risk, (2) forecasting hazard and impact, and (3) warning dissemination and early action (IFRC, 2020a).

The process of understanding risk begins by identifying suitable source of hazard and impact information, based on their type, availability, and reliability. Hazard forecasts can be done from regional to global scale, but localized interventions require a forecast of higher spatial resolutions. Some approaches may also use a combination of global and regional models in suitable times before an event. During the Bangladesh flood of 2021, a probabilistic model from Global Flood Awareness System (GloFAS) (Alfieri et al., 2013) was used 10-15 days before the event for pre-planning, while a national forecast model was referred to when the event was closer (OCHA, 2021a). In terms of damage history, the available global disaster databases, such as EM-DAT[2] and Desinventar Sendai[3], are great tools considering their systematic operations and coverage. But having said that, these data are often on a regional or national scale, and obtaining localized data is more helpful in planning a focused intervention. The granularity to which an impact model can provide forecast, is completely dependent on the resolution at which all the risk indicators are available.

There are several techniques being developed or used by agencies to integrate the risk information for financing early action: threshold-based, qualitative combination of hazard and vulnerability, impact modelling and climate-based models (Wilkinson et al., 2018). A threshold method identifies the limit beyond which an impact becomes problematic - a "*trigger*", which helps determine the time and the target areas of activating the early actions. With probabilistic or ensemble methods for hazard forecast, the models give not only the intensity of impact but also its likelihood, which is converted to a risk matrix (Figure 2) that is easily interpretable. The risk matrix allows the decision-makers to choose when these anticipatory actions must be triggered, depending on when an impact becomes a point of concern and certainty they require in their actions. If the forecasted impact does not materialize, then the action taken is considered as 'action in vain'.

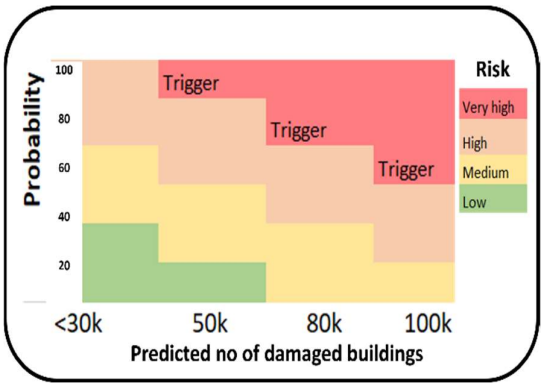


Figure 2 An Example of a risk matrix based on probabilistic damage

A trigger model can be designed utilizing different techniques, where some are automatic, while others are based on expert knowledge (IFRC, 2020a). Automatic approaches to prediction of absolute impact levels are generally done in two ways; (1) elementary modelling where a relationship between variables is constructed using mathematical equations, and (2) statistical modelling trains a machine learning algorithm to find patterns in the data and does not rely on rule settings. In both approaches the determination of impact level is done primarily with damage and hazard history along with social vulnerability or coping capacities, all integrated in different ways. The impact models used for humanitarian actions constantly improve with state-of-art prediction

[2] https://www.emdat.be/
[3] https://www.desinventar.net/index.html

algorithms, which allow better control over financial risk (Bierens et al., 2020). However, generating these models is still a challenge, since many areas have limited historical impact information, though researchers are focusing on reducing these gaps. Many promising studies have developed machine learning techniques for improved transferability of damage data in terms of location and event type (Valentijn et al., 2020; Wagenaar et al., 2021), and identifying damage patterns in areas where data are not abundant (De Perez et al., 2016). However, using such a black box technique makes it difficult to compare changes that might occur due to the nature of hazard or risk information. For example, the characterization of building vulnerability can have a different definition in two separate geographic locations. Moving forward, there needs to be a quantitative evaluation of how the availability of geodata and design of trigger model affects the prediction of impact and effectiveness of FbA strategies.

While designing these triggers, the time is an essential consideration. In the case of typhoons, for example, the spatial accuracy of the forecast increases as it gets closer to land (Yu et al., 2013), while intervention decision needs to be taken with several days lead-time. On the other hand, drought has a very long lead time, which is how the World Food Program (WFP) released funding for an early response of drought in 2015 in Southern Africa, the effect of which was felt later in 2016-2017 (WFP, 2016). Thus, there is always a choice between waiting for a last-minute accurate forecast or acting in advance with enough time to prepare. In either case the chances of acting in vain and if they outweigh the loss of acting late need to be evaluated, and these trade-offs may differ for each hazard. This balance can also be affected by the scale at which the early actions are targeted, for instance, an intervention done at the household level will need a more accurate forecast compared to a community scale planning. On the other hand, there are also restrictions in terms of time required to take actions. During a pilot study done on Bangladesh during the 2020 flood event, it was deduced that the intervention would not have been successfully completed if the pre-activation, 10 days before the event, were not done (OCHA, 2020a). The quantitative determination of optimal time and spatial coverage for these actions is still an open question.

The early action process usually begins with an implementation map which indicates the predicted impact and is used for prioritizing the intervention. The early actions are pre-designed and may be unique, depending on the implementation approach, hazard type, location, data availability, main economic activity, and organizational priorities. In the Philippines, for example, the Red Cross provides homeowners with a shelter strengthening kit before a major Typhoon strikes (Anticipation Hub, 2020), whereas in Bangladesh the Food and Agriculture Organization (FAO) provides them with materials to flood-proof agricultural assets (FAO, 2020). The prior evaluation of risk should keep this in consideration; to give an example, in a place where flood significantly affects crops, a detailed crop loss assessment must be done.

As we are shifting to automated decision making in the humanitarian context, it is also an added challenge to make these processes more interpretable. The General Data Protection Regulation (GDPR) adopted by the European Union, establishes that an algorithm-based decision makings must be non-biased and explainable (Goodman and Flaxman, 2016). Meteorological agencies are a very good example of recognizing the need to deliver the uncertainties in their forecast for fair and effective decision making (WMO, 2008). The same needs to be adopted in impact forecasts, which require a high level of confidence considering how the decisions need to be triggered in a very short span of time. Gevaert et al., (2021) call attention to the fact that the replacement of traditional techniques by artificial intelligence (AI) has reduced the transparency in humanitarian decision making. There are multiple studies that (1) emphasize how the disparity in the understanding between users and producers increase as models become more complex, and (2) advocate the role of visualization for better communication (Kirchhoff et al., 2013; McInerny et al., 2014). As described in the study by Mittelstadt et al., (2019), explainability or interpretability of decision-making algorithms can be evaluated in two ways (Figure 3); (1) transparency, which is the ability to decipher the internal functions of the model, including the architectures and parameters, and (2) post-hoc interpretation that gives an understanding of how and why a model behaved in a certain way. The latter is often done through explanations or visual and interactive interfaces.

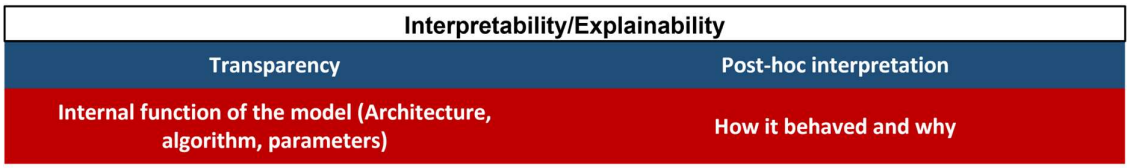


Figure 3 Ways to evaluate the interpretability or explainability of a decision-making algorithm as described by Mittelstadt et al., (2019)

In the framework of IbF, most commonly only an implementation map is communicated with the end-users, which could lead to a gap in understanding the uncertainties and limitations during the modelling process. In a case study by Bierens et al., (2020), the authors have very well highlighted a lack of insight in the local stakeholder about what basis the IbF models are built on and what they can or cannot do. The study recommends shifting to a co-creation of these models, which can increase the understanding and help meet the end-user requirements more effectively. A similar issue is identified in the study by van den Homberg et al., (2020), which underlines the need to benchmark different approaches of IbF, such as the use of AI against expert opinion for decisions. There need to be more evidence-based studies on not only evaluating the choices made in IbF, but also making them assessable and effectively conveying them.

1.2. Research Problem

Problem Statement: A quantitative evaluation and communication of the performance and uncertainties of trigger models over time and under different scenarios is crucial for effective scaling up of the forecast-based early action practices.

The IbF has allowed not only the generation of a systematic framework to use forecast information for preparatory action, but also provides a mechanism to validate the incentives. While the practice of IbF and FbA began with a very limited number of organizations and countries (Coughlan De Perez et al., 2015), by now it is being adopted across the globe for multiple hazards (Anticipation Hub, n.d.). However, to scale up, more robust evidence is needed on what kind of trigger model works in which area and in what context. Since investing effort in early action is a lot about different financial or managerial risks that various agencies are willing to take, it becomes crucial to increase the reliability of data-guided decisions and make it more transparent.

As already highlighted in the introduction, in designing the trigger models for the FbA there are multiple trade-offs to consider (Figure 4), and (1) the timing of these actions is crucial. This selection will further influence the (2) scale of implementation on whether to wait longer for spatially accurate forecasts, or to increase the spatial coverage. However, this trade-off in time and scale cannot be isolated, since several other indicators such as the (3) choice of the forecast, (4) availability of historical datasets and risk information, and (5) the technique used to integrate these will affect the accuracy of loss predictions and the success of a trigger model. There is also an associated question on (6) the ease of validation and interpretability of the results.

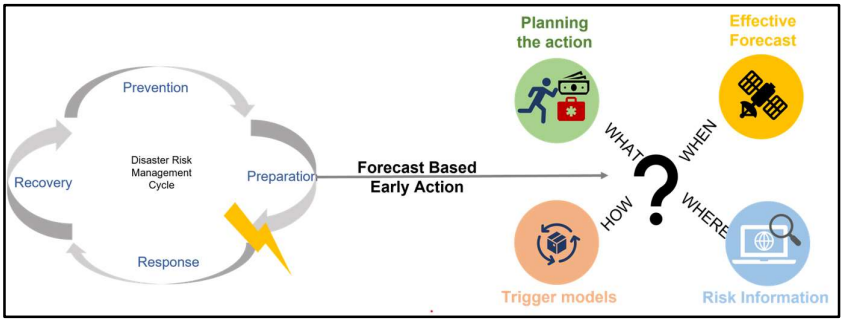


Figure 4 Conceptual framework

These are fundamental questions that many organizations are currently tackling to make the practice effective and sustainable, while very few empirical studies have been done on assessing the trade-offs. This research aims

to quantitatively measure this based on the accuracy of impact predictions, made with various forecast timestamps for a historical typhoon event using two different impact-forecasting technique. It further intends to investigate how this accuracy will vary with different choices of trigger models and how can it be effectively communicated.

1.3 Objectives

1.3.1 Main Objective

To evaluate two IbF trigger models based on (1) statistical and (2) elementary techniques, by retrospectively assessing their performance for Typhoon Kammuri (2019) in the Philippines, and while identifying dependent parameters for accuracy of damage estimation.

1.3.2 Sub-objective and Research Questions

1) To compare the input parameters of the two IbF trigger models

1.1 What is the difference in the way hazard forecast and hazard history are integrated in the two trigger models, and converted to the required spatial and temporal scale?

1.2 What is the difference in how the models incorporate physical and social vulnerability?

2) To validate the performance of the two trigger models

2.1 What is the impact of lead time on the performance of the two IbF models in predicting household damage?

2.2 What is the impact of choices made in trigger thresholds and municipality prioritization, on the performance of damage prediction?

2.3 What is the sensitivity of the two trigger models to changes in one of the predictors (in terms of hazard and vulnerability)?

2.4 What would be the most adequate combination of threshold and lead time for each trigger model to trigger early action during the event?

3) To compare the explainability of the models in terms of their transparency and post-hoc interpretation

3.1 What is the level of transparency in the two models in terms of interpreting individual parameters, algorithm and prediction results?

3.2 What is a suitable method to evaluate the choices and uncertainties in the models for a post-hoc evaluation through an interactive platform?

1.4 Thesis Outline

The chapters of this thesis are organized in the following ways: Chapter 2 gives a brief background of the study, diving into IbF in the context of the study area, particularly for typhoon hazards. This chapter also includes a literature review on approaches of IbF in two different trigger models. Chapter 3 outlines the methodology that was used for answering the research questions. Chapter 4 then presents the results obtained in each step. Chapter 5 includes a detailed discussion and implications of the results, in context of the research questions. Finally, chapter 6 highlights the identified limitations and recommendation for future work and the concluding remarks.

2. CONTEXT OF THE STUDY

When a forecast-based early action is considered for an area, there are multiple influential variables which differ as per the hazard type and local context. This section briefly looks into a few of these components to set up a mechanism for early action, focusing on Typhoons but also with relevant example of other hazards. Along with that, the theoretical framework of the two trigger models and background of the study area has been presented.

2.1 Trigger Development for FbA

As previously discussed in Chapter 1, the process of developing early warning and early action strategies has multiple stages, and different tools can be employed to achieve it, which is summarized in the Table 1.

Table 1 Key components in the timeline of an early warning and early action system along with its scale and assessment methods

Early Warning Early Action (EWEA) System component	Subcomponent	Spatial scale	Temporal scale	Assessment method
Risk knowledge	This involves gathering data on historical disaster impact, vulnerability, and exposure to examine the nature of past impacts. A detailed risk analysis is done to develop the trigger, but also to help in identification of suitable early action.	The risk indicators are often taken at the smallest administrative unit available. But this is highly dependent on the availability of regional, national or global databases.	Depending on the type of risk indicator, the effect of timeliness of the dataset varies. For example, topographic variables can be considered stable for a more extended period, compared to demographics such as population. The granularity of the intervention area also greatly influences the extent to which these dynamic variables impact the trigger model.	Sensitivity analysis can be done to test the effect of data quality in terms of timeliness, accuracy, reliability of source and granularity (van den Homberg et al., 2018).
Monitoring and warning	Identification of suitable forecast providers is done based on its type, availability, reliability and frequency.	The spatial scale of forecast often depends on the data provider, where national and regional forecasts have better resolution compared to global. In the context of Ibf, the forecast is converted to the desired spatial scale at which the early action is targeted, using different modelling and estimation techniques.	Lead time: Differs based on the hazard type, from short-term forecasting within minutes for hazards such as earthquakes, to a long-term forecast of years ahead for droughts or sea-level rise. At the same time, it also depends on the forecast capacity in place for a specific country or context (for example, Bangladesh has a shorter available lead time for tropical cyclones compared to the Philippines due to the differences in coastal geometry).	Skill analysis of a forecast provider can be done by evaluating the average differences in forecasted versus observed event at various lead times over a time frame. On the other hand, verification of forecast is something done after an event for a case-specific evaluation.
Impact Modelling	Based on risk indicators and hazard forecasts, impact predictions are made. Method of modelling can include	The scale at which predictions are made is often dependent on the granularity of available variables of risk indicators and impact history. The	The regularity of impact modelling within an event is dependent on the frequency of forecast data distribution. Trigger models are often directly linked to the forecast	Detailed skill assessment of the trigger model can be carried out based on historical forecast

	expert-based, statistical, or elementary.	scale is also determined by the modality of the trigger mechanism, which can be either targeted at an entire region of a country or at the smallest administrative level.	database, and impact can be modelled as soon as new forecast information is available.	and observation to see how often a trigger would be reached and the chances of acting in vain.
Communication and dissemination	While the impact forecasts are being monitored during an event, they are continuously communicated to relevant users. Once the impact has reached the pre-defined trigger, an implementation map is generated and circulated for prioritization of early actions.	The scale of the EWEA message is relevant at the level for which intended actions are targeted. For example, some can be for subnational level, to alert local authorities for strategizing preparedness. There could also be alerts for the community to timely evacuate.	The time for communication and dissemination of EWEA messages depends on the procedural aspects of information flow from the national to the local level and the type of communication channel used (radio, social media, TV, information portals etc.).	Assessment of an effective EWEA message can be done through several qualitative evaluation metrics. For example, how the warnings were understood, if they reached the intended users on time, or what actions were taken based on the warning.
Response	Based on the impact-based warnings and implementation maps, the pre-defined early actions are carried out.	The scale of implementing early action (the number of households, population or area that can be covered) depends on what is needed, the budget available and the chosen thresholds.	The implementation time of an action is the duration of time taken to carry it out. Action lifetime is the duration until which the preparedness measure will have an effect. For example, early harvesting of crops, once done, will remain relevant for an entire flood or typhoon season, whereas measures such as evacuation are short-term.	Monitoring and evaluation can be employed for measuring to what extent the early action contributed to minimizing impact, which often requires a detailed study (Bischiniotis et al., 2020; Gros et al., 2019; Lopez et al., 2020).

2.2 Typhoon Impact Forecasting and Validation

A tropical cyclone (also known as a Typhoon or hurricane, based on their origin) is a rapidly rotating storm that forms above warm tropical water (Montgomery and Farrell, 1993). Once formed, the global wind circulation causes the cyclone to move, where it continues to enlarge until it reaches land and starts dissipating energy, also known as the landfall. This weakening of the storm after landfall is caused by the change in temperature over land due to the influx of cooler and drier air. In the context of IbF, the lead time of tropical storms is when the forecast of landfall is made prior to its occurrence. Cyclones have destructive properties associated with extreme wind, rainfall and storm surge, followed by cascading effects such as floods and landslides, causing on average 43 deaths per day worldwide (WMO, 2020).

Tropical cyclones are forecasted from their formation based on movement (track) and wind speed (intensity). The position of the storm centre gives the track, while the intensity is calculated based on the maximum speed it sustains in a specific time, usually calculated in knots (1knot = 1.85 km/hour). The tropical cyclone forecast changed drastically since 1960 when the first meteorological satellite, TIROS-1, was launched (U.S Weather Bureau, 1961). Today many numeric predictions models are used to predict how a tropical cyclone will develop and move after its formation. A deterministic forecast for tropical cyclones gives a single value for intensity and

path at different lead times. On the other hand, a probabilistic forecast can either be dichotomous with binary results of the event, or an ensemble forecast which gives the probability of exceeding a threshold value instead of one best estimate (Tóth and Zhu, 2003). Most of the models today use an ensemble forecast technique that allows better predictions for longer lead time with inclusion of uncertainties in the model and climatic patterns (Wilkinson et al., 2018). The uncertainty is represented by a cone in the track based on historical errors in the official forecast (NHC, n.d.).

A global model predicts for the entire globe and can provide forecasts at a longer lead time than a regional model. In contrast, a regional model with a higher spatial resolution is specific to local geography, but is used only once the storm has reached a certain intensity. These models are either entirely statistically based on historical events, or statistical-dynamic, i.e. combined with environmental variables. Currently, there are numerous forecast providers, such as the European Centre for Medium-Range Weather Forecasts (ECMWF) (Molteni et al., 1996), United Kingdom's Meteorological Office (UKMET) (Radford, 1994), National Centers for Environmental Predictions (NCEP) (Tallapragada, 2016), which have been in place for a few decades now. Over the years the forecasts have improved much more in terms of storm track prediction, compared to its intensity, as concluded by Heming et al., (2019). The mean track error for 3 to 5 days lead time in UKMET forecasts has reduced by several hundred km in the past 3 decades, whereas the positional error in ECMWF has reduced by almost half since 2008 (Heming et al., 2019). Modelling techniques have also improved in forecasting not only typhoon track and intensity, but also its associated extreme weather such as storm surge and rainfall (Chen et al., 2020).

To convert these forecasts into impact-based warnings, timing and location of hazard are highly crucial, along with the exposure data. There needs to be a strong coordination between national meteorological and hydrological services (NMHS), who are responsible for providing timely weather warnings, and relevant agencies for vulnerability and exposure information (WMO, 2015). In 1980, the World Meteorological Organization (WMO) launched the tropical cyclone program (TCP) to increase coordination between local and global forecasters for effectively using tropical cyclone warnings to reduce its impact (WMO, 2021b). The warning of tropical cyclones is given in categories 1 to 5, based on different thresholds for maximum sustained windspeed that vary in space and time (WMO, 2015). These warnings are dissipated through an impact matrix along with a probabilistic map for typhoon track, intensity, and its likely impact. There are still hydrometeorological forecast uncertainties, such as storms that undergo rapid intensification or are slow moving, which limit the performance in these models (Grimes and Mercer, 2016).

IbF for typhoons can be assessed in terms of materialization of forecasted event and the impact. The skill of a tropical cyclone forecast is given by the similarity in the forecasted storm path and intensity with the observations. The observed track and windspeed vary within agencies, as they use a different metric for calculation, and the International Best Track Archive for Climate Stewardship (IBTrACS) gives the mean location and intensity estimated (Knapp et al., 2010). There are several methods used for the spatial verification of forecasts depending upon the application (feature-based, scale-decomposition and neighborhood-based). Casati et al. (2008) demonstrate a detailed study of how verification of forecast methods has dramatically advanced in the last decades, from traditional techniques with limited observational parameters to more user-specific and meaningful interpretations. Several studies describe statistical methods to quantify the economic value gained from a forecast to judge the value of forecast over its quality (Katz and Lazo, 2012; Mylne, 2002; Thornes and Stephenson, 2001). On the other hand, the validation of predicted loss is often done using the observed damage data collected after the event through various sources (Example: government repository or satellite-based damage assessment).

In this study, the evaluation of IbF model performance is limited to verifying the predicted damage against the actual (observed) damage.

2.3 The Study Area

After a brief context of different influential factors of FbA, this section will focus on understanding more of the study area. The typhoon in the Philippines was chosen as the case study for this research based on the country's high susceptibility to disaster impacts, while also having good examples of preparedness strategies adopted for it.

2.3.1 Disaster Risk Management in the Philippines

As per the records of past 20 years, the Philippines ranks top 5 in the number of disaster events globally, with typhoons being predominant (EM-DAT, 2022). Yearly, on average, 20 storms enter the Philippines Area of Responsibility (PAR), out of which 5 are destructive (Santos, 2021), and the frequency of such extreme weather events is expected to rise in the future (Masson-Delmotte et al., 2021). With the growth in urbanization, the impact from typhoons will also naturally grow, including associated risks of flood, storm surge and landslide due to heavy rainfall. Figure 5 shows a map of the Philippines and areas affected by typhoon Kammuri.

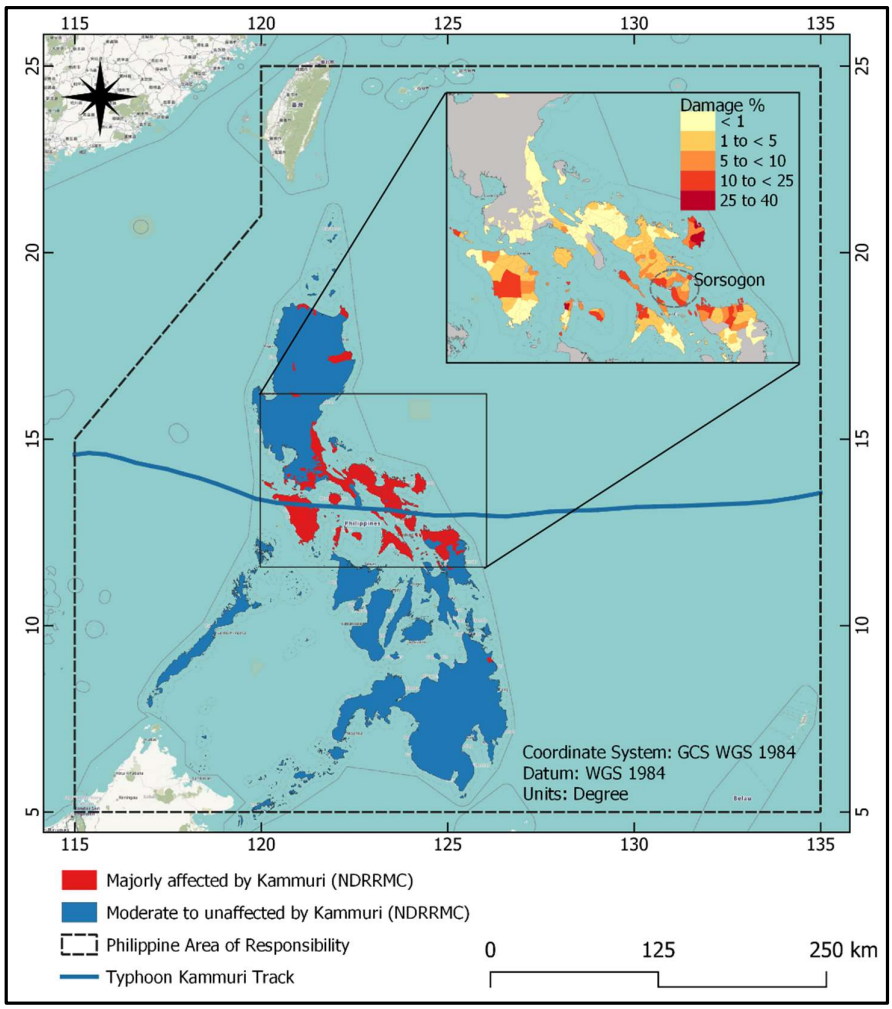


Figure 5 A map of the Philippines, highlighting areas affected by Typhoon Kammuri (2019)

As of 2005, the country's disaster management policy was largely limited to response mechanisms, with minimal action before an event (Bank, 2005). After recent devastating events, in particular Haiyan (2013), Hagupit (2014), and Mangkhut (2018), the approach towards typhoon risk management in the Philippines largely shifted to pro-active preventive measures, such as pre-evacuation (ADRC, 2018). A recent revision in the government policy has also permitted the local administration to use the response funds for preparedness, if 15% of population

are predicted to be impacted based on forecasts (Bierens et al., 2020). The National Disaster Risk Reduction and Management Council (NDRRMC), which is formed under the Philippines government in partnership with several other stakeholders, is responsible for coordinating all pre and post disaster activities in the country (NDRRMC, n.d.). The Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA) is a dedicated body for disseminating flood and typhoon advisories in the region since 1972 (GOVPH, n.d.). The department of science and technology (DOST-ASTI, n.d.), functioning under the government, also works actively in conducting research and product development to support the resilience activities.

2.3.2 Typhoon Early Action Protocol

The preparedness strategies of the government were adapted to more evidence-based actions under the FbA framework of the Netherlands Red Cross, as Typhoon Early Action Protocol (T-EAP), in 2018 (PRC et al., 2020). The primary implementation of the EAP (PRC et al., 2020) is done under the Philippines Red Cross (PRC) in partnership with the German Red Cross, Finnish Red Cross and 510- initiative of the Netherlands Red Cross. Currently, the EAP is targeted to support 19 regions that were identified to be at the highest risk, and targeted explicitly to smallholder farmers, the fishing community and houses made of lightweight material. The EAP is performed with three pre-identified early actions: providing temporary shelter strengthening kits, cash for an early harvest of crops, and support for livestock evacuation. Evacuation of people is not considered, because the Philippines government is believed to be already implementing this quite efficiently. Figure 6 shows an overview of the operational process in this trigger model.

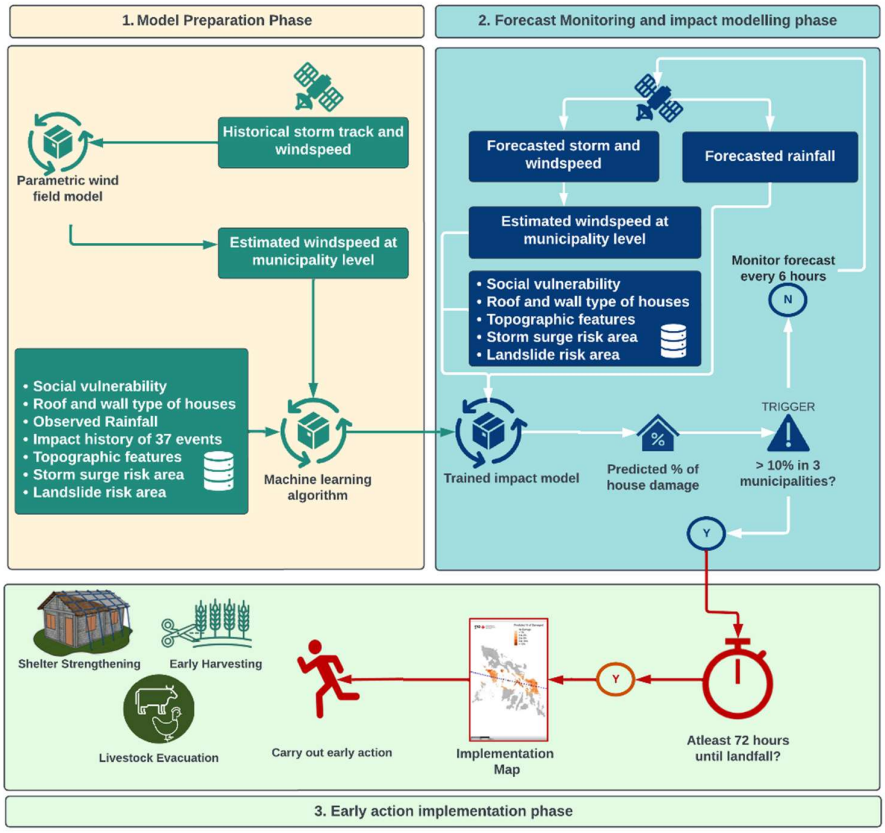


Figure 6 An overview of the steps used in impact-based forecasting for Typhoon early action in the Philippines

The early action trigger in its first implementation was agreed at 10% or more houses predicted to be damaged, affecting at least 3 municipalities and at least 3 days before the expected landfall (PRC et al., 2020).

The damage predictions are made using a statistical impact model developed by 510, (2020). The global ECMWF ensemble forecast for track and intensity is used in combination of a parametric wind field model based on Holland, (2008) to calculate windfields per municipality. This is done using limited storm parameters (intensity, track, forward speed, and structure in terms of radius) and landfall locations; however, the terrain effect is not considered. These parameters from historical typhoon events are used to fit the profiles of wind and pressure, and then converted to estimated windspeed based on the centroidal position of municipalities. This estimated windspeed is then combined with several pre-disaster indicators, topographic variables, rainfall data and landslide susceptibility maps to train the impact model using historical impact data. Once PAGASA has issued a warning for an approaching typhoon, the 510 team runs an operational pipeline using forecasted typhoon parameters on the trained models. This process is run for every 6 hours forecast timestamp to continuously monitor if the impact level reaches the trigger threshold at least 72 hours before the expected landfall.

2.3.3 Typhoon Exposure and Vulnerability

The exposure to typhoons in terms of their frequency and the susceptibility to impact differs spatially across the Philippines. Figure 7 shows the four super regions of the county divided from an economic standpoint (LawPhil, 2009), and the section below discusses their exposure and vulnerability in terms of typhoons, based on the community risk assessment dashboard (510 Global, 2022) which can be found in Annex 1.

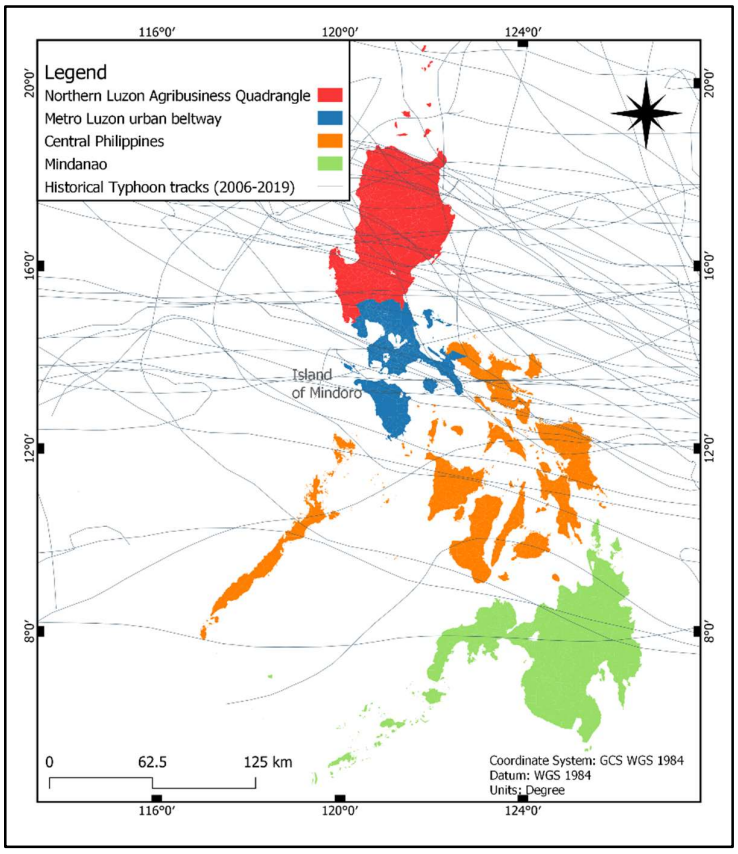


Figure 7 Map showing the four super regions of the Philippines categorized based on economic standpoint

Northern Luzon Agribusiness Quadrangle: Experiences a very high frequency of typhoons every season. Population density is, however, comparatively lower than Metro Luzon. The house types are mostly of solid roofs and walls, making them more resilient to wind damage, and fewer people below the poverty line or belonging to the dependent age group. It also has a relatively higher coping capacity.

Metro Luzon urban beltway: Higher population density and higher frequency of typhoons, but the social and physical vulnerability is less, coupled with higher coping capacity. However, the island of Mindoro (as highlighted in Figure 7) does not share the exact characteristics of higher social and physical vulnerability, even though it belongs to Metro Luzon (Annex 1).

Central Philippines: This region also experiences a very high number of typhoons. The physical and social vulnerability is higher than in Luzon, but it still has a better coping capacity than Mindanao.

Mindanao: This region is exposed to relatively fewer storms every year, but has houses with weaker wall construction material and a significantly high number of people rank low in terms of social vulnerability. So even with a lower wind intensity, the region gets easily affected.

The damage due to wind on a house structure can be primarily attributed to the construction material. A detailed damage assessment report from Build Change for each house type after the devasting Typhoon Haiyan was conducted in the Bohol province of the Central Philippines (Build Change, 2014). The study findings highlight that most houses built with timber were blown away. Those made using a reinforced concrete frame had the walls intact with relative damages in other sections. Structures constructed with confined masonry were seen to have performed better than other designs in terms of impact. A similar classification has also been done by Pacheco et al. (2014) while creating a vulnerability curve across the Philippines for different housing types. Both the studies highlight the importance of solid wall material to prevent the impact of strong winds.

2.3.4 Lessons from the past

Different typhoon events of various characteristics that the Philippines has experienced in the past can be seen as an example of the complexities involved with this hazard, which needs to be considered for effective preparedness.

Case under study - Typhoon Kammuri (2019)

For this research purpose, Typhoon Kammuri in the Philippines was used as a case study, as this was the first time when the T-EAP was triggered and also has a detailed documentation of its implementation strategy. According to the report of NDRRMC, the typhoon, locally known as Tisoy, entered the PAR on 30th November and made its first landfall on 2nd December in Sorsogon province of the Central region (NDRRMC, 2019). The report further suggests that it made 3 other landfalls the next day before leaving the area, followed by intense rainfall and flooding, affecting close to 2 million people.

The T-EAP, which was still under review, was tested for the first time during this event. The 510 team started monitoring the event 120 hours before the landfall, when the first impact map was released (Figure 8) and updated every 6 hours (IFRC, 2020b). The trigger was reached 81 hours before landfall, predicting four municipalities exceeding the threshold, out of which two were selected for piloting the early action. Even after the trigger, 510 continued sending the impact prediction maps every 6 hours to support monitoring and planning for the event (510, 2019). The EAP supported the community through livestock evacuation, early harvesting and shelter strengthening. The damage predictions were limited to within 100 km from the forecasted track, which deviated considerably within 24 hours after the trigger. A post-event review of the EAP suggested that even though the trigger was reached within the expected lead time of 72 hours, the area of target could have been widened more (IFRC, 2020b).

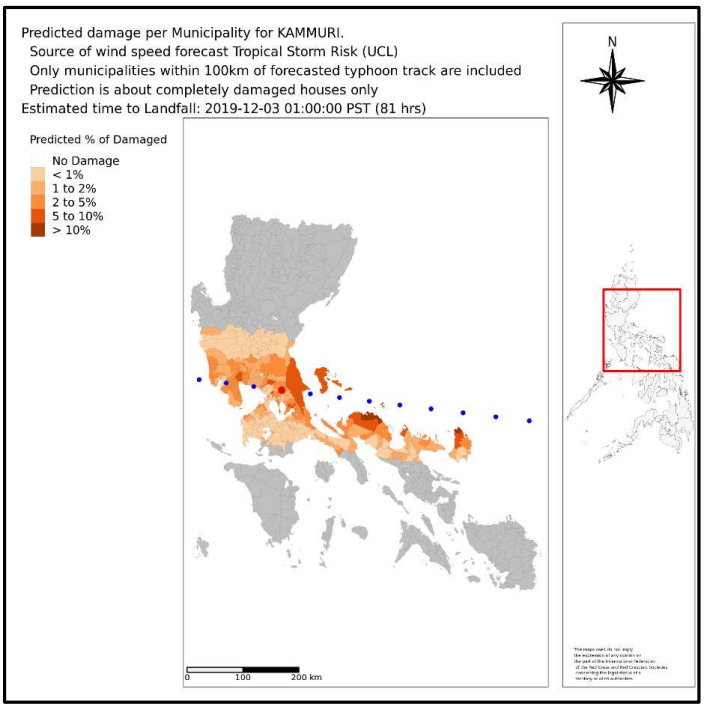


Figure 8 Implementation map with predicted damage at municipality level, that was circulated before typhoon Kammuri to prioritize municipalities for early action. (Source: 510)

Dealing with unexpected events

Research has shown that around 31% of tropical cyclones undergo a rapid intensification (RI) once in their lifetime (Kaplan and DeMaria, 2003). There is no single agreed definition, but most widely a storm is believed to go through RI when its intensity change in 24 hours is similar to the 95 percentiles of typhoons studied over a time period. The incidence of such storms in the past decades has been increasing, which has commonly been attributed to climate change (Masson-Delmotte et al., 2021).

The impact models heavily rely on hazard forecast accuracy and cannot perform better than that. For the research community currently, the biggest challenge in forecasting such events is identifying characteristics that distinguish the RI and non-RI storms (Grimes and Mercer, 2016). One of the reasons why current forecasts cannot locate such anomalies is the relatively low number of RI events for the forecasting models to learn from.

A rapid intensification also has an additional effect on the impact itself, which is not only because of stronger windspeed, but because the information does not reach the people in time. During Typhoon Goni (2020) in the Philippines, the storm, which was initially identified as a tropical depression, went through rapid intensification only a few hours before the landfall, which meant that the model could not predict it with enough time for early action (The Philippines Humanitarian Country Team, 2020). The shock due to these storms is usually more intense when they occur in a less active season (December to February), also commonly termed a Christmas Typhoon. The most recent example of this is Typhoon Rai. In the past decade, the Philippines has seen a 70 % increase in such occurrences in the Western North Pacific (Basconcillo and Moon, 2021). The study reveals that even though the annual cost of typhoon damage in the Philippines has reduced after Haiyan, this is not true for those occurring in less active seasons. Additionally, some storms are slow-moving and bring intense rainfall followed by floods and landslides. The tropical storm Tembin in 2017 that affected Southern areas of the Philippines (IFRC, 2017) is a good example. It becomes crucial to weigh these complexities of storms and the unique nature of their impact, during the planning of FbA.

A recent example - Typhoon Rai (Odette) - 2021

Typhoon Rai is a very recent example of a devastating super typhoon to hit the Philippines in the December of 2021, which underwent rapid intensification. The effect due to this typhoon is believed to be comparable to that of typhoon Haiyan (2013) in terms of infrastructural damage and number of provinces affected. According to the internal reporting from the team of 510, the impact model did not reach its expected trigger before 72 hours of lead time, but continuous monitoring of impact was done until the typhoon made landfall. The retrospective map suggested that the trigger would have reached at a later lead time of 36 hours, but this would not be enough time to carry out targeted interventions under the DREF framework. The humanitarian organization Oxfam was able to support 2600 families with cash support under their B-READY project, with which they are piloting the impact of digital cash transfer before the typhoon to the most vulnerable groups (OXFAM, 2021).

A local humanitarian expert present in the field during the event was consulted to understand the forecast time dilemma for early action (Olaf Neussner, Personal Communication, January 18, 2022). Even though the institutionalized FbA mechanism failed due to an unexpected turn of events, the early action in terms of evacuation did take place, with a timely warning from the government. However, the practicality of taking the actions in time less than 72 hours seems very difficult, according to him. Firstly, the logistics of transporting materials for shelter will take time, especially in remote locations, and a lot of ship transportation is also closed near the landfall time. Once transported, it also needs to be brought to individual houses and put to use. Still, a more community-driven activity, such as a cash transfer carried out by Oxfam, can be feasible in time closer to the event, and people can utilize this money to buy materials from their local stores to strengthen the houses Yet, the effectiveness of cash-based support in minimizing impact, still remains mostly unexplored (Willitts-King et al., 2020).

Strengthening Data Capacity

In 2012, the government of the Philippines launched the Nationwide Operational Assessment of Hazard (NOAH) program to assist in getting a timely warning for approaching floods with the help of WebGIS tools (Langmay et al., 2017). This was a significant step by the government in terms of obtaining risk maps and conduction hazard modelling to support early warning. However, during Typhoon Haiyan in 2013 it was realized that the quality of data was not sufficient to incorporate the variability in the landscape, which resulted in unforeseen impact due to storm surges. As learning from this event, project NOAH created various storm surge hazard maps which are currently used at the national level for evacuation planning (Lagmay and Kerle, 2015). Along with that, the government also funded a project for nationwide lidar mapping, which has assisted in producing a more localized hazard map and can be requested through the LiPAD portal[4]. Additionally, several risk assessment platforms[5] [6] are under development to support real-time decision making. This is evidence that the government of Philippines has been very progressive in handling disaster events by building a strong data capacity.

2.4 An Alternative Trigger Model

Currently, under the DREF there are 3 fully approved EAPs for wind-related hazards. One of them is for the typhoon in the Philippines, which was discussed in section 2.2.2, and one another EAP is for cyclone early action in Bangladesh. The trigger model for this EAP was found to have been developed using a different approach as compared to T-EAP, which is demonstrated in the sections below.

[4] https://lipad.dream.upd.edu.ph/
[5] https://hazardhunter.georisk.gov.ph/map#
[6] https://geomapper.georisk.gov.ph/

2.4.1 Background of the region

Bangladesh lies in the junction of three major river systems and north of the Bay of Bengal, which is known as a hotbed for the most catastrophic tropical cyclones. The country which experiences around four cyclones every year is identified as highly vulnerable to it contributing to the most number of causalities from cyclones worldwide (Saha et al., 2014). The Bangladesh Meteorological Department (BMD) is responsible for the forecast of formation and movement of cyclones in the region and for providing timely warnings to the community. As of 2015, the warning of BMD were not executed into effective actions such as evacuation, as concluded by Roy et al., (2015). Factors such as the unreliability of forecast, ambiguity in warning message, unmanaged and inaccessible evacuation centers were the main reasons why people refrained from acting on warnings. A study by Tanner et al., (2019) also point out that the warnings of BMD are not very user-oriented. Recently many humanitarian agencies such as FAO and IFRC have been supporting the Bangladesh government in institutionalizing early actions for typhoons and floods.

2.4.2 Cyclone Early Action Protocol

Also funded by the DREF, the Cyclone Early Action Protocol for Bangladesh, which subsequently will be referred to as C-EAP, was approved in 2018 (BDRCS et al., 2021). This EAP, which is targeted on 13 out of 16 coastal districts, is carried out at the community level through distribution of food, water, first aid and support for evacuation of people and livestock or other assets. The actions are carried out based on pre-defined hazard thresholds with the minimum expected lead time of 30 hours. Employing the learnings from cyclone Amphan (2020), the EAP has recently been revised and now includes global forecasts with a longer lead time to allow a preparatory phase at least 72 hours before the event. However, according to an expert from the development team of this EAP (Dr. Ahmadul Hassan, Personal Communication, January 17, 2022), within this lead time, the intervention is possible only at a community scale, and it is too late to implement activities such as individual cash transfer.

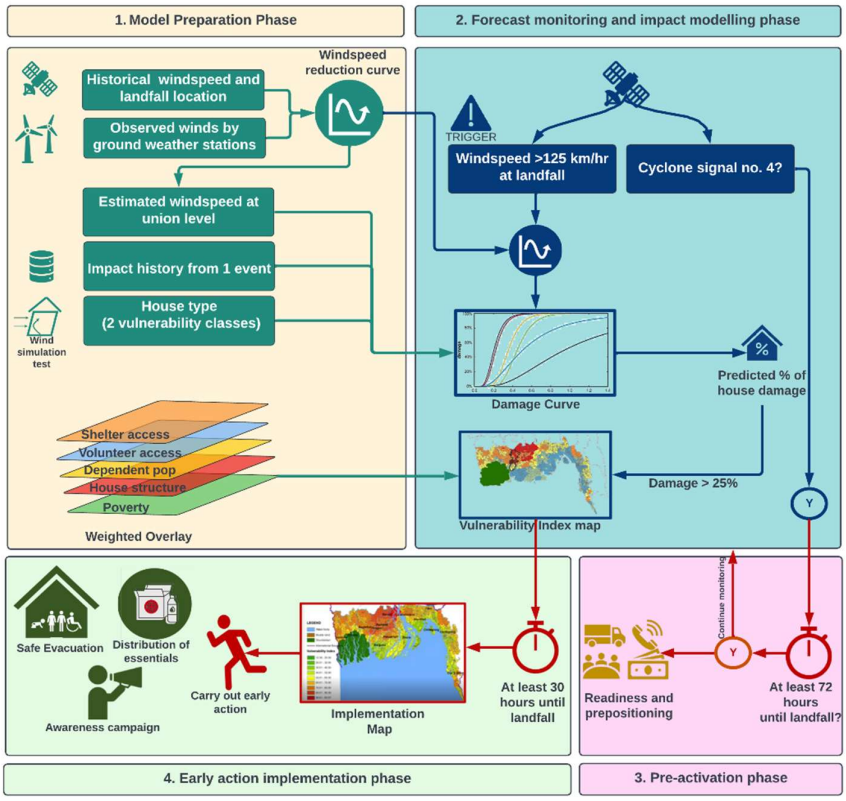


Figure 9 An overview of the steps used in impact-based forecasting for cyclone early action in Bangladesh

As shown in Figure 9, forecasts from BMD, ECMWF, India Meteorological Department (IMD) and Global Forecast System (GFS) are used in this trigger model, and employs a relatively straightforward method for impact predictions, combining damage curve and composite index. The trigger is activated when more than 125 km/hr of wind speed is expected at landfall in Bangladesh. Once activated, a map with predicted windspeed is prepared for the union level - the smallest administrative division. Bangladesh does not have a long term record of cyclone intensity and frequency, limiting the analysis of historical data for wind estimation (Fakhruddin et al., 2022). Currently, this estimation is made by generating a wind speed reduction factor (Figure 10), where 20 field observations of reduced wind speed are plotted against the distance from landfall. Once the expected windspeed is obtained, the impact is predicted with the help of damage curves that have been prepared using one historical event and a simulation done on buildings for identifying damage at different wind speeds. All the unions expected to have at least 25% of houses completely damaged are considered for the intervention. However, due to limited capacity, the unions are then prioritized based on a vulnerability index that is calculated by a weighted overlay method.

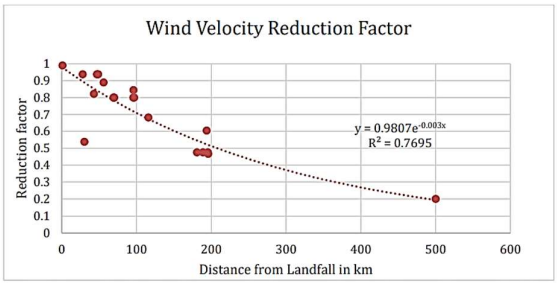


Figure 10 The reduction factor curve that was built using cyclone observations in Bangladesh and is currently used for the operational trigger model for cyclone early action (BDRCS et al., 2021)

2.4.3 Comparison of the two trigger models

The literature above clearly demonstrates that the two trigger models, T-EAP and C-EAP, have significant differences not only in the modelling method of impact but also in the modality of operation (Figure 6 and Figure 9). The differences between the two, based on different indicators, are listed in Table 2 below.

Table 2 A comparison table between C-EAP and T-EAP in Bangladesh and the Philippines, respectively, in terms of different indicators influencing the implementation of early action

	C-EAP (Bangladesh)	T-EAP (Philippines)
Minimum Lead Time	72 hours+ pre activation 30 hours+ activation	72 hours
Activation Budget	177,405 Euro	242,040 Euro
Target	40000 people All sea-facing unions (the smallest administrative unit)	1275 – 1950 households Municipalities from 19 chosen provinces
Forecast for pre-activation	When BMD raises signal 4 warning OR any of the two agencies (BMD, IMD, ECMWF, GFS) forecasts the windspeed to exceed 125 km/hr	-
Forecast for activation	BMD raises danger signal no. VII (out of 11 signals) or forecasted windspeed from BMD/IMD to exceed 125 km/hr	PAGASA informs a typhoon to have entered PAR Probabilistic forecast ECMWF at different timestamp used for models
Accuracy of forecast - track	Roughly the error in track is 100 km in 30 hours lead time with 3 hours error in the time of landfall. Based on cyclone Amphan, 50-250	Avg error of 110 km in 24 hours lead time and 210-300 km in 48-72 hours lead time

	km error for 48-72 hours lead time and 50km for 24 hours lead time	
Trigger Activation (When)	If windspeed greater than 125 km/hr is forecasted at landfall	If predicted damage exceeds 10% for at least 3 municipalities
Intervention area (Where)	Unions with damage exceeding 25% and with priority ranking based on vulnerability index (as many as possible) – capacity of 40000 people	All municipalities identified with more than 10% damage
Windfield calculation	Reduction factor for wind surface velocity (single value) based on the distance of union centroid from landfall using 20 ground observations	Calculates radial profile of wind based on historical data on several wind parameters using Hollands 2008 parametric modelling
Impact calculation	Based on 3 damage curves (for western, eastern and central region) constructed using one historical cyclone and adapted for two different house construction type	Typhoon impact model based on machine learning algorithm using several explanatory variables (related to social, physical vulnerability, geography and weather) with damage history from 37 typhoon datasets
Frequency of impact calculation	Just once when the trigger level is reached	Every 6 hours when the storm enters the PAR until the trigger is reached

3 DATA AND METHODOLOGY

After briefly introducing the study area and describing the principle behind the two IbF models for tropical cyclones, this section will explain the method employed to answer the research questions and the supporting data used.

3.1 Methodological Overview

The foundation of this research was based on evaluating the performance and explainability of models that are operationalized for triggering early action before an imminent crisis. This was done using two different impact modelling approaches and testing their prediction accuracy under different scenarios. The first method was the machine learning model used by 510 for informing early action during typhoon Kammuri in the Philippines. The method was in section 2.2.2. and will from here on be referred to as the statistical modelling. The second approach used was the method of impact prediction for cyclone early action in Bangladesh, discussed in section 2.3.2. Through literature, it was identified that this method uses a combination of damage curves and composite index for impact predictions, and will be referred to as an elementary modelling. Figure 11 demonstrates an overview of how the research was carried out, and Annex 2 lists the sources of the dataset and model codes.

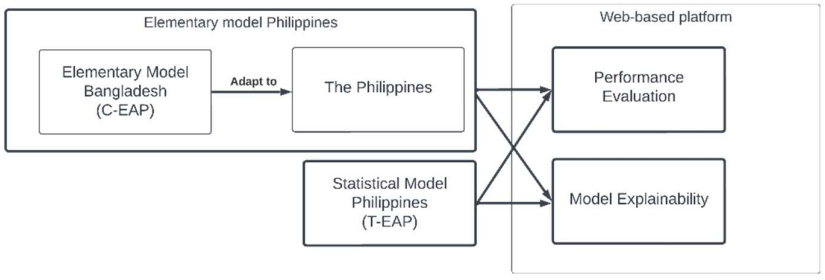


Figure 11 Overview of Research Design

In the first stage, all the input and damage dataset for the study area was prepared, which was used in the impact models and to validate predictions. The performance of the statistical model was tested on baseline conditions and on improving the quality of the demographic variables using the observed weather data. Following that, the damage predictions made during Kammuri at different lead time hours before the event (forecasted weather) were tested for accuracy. In the case of the elementary model, if the evaluation was done based on its performance during a cyclone event in Bangladesh, it would be rather challenging to benchmark the two models against each other. Hence, this modelling approach was adapted for the Philippines and tested on the same event - typhoon Kammuri.

Comparing the prediction results from the two, primarily the effect of data availability and method complexity on the predictions were tested. At the same time, the integration of risk indicators in both methods was assessed to interpret the model performance and discuss their transparency. To achieve that, an interactive decision support tool was built, intended as a prototype for practical interpretation and evaluation of IbF models. The following section describes in detail how each of these steps were achieved.

3.2 Statistical Impact Model

3.2.1 Model setup and input variables

Currently, the statistical model is based on the historical damage dataset of 37 typhoons recorded in the Philippines. This is used as the target variable, which in the context of machine learning is the variable expected to be modelled or predicted. Along with that, there are multiple explanatory variables related to social and physical vulnerability, topography, and weather parameters. An explanatory variable assists a model in finding a trend in the target variable to make predictions. The variables relating to household and vulnerability are used as a percentage of total instead of absolute numbers in the model input. The trained impact model is fed into an operational pipeline, where it automatically pulls the forecasts of an approaching typhoon and emails the prediction results to the concerned authority, as illustrated in Figure 12. The codes for the model were retrieved from the public repository of Red Cross and run locally using the docker platform.

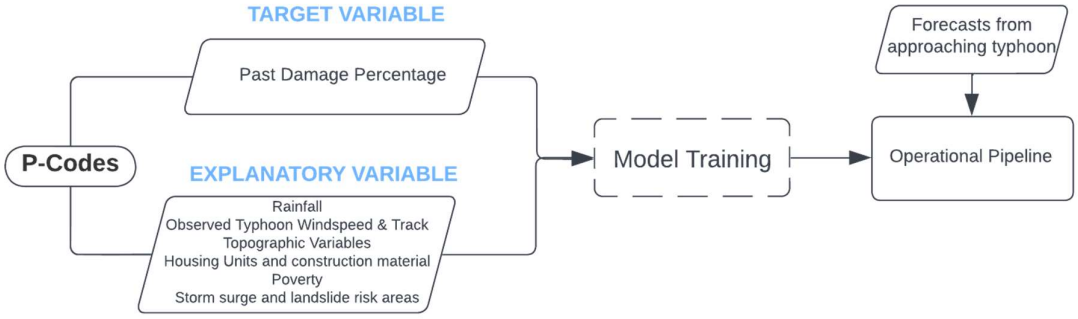


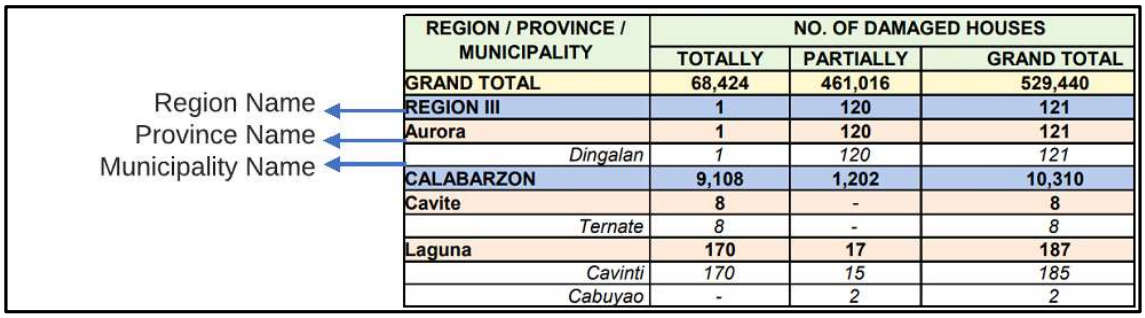
Figure 12 Workflow of the statistical impact model

All the input variables for the model require a unique I.D. (P-codes) to link the datasets and extract new information (more information can be found in Annex 3).

Damage dataset

The target variable used in the impact model, the actual reported damage data, is collected by the Philippines' government and circulated through NDRRMC. For the model input, 510 already has compiled this dataset for 37 different historical typhoons, and it was mostly used as-is in this research. During the communication with developers of the model, it was found that the extraction of these damage data is partly automatized, but also requires manual correction. Consequently, for typhoon Kammuri, which was the case being looked at in this research, this information was re-generated from the raw data of NDRRMC. The reason for doing this was to

increase reliability in the actual damage data, which is crucial for calculating the accuracy metrics of the two model predictions.



REGION / PROVINCE / MUNICIPALITY	NO. OF DAMAGED HOUSES		
	TOTALLY	PARTIALLY	GRAND TOTAL
GRAND TOTAL	68,424	461,016	529,440
REGION III	1	120	121
Aurora	1	120	121
Dingalan	*1*	*120*	*121*
CALABARZON	9,108	1,202	10,310
Cavite	8	-	8
Ternate	*8*	*-*	*8*
Laguna	170	17	187
Cavinti	*170*	*15*	*185*
Cabuyao	*-*	*2*	*2*

Figure 13 A sample section of the situational report of NDRRMC, which lists total damages recorded at the individual municipality level

The ground truth data for household damage caused by Typhoon Kammuri were circulated through situational report no. 19 at the municipality level (NDRRMC, 2019). As a mandate of this EAP, the impact is considered as number of houses 'totally damaged' in a municipality. The data by NDRRMC are provided in a tabular form (Figure 13) in a hierarchical order of region, province, and municipality. The only form of identification, in this case, is the name of the municipality.

While investigating the dataset, it was found that extracting the P-codes directly based on municipality names raises multiple concerns. One of the problems is that the names of region, province, and municipality are inconsistent and do not match within multiple datasets. Another issue is that of duplication, where many municipalities have the same name but lie in different regions and provinces. (Examples can be found in Annex 3)

Considering this, a two-step process was employed where all the mismatches in names were first manually corrected to have a uniformity in naming convention. After that, a unique I.D. was created with a combination of each administrative name in both the datasets (municipality name + province name + region name), using which the P-codes were assigned for linking. This avoided any source of miscalculation due to duplication in names. Once these damage data were obtained, it was cross verified with the operational dataset being used in the statistical modelling, the results of which will be discussed in Chapter 4.

Household dataset

Another set of data used in the model as an explanatory variable is the number of households in each municipality segregated based on roof and wall construction materials. Currently, the historical events in use range from 2006 (Typhoon Durian) to 2020 (Typhoon Goni). However, the pre-disaster variables associated with household numbers are based on the official census conducted in 2015. The timeliness of dataset, in terms of how often it is updated, is a relevant test for quality of model input (van den Homberg et al., 2018). As this study intends to test the effect of data quality in impact predictions, the household number was replaced with statistics based on census closer to the event date as explained below.

The census in the Philippines is conducted every five years, and since 2015 it has been administered by the Philippines Statistical Authority (PSA, n.d.). The dataset for the official population of three different years (2010, 2015 and 2020) was retrieved from PSA, (2021), and each Typhoon event was assigned values from the closest census possible (Table 3). An issue of duplication and non-uniformity in the municipal names was found in this dataset as well. Considering that, while assigning the P-code, a similar approach for unique ID generation was utilized in combination with manual correction, as discussed earlier in this section.

Year of Typhoon Event	Year of Census
2006, 2008, 2009, 2010, 2011, 2012	2010
2013, 2014, 2015, 2016	2015
2018, 2019, 2020	2020

Table 3 A list of years when different Typhoons existing in our dataset made landfall and the census data assigned

Once the population statistics for the different years were obtained, the next step was to generate the total household number for each municipality segregated based on construction material. For this, the relation between population to household numbers for 2010 and 2020 had to be generated based on existing statistics from 2015. There is a possibility that major events such as typhoon Haiyan could have been a catalyst in bringing significant changes in construction practices over a decade. This would mean that the distribution of houses in categories of construction material may be different in 2020 as compared to 2010. For this reason, stakeholders working in the reconstruction programs in the Philippines were consulted. Through them, it was confirmed that there is minimal change in construction behaviour due to several reasons: no national programs effectively promote disaster-resilient construction, implementation of the building code is very poor, and low-income families rely on cheap and poor-quality materials (Build Change, Personal Communication, January 31, 2022). A surge team deployed after Typhoon Odette in the Philippines also confirmed, that the biggest challenge in the field is the message of building back safer not being incorporated well during response activities. Instead of a safer choice, the homeowners opt for a faster way of rebuilding their houses (510 Internal Reporting, March 17, 2022). An assumption was thus made that the number of houses in each construction typology has remained largely similar over the decade. Therefore, the distribution of roof and wall type for 2010 and 2020 was obtained by interpolation from the existing statistics of 2015.

3.2.2 Performance Metrics

Once the dataset was prepared, the model was evaluated for two scenarios: existing conditions and a modified dataset based on section 3.2.1. The model works on the binary classification of municipalities based on the percentage of damage noted in the trigger threshold. This existing architecture was used for the research, making changes in the script where required. Two optimised machine learning algorithms are currently utilized for impact prediction: Random Forest and XGBoost. Both of these algorithms are tree-based methods designed for regression and classification problems (Breiman et al., 2017). The optimised hyperparameter generated using K-fold cross-validation was applied to the validation set to test the accuracy. Three accuracy metrics were used for evaluating the results, defined as the following:

$$\text{Precision} = \frac{|TP|}{|TP|+|FP|} \qquad \text{Recall} = \frac{|TP|}{|TP|+|FN|} \qquad \text{F1} = \frac{2\times Precision\times Recall}{Precision+Recall}$$

T.P. True Positive
TN True Negative
FP False Positive
FN False Negative

Figure 14 Accuracy metrics for testing the performance of the models based on the actual and classified damage

A precision score is based on the classification, and shows, out of all the predictions of the model, how many are actually true. On the other hand, a recall score is based on ground truth and shows the ability of the model to identify the municipality with actual damage. An F1 score represents the harmonic mean of precision and recall and can be used to assess the trade-off between the two.

The operational pipeline can be run automatically once the Typhoon has entered the PAR, starting by retrieving forecast information from the ECMWF database. In case when there is no active typhoon, the forecast information is extracted from a remote directory where ECMWF stores predictions for past storms. However,

this information is only stored for a duration of 180 days, which meant that for typhoon Kammuri, this dataset was unavailable limiting our ability to test the model performance on forecasted windspeed. In that case, for comparison purpose with the elementary model, the existing database for predictions made during Kammuri in 2019 were used.

3.3 Elementary Impact Model

As already discussed above, an elementary technique used in Bangladesh was adapted for the study area (the Philippines) and evaluated for a historical event. Figure 15 shows an overview of four major steps taken in preparation of the trigger model on the basis of model description given earlier in Figure 9.

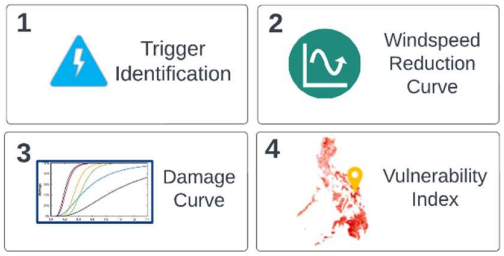


Figure 15 Steps in preparation of an elementary trigger model

3.3.1 Trigger Identification

The first step for this elementary modelling technique is to identify the hazard threshold above which the early action would be triggered. The example case of Bangladesh uses a trigger of 125 km/hr of windspeed predicted during the landfall. As per the IFRC mandate, an event is defined as extreme when it has a return period of 5 years, or has a 90th percentile of magnitude in all historical scenarios (IFRC, 2020a). This is considered the point where substantial damage can occur, and this way, an EAP is expected to be activated at least once every five years, which otherwise is revised.

To identify this trigger value, a study on return periods of typhoons done by Espada, (2018) was referred to. The study looks at all the major storms in the Philippines in the last 45 years, and investigates the maximum sustained wind and minimum central pressure at different return periods, along with damage and death counts. Based on this literature, the one in 5-year return period maximum sustained wind and the 90th percentile of all windspeed in that time frame was identified. The threshold of hazard for triggering action was then considered to be the minimum of the two values.

Alternatively, C-EAP also starts monitoring, as a pre-activation stage for EAP, when BMD raises a level 4 warning, which is defined as: "The port is threatened by a storm, but it does not appear that the danger is as yet sufficiently great to justify extreme precautionary measures." - (BMD, 2014). A similar warning level was identified for the Philippines based on PAGASA's recently revised wind warning signals (PAGASA, 2022). This warning level was concluded to be the pre-activation stage at least 72 hours before the predicted landfall to plan and implement any preparatory actions that can support in the later stage.

3.3.2 Windspeed Reduction Curve

Once the hazard is identified, the next step is to develop a factor to estimate the wind speed experienced at each municipality level: the intervention scale. The reduction factor in this case is defined as the fraction of windspeed that is felt at a certain distance in comparison with that reported at the landfall location.

$$Reduction\ factor\ at\ distance\ x\ from\ landfall = \frac{Actual\ windspeed\ reported\ at\ lanfall}{Observed\ windspeed\ at\ distance\ x\ from\ the\ landfall} \quad (1)$$

Wind Observations:

The actual (1-minute averaged) windspeed reported for different typhoon event was obtained from IBTrACS. The observations for wind that reached different location were taken from ASTI-DOST, (2022), which has deployed 255 automated weather stations (AWS) across the country for measurement of various parameters, including wind speed. ASTI also uses an automated quality control system to ensure consistency in crucial elements such as the range and spatial correctness of these datasets, since they are being used in very sensitive decision support, such as disaster risk (DOST-ASTI, 2018).

To construct the reduction factor curve, observations had to be taken from stations at various distances from the landfall location, and at a time very close to when the landfall occurred. For this purpose, the observations were made for multiple typhoons with different intensity and landfall regions to have a better generalization capacity (Figure 16). Though these wind observations are publicly available for use, an issue to note is that the datasets are not in a downloadable tabular format, which would have allowed a faster retrieval. Given that, the stations had to be individually examined for records around the landfall time. While compiling this dataset it was also found that there are very limited stations that are close to the landfall, and some of them have missing records during the day of the storm. The stations were initially deployed in 2010, but it was found that many of them have stopped operating for the past few years, and some have gaps in between. A total of 45 observations were collected for nine different historical typhoons for the construction of reduction curve.

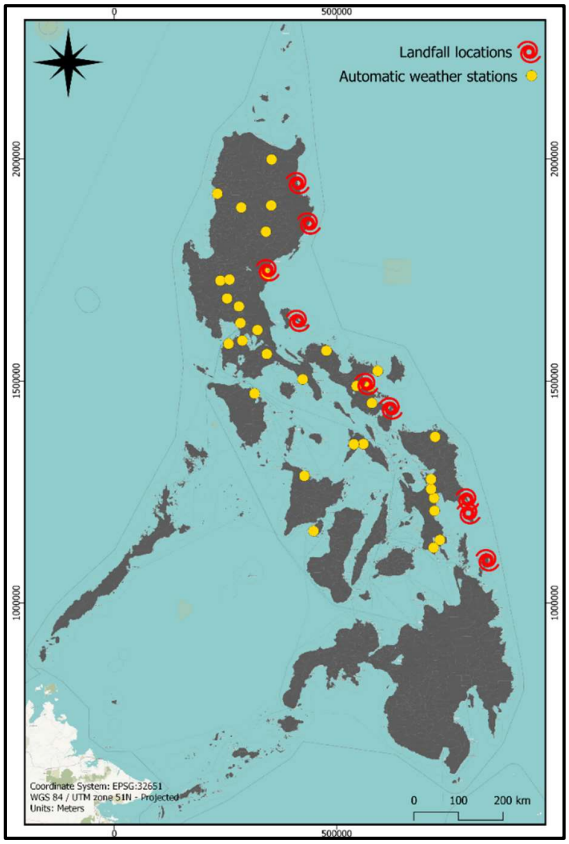


Figure 16 A map of the Philippines representing the location of automated weather stations and historical typhoon landfall locations used for constructing the reduction factor curve

The geo-coordinates of observed landfall are not available in any of the repositories. Hence, manual extraction was done by intersecting the typhoon path (Knapp et al., 2010) with the coastal boundaries of the country. The locations were then cross validated with the names of municipality where the landfall was reported, and

adjustments were made accordingly. The shortest distance between each observation station to the corresponding typhoon landfall was then calculated in ArcGIS software.

Normalization of Observations:

Since the observations are taken from multiple stations, factors such as the exposure at the site, the elevation where it is placed, and averaging time of windspeed, may affect the value of recorded observations. To normalize this value, the approach of Powell et al., (1996) was adopted, which describe a method to correct these differences and obtain a common framework for usage.

The normalized observation is given by the following equation and explained in Figure 17:

$$\frac{U10}{Uz} = \ln\left[\frac{10-Zd}{Zo}\right] / \ln\left[\frac{Z-Zd}{Zo}\right] \quad (2)$$

Uz	= 10mins sustained mean wind measurement at a height z
U10	= Wind speed at 10m height
Zd	= 75% of the height of the blocking object (H)
Zo	= Roughness coefficient (range 0 – 1)
Z	= Elevation of the station from the mean sea level

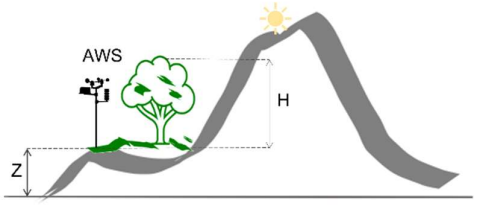


Figure 17 An illustration to describe the parameters used for normalizing the windspeed of AWS

The first step was to adjust the measurement to a 10m elevation from sea level. In an implementation of this method, Powell et al., (2010), consider 'Zd' to be 0, unless there is very poor exposure at the site, and the same was adopted for this research. The study also describes the surface roughness (Zo) as a subjective estimation, based on satellite images of the area where the stations are placed, and a quantitative method to do that is provided by Wieringa, (1992). A visual inspection in Google Earth imageries was done for all 45 stations and their surroundings to make an estimation of the roughness coefficient for each. It was also ensured that the date of the historical imagery referred to was around the same time as that of the event date. Using equation (2), windspeed adjusted to 10 meters height was obtained. The literature also suggested the observations to be converted to 1-min sustained wind speed which was done by multiplying the observation by in land conversion factor of 1.21 as per WMO, (2010).

Using the normalized observed windspeed and the actual windspeed at landfall, a reduction factor was computed for all 45 observations using equation (1). Finally, a windspeed reduction curve was then obtained by plotting this reduction factor against the distance from its corresponding landfall. Along with that, the resulting curve was also evaluated against the one built for C-EAP (Figure 10). This will be discussed in the result section 4.3.2.

3.3.3 Damage Curve

The next step was to obtain a damage curve that employs an empirical method of understanding the physical vulnerability based on historical damage statistics and its relationship with wind magnitude at each municipal level. The C-EAP uses the impact data only from one historical cyclone based on a very intensive field survey of the damage. For the Philippines, there exists a considerable amount of past impact ground truth information, as explained in section 3.2.1. Using that dataset, the percentage of damage for each municipality per event was calculated.

$$\text{Percentage of damage} = \frac{\text{Number of houses fully collapsed in Municipality X}}{\text{Total number of houses in Municipality X}} \quad (3)$$

For each event, the actual wind speed experienced at the municipal level was then estimated based on the factor of windspeed reduction obtained in section 3.3.2. The distance from the centroid of each municipality to the landfall location was calculated for that purpose.

Windspeed in municipality X = Factor of windspeed reduction * Maximum Windspeed at landfall (4)

Before constructing the damage curve, the aggregation of damage in different categories was done to have a better fit in the estimations.

Categories for Damage Curve

The vulnerability and typhoon exposure in the Philippines was discussed in section 2.2.3 based on the classification of super regions. It was observed that the four areas have stark differences in terms of the number of typhoons they experience, and also the extent of the impact. Consequently, an individual damage curve was prepared for each of these four regions to account for the differences.

From the literature it was also noted that even though the super regions follow a general characteristic, there are still areas, such as Mindoro, that do not follow the same trend in terms of vulnerability. To account for the differences in physical vulnerability, another classification was made on the basis of construction material used to build houses. For C-EAP, a detailed model simulation was done to test different typologies of houses, and at what wind speed they entirely collapse. Since doing that was not feasible in this research, the classification was obtained based on literature, which has also been described earlier in section 2.2.3.

Through the works of literature, it was observed that the material used for the wall construction in a house is very crucial in determining its likely impact level. Considering this, the houses with stronger wall type were put in a less vulnerable category, and everything else was assigned as highly vulnerable, as shown in Table 4. The housing unit data for the Philippines are consolidated at the municipality level, and there is no damage information available for individual houses. Hence, a sum of house types in each municipality was calculated in the two damage likelihood classes, and they were assigned a value of the majority class. This way each municipality had a corresponding damage curve based on its region but also the predominant house construction type.

Table 4 A list of house categories (based on construction materials) and the vulnerability class assigned to it for the damage curve

House type	Damage likelihood
Strong Roof/Light Wall	Highly vulnerable
Strong Roof/Salvage Wall	
Light Roof/Light Wall	
Light Roof/Salvage Wall	
Salvaged Roof/Light Wall	
Salvaged Roof/Salvage Wall	
Strong Roof/Strong Wall	Less vulnerable
Light Roof/Strong Wall	
Salvaged Roof/Strong Wall	

The percentage of damage were plotted against their corresponding wind speed after which a best fit line for each category was generated using a linear model function in R programming. A total of 12 different damage curves were obtained: 4 per region with 2 damage likelihood class and one overall curve. The trend of these curves in each category is reflected upon in section 4.3.3.

3.3.4 Vulnerability Index

In the final step of this trigger model, there needs to be an assessment of the vulnerability index which will be used to prioritize the areas for intervention. This index was generated based on a weighted overlay method where multiple indicators and the quantification of the relative importance of each can be used for decision making (Greco et al., 2018). The datasets for the indicators were prepared as explained below:

Poverty: Poverty incidence is defined as "the proportion of families/individuals with per capita income/expenditure less than the per capita poverty threshold to the total number of families/individuals" (PSA, 1997).

Vulnerable Age Group: This was defined as a sum of the number of people in a municipality that belong to the following category – Child headed, solo headed, disabled, single parent, or older population.

Evacuation Center: Based on the total number of evacuation centers in each municipality and assuming a capacity of 10000 per center, the unserved percentage of the population was calculated.

House Typology: This was defined with a similar approach as explained earlier in section 1.3.3; however, classified into four instead of two categories. Each category of houses was given different weightage (adapting the C-EAP) in terms of their vulnerability to different wind speeds. The index was the sum of houses in each category within a municipality, multiplied by its corresponding weightage as shown in Table 5.

Table 5 A list of house categories based on construction materials and the corresponding vulnerability class with weights assigned for physical vulnerability index calculation

Vulnerability to wind	Weightage	House Types Included
Very High	50	Salvaged Roof/Salvage Wall
High	30	Strong Roof/Light Wal, Strong Roof/Salvage Wall, Light Roof/Salvage Wall, Salvaged Roof/Light Wall
Moderate	15	Light Roof/Strong Wall, Salvaged Roof/Strong Wall
Low	5	Strong Roof/Strong Wall

The index obtained from all the four indicators was normalised to a scale of 0 to 1 and multiplied with different weights (adapting the C-EAP), as shown in Table 6, to calculate the overall vulnerability index at the municipality level.

Table 6 The weights assigned to each indicator for calculating the overall vulnerability index at the municipality level

Indicator	Weightage
Poverty	35
Vulnerable Age Group	15
Evacuation Center (Unserved)	15
House Typology	35

This vulnerability index of each municipality was used along with the predicted impact for prioritization.

Once all four variables of the elementary modelling were obtained, the integration of hazard, vulnerability and exposure was evaluated against the existing statistical model to achieve Objective 1.

3.4 Model Evaluation

As the next step, the two modelling methods were then tested for their prediction accuracy and explainability on different scenarios (Objective 2). In context of this research, the explainability has been tested in two ways; (1) case specific post-hoc interpretations based on the performance during the event of Kammuri, and (2) effect of change in input parameter on the prediction accuracy. To effectively communicate these trade-offs, an interactive web-based platform was built, which did not only make this process efficient, but can also act as a potential decision-support tool to evaluate various alternatives (Objective 3). The predictions made from the existing statistical modelling method were used in this platform as a static input, only to visualize the predictions

and accuracy metrics. In the case of the elementary impact modelling method, which was developed as a part of this research for the Philippines, the entire process described in section 3.3 was automated and visualized in an interactive decision support tool developed using the R package "shiny" v.1.5.0 (Chang et al., 2020). The tool allows the users to input the forecasted hazard information and adjust different variables for a quick calculation and visualization of impact predictions. Each element is meant to serve a purpose of increasing the transparency of the prediction algorithm and at the same time testing the choices or uncertainties involved an IbF. A need for inclusion of the following components was identified, and their importance have been briefed below:

Uncertainties: The platform incorporates forecast uncertainty in terms of margin of error in forecasted location and intensity of storm which are the two dependent factors in the elementary modelling. Along with that, the number of historical events in preparation aggregation of damage curve is also added to highlight the effect of quantity and quality of historical loss data.

Alternatives: This includes dynamic variables which permits the user to make certain choices and visualize the change in the prediction based on that. The prototype dashboard currently includes the two-modelling methods, trigger threshold and number of municipalities to intervene as choices for the user.

Intermediate Results: Damage curve and vulnerability index maps were also incorporated in the platform. These elements are added to permit the users to view the intermediate results which is driving the predicted losses making it better interpretable.

Actual and predicted maps: This allows the users to easily visualize and interpret the predicted loss against what was recorded in the ground. Along with that, multiple past typhoon events were included in a drop box. Each event will have unique characteristics in terms of how they progressed, and an evaluation of those scenarios may also be highly relevant in context of model performance. This will further support in understanding how and to what extent does forecast uncertainty propagate to the model.

3.4.1 Modelling technique

The study primarily intended to compare the accuracy metrics obtained for the same event using the two methods: statistical and elementary modelling.

For elementary modelling, the first step was to obtain the landfall location at the time when the wind speed was predicted to exceed the trigger level identified in section 3.3.1. The data were retrieved based on different reporting made through media and news platforms. The actual time of landfall of this event is known to us, i.e. 2nd December 2010 at 23:00. This time stamp was used to decide the lead time at which the prediction of our trigger threshold was made, which is shown in Table 7 along with landfall locations at different lead time is illustrated in Figure 18.

Table 7 Predictions of landfall location and maximum wind speed made for Typhoon Kammuri at different lead time hours

Maximum Windspeed	Date	Lead time	Location	Geo- Coordinates	Source
215 km/hr (Actual)	2nd Dec 2019 23:00	0	Gubat Municipality, Sorsogon	124.119 E 13.005 N	NOAA historical hurricane track
185 km/hr (Forecast)	1st Dec 2019	24	Sagnay Municipality, Camarines Sur	123.550320 E 13.595774 N	News report, JWTC[7] forecast
166 km/hr (Forecast)	29th Nov 2019	72	Pandan Municipality, Catanduanes	124.214001 E 14.086363 N	News report, JWTC forecast

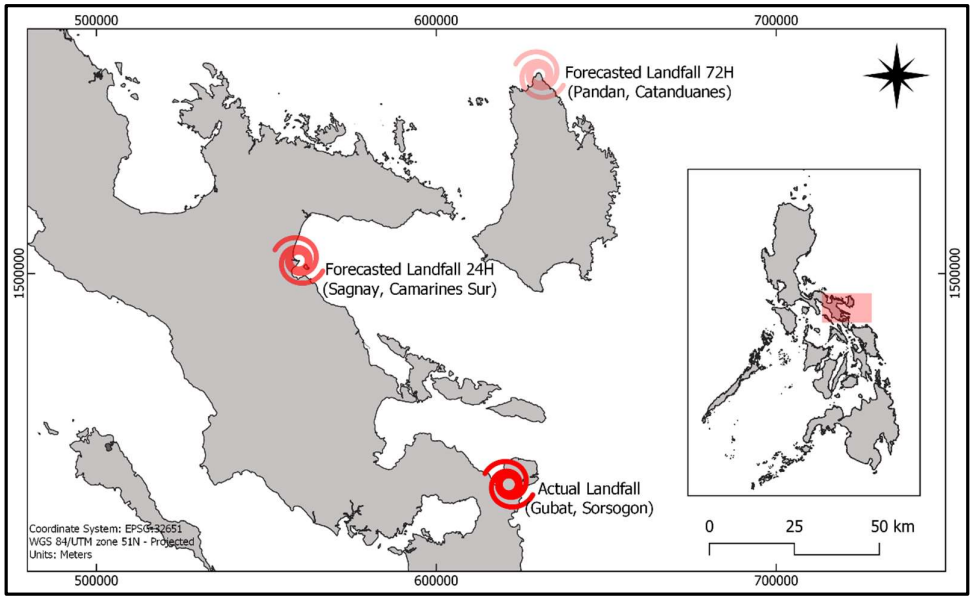


Figure 18 A map representing the actual landfall location of typhoon Kammuri along with forecasts made at different lead time hours

The distance from all municipality centroids to the landfall location was calculated as a next step. Based on this distance and the forecasted wind speed, the factor of reduction obtained through section 3.3.2 was used to estimate the wind speed at each municipality. Finally, each municipality's expected impact (in terms of percentage of house damage) was calculated using the damage curves obtained through section 3.3.3. For each municipality, the corresponding curves based on its region and majority house type was used. All the municipalities that exceeded the damage threshold were selected for the intervention in this scenario.

A common damage threshold was then applied to the actual damage reported for the event, as explained earlier in section 3.2.1, as well as the modelled results (from both elementary and statistical technique). A binary classification was done where any municipality exceeding this threshold is assigned a value of 1, while the rest is set at a value of 0. The performance of the model was then tested using the accuracy parameters discussed earlier in section 3.2.2, explained as:

True Positive (TP) Exceeding the damage threshold in both actual and modelled results

False Positive (FP) Exceeding the damage threshold in modelled result but not in reality

[7] Joint Typhoon Warning Center (JTWC)

True Negative (TN)	Does not exceed the damage threshold in both actual and modelled results
False Negative (FN)	Exceeds the damage threshold but modelled otherwise

3.4.2 Threshold for selection

One of the criteria of selection in the trigger models is the damage threshold above which impact is deemed to be problematic. Along with that, the elementary modelling technique also requires a specification for the number of municipalities to be prioritised for action. Both selections might be affected by various criteria, including the financial policies of the organisation, scale of implementation or even the vulnerability of the target community. Considering that, a flexible range of 0 to 50 % for damage threshold, and 0 to 100 for priority municipality, was tested at every interval of 5 steps. This will cause no change in the impact prediction, but the accuracy metrics that depend on these factors, as explained in section 3.3.5, will be affected.

3.4.3 Lead time

Lead time before the typhoon landfall at which the forecast information is used is also a crucial factor for FbA, which has already been highlighted in section 2.1.1. To demonstrate this, the accuracy of impact was tested using the predictions made during the event, which was available for the following lead time: 81 hours, 60 hours, 51 hours, and 45 hours before the expected landfall.

For the trigger model using an elementary technique, in a real scenario, there is only one lead time at which the prediction is made; once the hazard threshold is reached. However, forecasts at two different timestamps (Table 7) were tested for the purpose of this research: 72 hours, 24 hours and 0 hour (the observed data).

3.4.4 Quality of data

Forecast Data

The skill of the forecast provider is a crucial factor for IbF, because even the most accurate impact prediction models cannot do better than the weather forecast models. The elementary modelling technique relies mainly on the location and maximum windspeed predicted for landfall. To allow an evaluation of forecast uncertainty in this approach, a landfall location error ranging 0 to 500 km and maximum windspeed error running from -50 to +50 km/hr was tested. For example, for an error of 200 km, an assumption will be made that the landfall can occur anywhere within 200 km radius from the forecasted location. Hence the distance to landfall from all the municipalities lying within this buffer will be considered as 0. Similarly, a selection of +20 km/hr of windspeed uncertainty would mean that this would be added to the forecasted windspeed before performing the impact calculation.

Damage History

Along with hazard forecast, the quality of the pre-disaster dataset can potentially impact the performance of the model. In the elementary modelling method, the most critical historical dataset used for prediction is the actual damage data reported for past typhoons. As explained in 3.3.3, the Philippines has an extensive record of damage dataset compared to Bangladesh. So instead of using only one event, different number and combination of past typhoon data was used to test the construction of the curve. A total of 7 events used to build the reduction factor curve have been put in this list based on the order of highest impact. The difference in the fit of the curve in each scenario and its effect on the accuracy was tested and critically evaluated.

In summary, the trigger model of C-EAP was adapted for the Philippines, generating the required equations and this, along with statistical model, was tested for prediction accuracy of damage during typhoon Kammuri in different scenarios. The upcoming sections will present the key findings of the study.

4 ANALYSIS AND RESULTS

In this chapter, the individual analysis results are provided.

Section 4.1 presents the results of the statistical modelling. It includes findings on the quality of input data, specific to population and damage statistics. Along with that it also describes how the model's performance is affected by the change in these datasets.

Section 4.2 examines the results of the adaptation of the elementary modelling method for the same study area. It covers the derivation results of various risk indicators.

Section 4.3 compares the elementary and statistical modelling techniques and their impact predictions across the lead time and across various scenarios. It also highlights the implication of the decision support tool developed for making this evaluation.

4.1 Statistical Impact Model

4.1.1 Quality of input variables

Figure 19 outlines the trend in population over the years, calculated for super regions based on the census statistics obtained for 2000, 2010, 2015 and 2020.

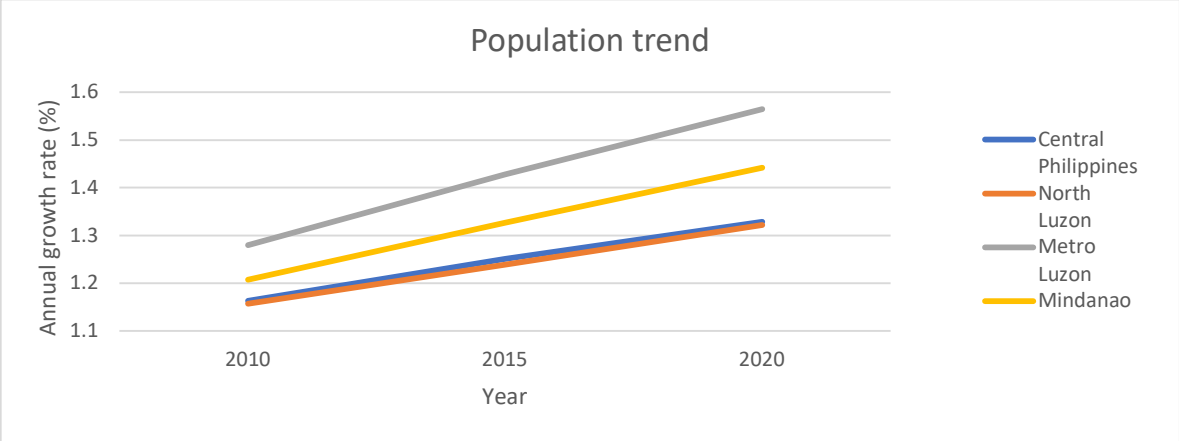


Figure 19 The population growth rate in four super regions of the Philippines over the past two decades

As can be seen, the growth rate in the population has been sharply increasing over the years, the highest being in Metro Luzon. The existing input dataset of the model were assessed for the degree to which it is influenced by these changes in population. Out of 272 municipalities affected by typhoon Kammuri, 33 exceeded the 10 % damage threshold based on the household dataset from the 2015 census. When this value was recalculated using the statistics of 2020, the maximum change observed in the individual municipality was 2% from the threshold.

The ground truth of damage from Kammuri, upon validation with the current model input, also showed a considerable discrepancy. There were 49 municipalities (out of the 272) that had an error in the existing input database and are factored as having no damage. Out of these, 13 municipalities had damage exceeding the 10% threshold, which can affect the identification of target areas. Upon examining, it was found that most of the errors corresponded to municipalities either having duplication or mismatch in the naming convention, which was resolved through manual editing during this research (Examples in Annex 1).

4.1.2 Model Prediction

Table 8 below shows the accuracy metrics of this model, with two different qualities of input parameters. With an improved dataset, one would assume that it will result in a change in model's performance. However, the model's accuracy did not appear very sensitive with these adjustments relating to the dataset proximity to the event and a revised damage count as shown in Table 8.

Table 8: Accuracy metrics of the two algorithms for the baseline dataset versus one with an update in demographic variables

	Census data - 2015			Census data - 2010, 2015, 2020		
Models	**F1 score**	**Recall**	**Precision**	**F1 score**	**Recall**	**Precision**
Random Forest	0.62543	0.614588	0.63666	0.626434	0.614588	0.638746
XGBoost	0.585663	0.518685	0.672504	0.581651	0.512382	0.672577

4.2 Elementary Impact Model

The following section will present the results of each step in designing a trigger model for the Philippines, by adapting the technique of C-EAP.

4.2.1 Trigger Identification

Based on the study of Espada, (2018), the one in 5-year return period maximum sustained wind was found to be approximately 165 km/hr. Similarly, the 90th percentile of all windspeed in 45 years duration was observed as about 180 km/hr. For defining a trigger, the minimum of the two values was considered, to allow a higher margin of safety. Hence using this scenario, early action would be activated once the maximum sustained wind speed at landfall is predicted to be 165 km/hr or more.

A pre-activation trigger was also identified based on a corresponding warning signal level for storms in the Philippines. When compared with PAGASA, the warning resembles signal no:2, where a wind speed ranging 62-88 km/hr is expected in the next 24 hours, likely to cause minor to moderate threats to lives and infrastructure. The release of this signal will be the pre-activation stage in this trigger model.

4.2.2 Windspeed reduction curve

For determining the factor of windspeed reduction, a total of 45 observations were identified from the AWS, corresponding to 9 different storms (Haiyan, Sarika, Haima, Yutu, Kammuri, Phanfone, Goni, Vamco, Odette). The distance of landfall location of the storms from the observation stations ranged from 9 to 453 km, and the maximum windspeed recorded during landfall by a station was 117 km/hr.

During the normalization of observations, it was noticed that the stations had substantial differences in elevation at which they were placed (5 to 2881 meters) as well as their surrounding environment (Roughness coefficient 0.1 – 0.75). Annex 4 can be referred to for a few examples of google earth imagery of the station location and the coefficient assigned to it after visual inspection.

Figure 20 shows the reduction factor curve constructed from the recorded windspeed at different distances from landfall, after normalization. A comparison of the curves before and after normalization can be found in Annex 3.

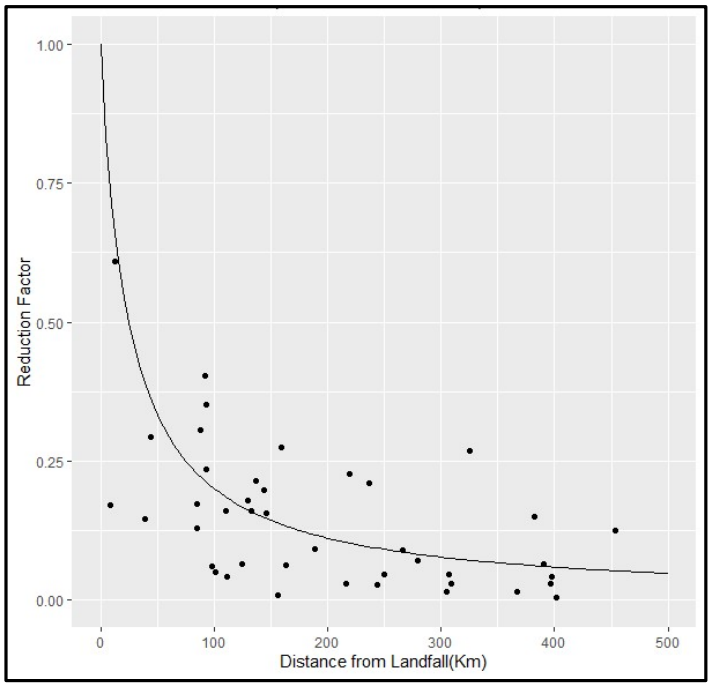


Figure 20 Windspeed reduction factor curve obtained for the Philippines, with factor of reduced windspeed (AWS - observations) against the distance from landfall plotted after normalizing

From the curve, it can be noted that majority of values are at the bottom of the graph, suggesting a very high reduction in wind speed. Within 50 km from the landfall, there is a steep decay in the curve, where only 30 per cent of the intensity is felt, after which it shows a more gradual drop. Furthermore, at a distance of 100 km, the values of reduced wind ranged from 5 to 35 %, indicating a variations in observation.

The results obtained for the Philippines were compared with that being used operationally for C-EAP, as shown in Figure 10. As can be seen, the fit of the two curves is quite dissimilar even though both have the same characteristic hazard. The curve constructed for Bangladesh shows that most of the areas that lie within 100km from landfall experience almost 80 percent of the landfall windspeed. This curve also shows a better fit with fewer deviations in observation.

The reduction equation from the normalized curve was further used in the next steps to derive the windspeed for all the municipalities based on the landfall location of the windspeed events.

4.2.3 Damage Curve

Figure 21 shows the damage curves obtained for the four super regions and two physical vulnerability classes of each. The curves are based on loss data from 7 different historical events for which the windspeed at each municipality was estimated based on the results of 4.2.2.

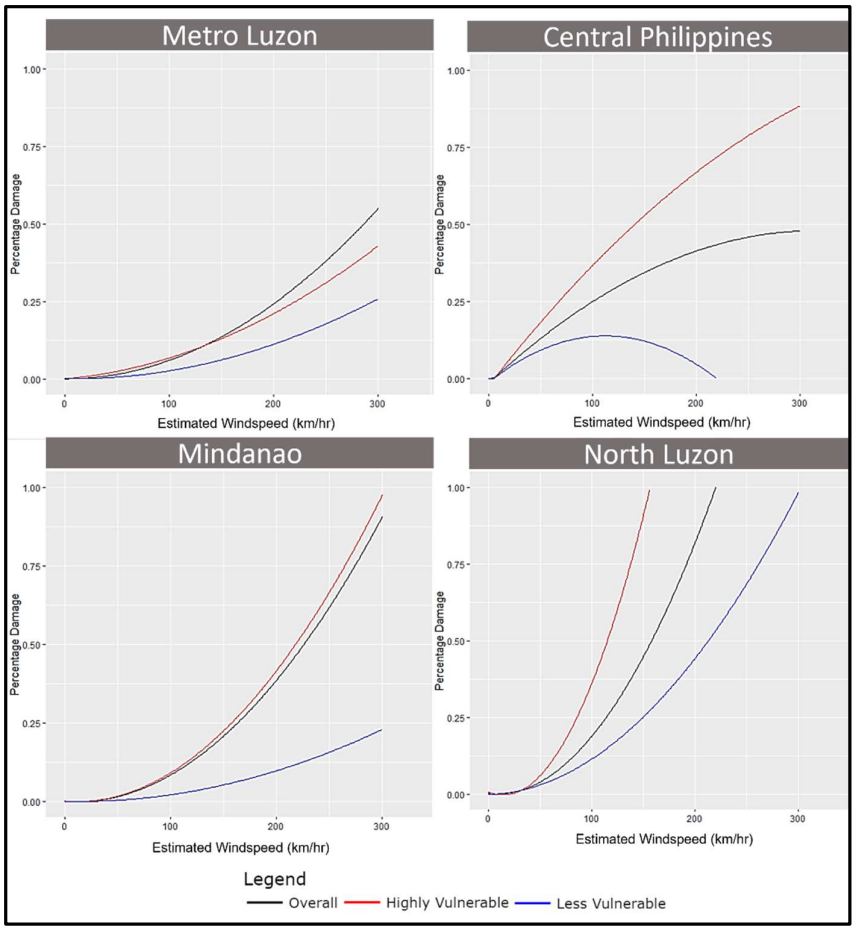


Figure 21 Damage curves from 7 historical events for 4 super regions based on the windspeed experienced by each municipality and the fraction of houses that were totally damaged

The curves show that at a same windspeed, municipalities with predominance of highly vulnerable buildings show a larger fraction of damage as opposed to those with less vulnerable structures. Similarly, in terms of regions, metro Luzon shows higher resistance to windspeed compared to the other three, and a significant damage, higher than 10%, is seen only after windspeeds of 150 km/hr. In contrast, for North Luzon and the Central Philippines, this damage corresponds to a windspeeds of 50 km/hr to 100 km/hr. The curve for the Central shows an unusual behaviour for the low vulnerability class, where the damage is seen reducing for higher windspeed, which does not appear reasonable. Annex 5 shows all the data points that were used to fit this particular curve, which indicates that there is not enough information at higher windspeed. In the case of Mindanao, theoretically, it is expected to have a higher vulnerability, but the curve shows steepness only after the windspeed crosses 100 km/hr. It is, however, important to note that this region gets very few typhoons; hence there were only a limited number of observation available for fitting the curve.

The behaviour of the curves was also different when they were aggregated based on a different number of events, as shown in Annex 5. Using a single historical storm, the fit of the curve is poor, especially in regions that are far from its track and have no damage record. The implication of using these curves on the prediction accuracy, is discussed later in section 4.3.3.

4.2.4 Vulnerability Index

After obtaining the damage curve, the final step for the selection of municipalities for intervention was multiplying this impact with the vulnerability index for each municipality. The vulnerability index obtained based on the composite overlay method will now be discussed in this section.

Figure 22 shows the overall vulnerability index at the municipal level, while Annex 5 shows the index obtained for each indicator and its detailed description. All the values were normalized before comparison, and ranges from 0 (low) to 1 (high).

The overall vulnerability indicates higher values in the southern region of the country, i.e., Mindanao, and in the coastal areas of the Central Philippines. The areas of Northern and Metro Luzon have a comparatively lower vulnerability in terms of the four indicators the calculation is based on. These results show a very close agreement with the overall vulnerability of these areas, as previously identified through the literature and discussed in section 2.2.3.

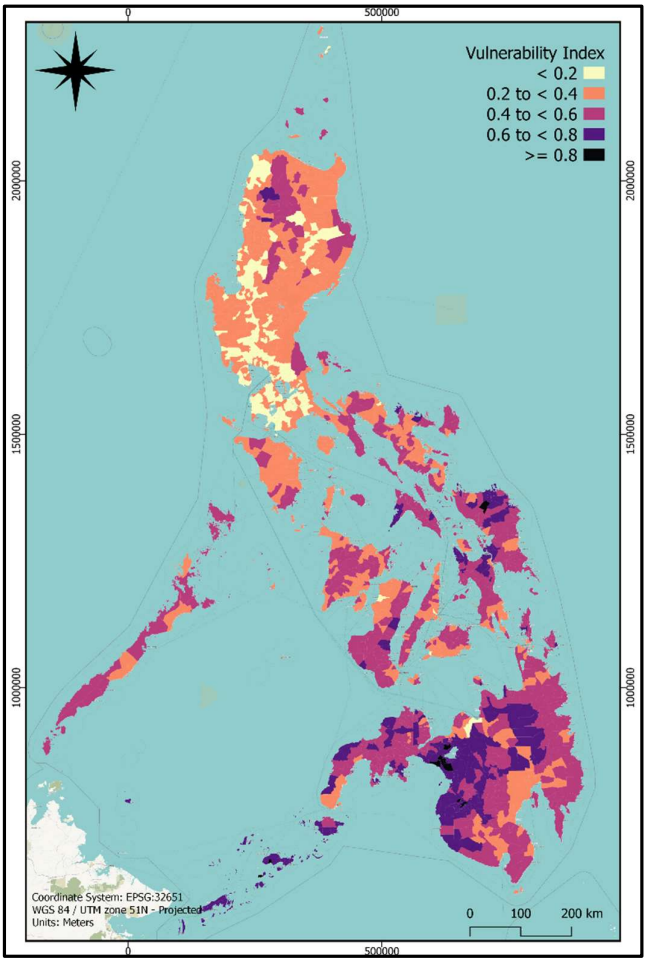


Figure 22 Overall vulnerability index at municipality level of the Philippines, obtained through a weighted overlay of four indicators

4.3 Model Evaluation

Once the two models were set up, the evaluation was done through an interactive dashboard. Figure 23 shows a screen capture of the decision support tool designed as a part of this research, which can also be publicly accessed through the web[8]. Several variables have been assigned a default value, but the users have the flexibility to make selections and dynamically view the changes. The details of each section on the dashboard and its functionalities can be found in Annex 6. The two maps visualize predictions against the actual damage by

8 https://bit.ly/IBFScenarioPortal_Sedhain Note: The platform may show an error at the start and takes around 10 seconds to load

typhoon Kammuri, for an effective post hoc evaluation. Predictions from both elementary and statistical modelling methods are available along with the corresponding accuracy metrics. However, as previously mentioned, most of the dynamic variables currently do not support the predictions made through statistical modelling, as they were only fed as static input into the system.

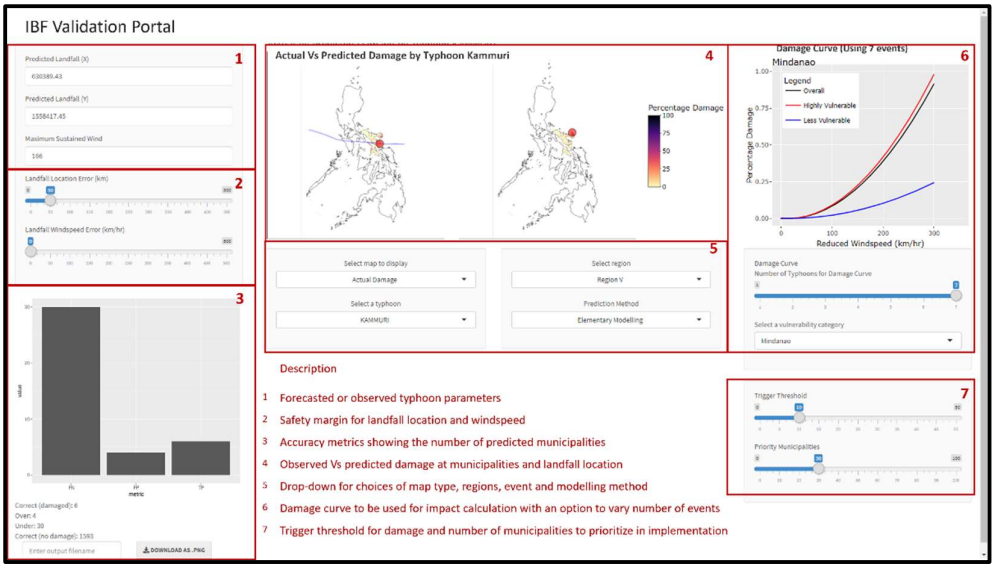


Figure 23 A screenshot of the decision support tool designed for effective comparison of trade-offs in impact forecasting tools

The tool allows the user to input the details of forecasted or actual storm parameters to obtain impact predicted at the municipality level, which can be tested for various lead times. As shown in Figure 24, the two maps can interact in parallel, to get more details, such as the predicted damage percentage and windspeed calculated after the reduction factor curve for each municipality. This, along with the damage curves allows the user to easily interpret how impact to certain municipalities is assigned. Similarly, it also includes vulnerability maps and accuracy metrics, all of which can be downloaded for documentation purposes. The option of region selection not only reduces the processing time, but also allows an area-specific comparison. The section below summarizes key results of accuracy metrics tested in various scenarios for both models.

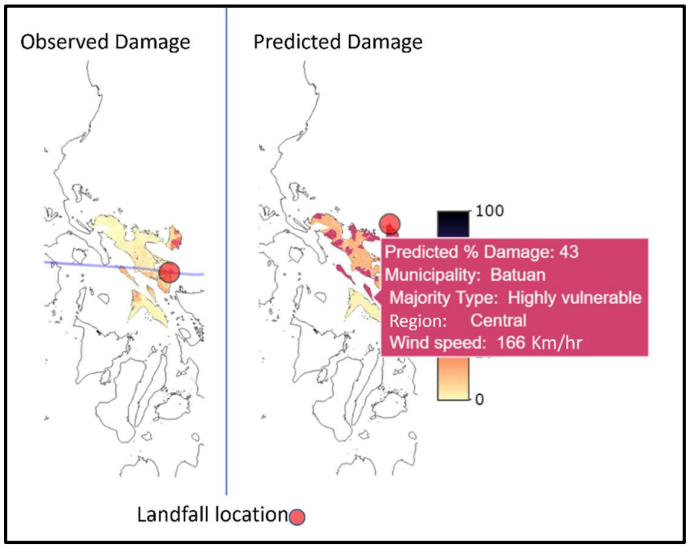


Figure 24 A section of the dashboard where the actual damage (left) and predicted damage (right) can be compared and interacted with in parallel

4.3.1 Modelling technique

In this section, the results of the impact prediction for Kammuri will be discussed, using the baseline scenario for both trigger models:

Elementary model: Trigger activated at 165km/hr with 10% damage threshold and margin of error at landfall location set as 50 km.

Statistical model: Trigger activated when at least 3 municipalities exceed 10% damage threshold

Table 9 Accuracy metrics of two impact models based on the number of municipalities predicted to have exceeded the damage threshold of 10 %, using the forecast at the lead time when the trigger would have been reached

	Statistical Modelling (81 hours before landfall)	Elementary Modelling (72 hours before landfall)
TP	1	6
FP	3	4
FN	34	30
Precision	0.25	0.60
Recall	0.03	0.17
F1 Score	0.05	0.26

The result in Table 9 shows that fewer losses would have been predicted correctly using a trigger model based on statistical technique. This model also appears more conservative, indicating fewer municipalities with damage, compared to the elementary model. However, both model performances have a low recall score, suggesting a poor ability to identify all the impacted municipalities. The elementary model has a higher precision score, hence considering all predicted values it has a better chance of being correct.

4.3.2 Threshold for selection

The accuracy metrics differ when tested against two different damage thresholds. In the case of elementary modelling, a suitable threshold choice is also made to prioritise the municipality for intervention based on the vulnerability index. As can be seen in Table 10, when 5 municipalities are selected for intervention from those predicted, the elementary model's ability to correctly identify gets reduced. However, when the damage threshold is increased from 10 to 25 %, then the overall performance improves, showing that the model is good at predicting the higher damages classes. In the case of the statistical model, it performed poorly and predicted none to have damage of 25%, when in reality there were 7 municipalities affected, as shown in Table 11.

Table 10 Number of municipalities predicted by the elementary impact model to have exceeded two different damage thresholds, based on the forecast at 72 hours lead time before landfall

	Damage threshold – 10%		Damage threshold – 25%
	Total prediction	Prioritizing 5 municipality	Total prediction
TP	6	3	2
FP	4	2	3
FN	30	33	5
Precision	0.60	0.60	0.40
Recall	0.17	0.08	0.29
F1 Score	0.26	0.15	0.33

Table 11 Number of municipalities predicted by the statistical impact model to have exceeded two different thresholds based on the forecast at 81 hours lead time before the landfall

Statistical Modelling		
Damage threshold	**Using 10% threshold**	**Using 25% threshold**
TP	1	0
FP	3	0
FN	34	7
Precision	25%	0
Recall	3%	0
F1 Score	5%	0

4.3.3 Lead time

Based on Table 7, showing the predictions made for the event at various time steps, it was found that in the baseline scenarios, using the elementary model, the trigger would have been reached at 72 hours. Similarly, for the statistical model Table 11 shows that 4 municipalities are predicted to exceed the damage threshold, indicating the activation trigger. Both methods were further tested for the same damage threshold of 10% along the lead time, and the results are summarized in the Table 12 and Table 13 below.

Table 12 Accuracy metrics using statistical impact modelling based on the number of municipalities exceeding the trigger threshold of 10% damage

Elementary Modelling			
Lead time	**0 hours (Observed)**	**24 hours (Forecast)**	**72 hours (Forecast)**
TP	7	1	6
FP	20	44	4
FN	29	35	30
Precision	0.26	0.02	0.60
Recall	0.19	0.03	0.17
F1 Score	0.22	0.02	0.26

Table 13 Accuracy metrics using elementary impact modelling based on the number of municipalities exceeding the trigger threshold of 10% damage

Statistical Modelling				
Lead time	**45 hours**	**51 hours**	**60 hours**	**81 hours**
TP	0	0	3	1
FP	3	8	9	3
FN	35	35	32	34
Precision	0	0	0.25	0.25
Recall	0	0	0.09	0.03
F1 Score	0	0	0.13	0.05

The performance of the elementary model is the lowest at 24 hours lead time where it identifies 1 more municipality correctly, but also predicts many false negatives, so the overall performance reduces. The prediction ability of the statistical model is seen to be highest at 60 hours before the expected landfall, but still with a relatively lower precision score. Figure 25 illustrates the actual and predicted impacts across lead time by the two models.

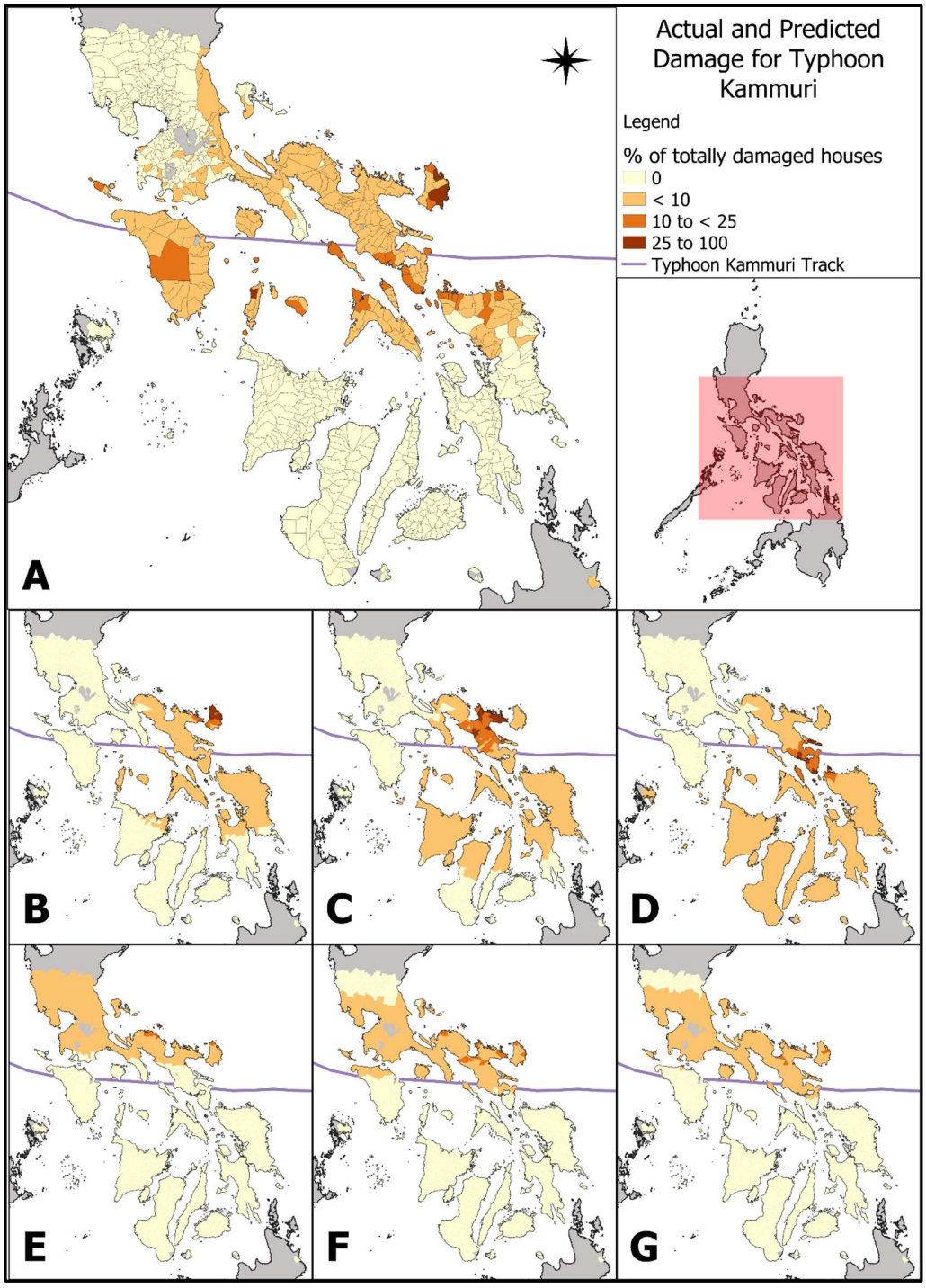


Figure 25 Damages due to typhoon Kammuri (actual and predicted values by the two models)

Map A	The actual damage of typhoon Kammuri as reported by NDRRMC		
Map B	Prediction by elementary model using forecast at 72 hours	**Map E**	Prediction by statistical model using forecast at 81 hours
Map C	Prediction by elementary model using forecast at 24 hours	**Map F**	Prediction by statistical model using forecast at 60 hours
Map D	Prediction by elementary model using observed weather	**Map G**	Prediction by statistical model using forecast at 45 hours

4.3.4 Quality of data

The effect of forecast data quality was tested in terms of the error at the landfall location and track as a safety margin incorporated in the elementary model. As seen in Figure 26, the allowance of an error of 50 km increases the overall performance by 10%, after which it remains constant. But it should also be noted that having a higher safety margin has reduced the precision score substantially, as the model is predicting many more municipalities having an impact. An optimum in this case would be to allow a 50 km buffer to maximize all three-accuracy metrics. In the case of error margin for landfall windspeed, the increase in overall accuracy is not as prominent, and a 20 km/hr addition in forecasted windspeed gives the best in terms of F1 score.

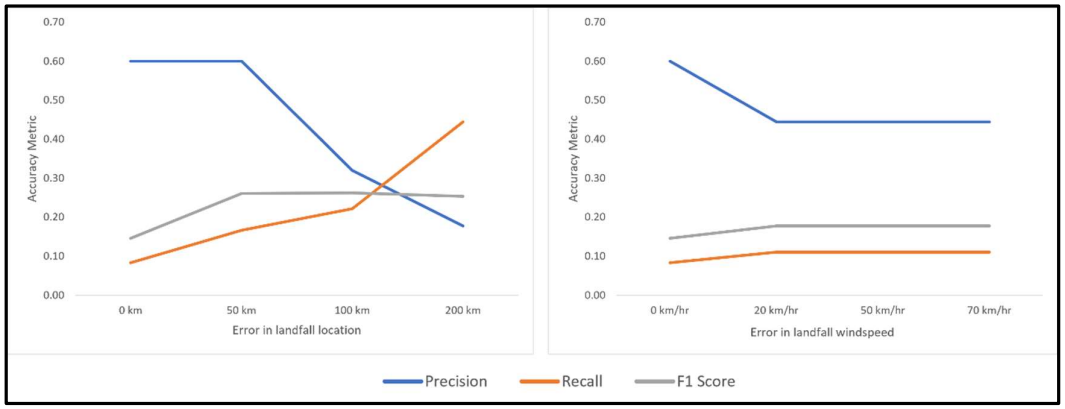


Figure 26 Change in accuracy metrics for elementary impact model, for allowance of different error margin in predicted landfall location (left), and landfall windspeed (right)

Along with that, impact accuracies based on damage data for different events were also looked at in the following order of the total damage: Goni, Phanfone, Haima, Vamco, Yutu, Sarika, Haiyan (Annex: 5). As can be seen in Table 14, the effect of increasing the quantity of damage history data did not have a substantial impact on the accuracy metrics. However, excluding Haiyan led to a slight improvement in model performance, both in terms of precision and F1 score.

Table 14 Accuracy metrics using elementary impact modelling, predicted using a different number of historical events, ranked according to the highest impact

Elementary Modelling				
Number of Event	**7 events**	**6 events**	**4 Events**	**1 Event**
Name of the events	**Goni, Phanfone, Haima, Vamco, Yutu, Sarika, Haiyan**	**Goni, Phanfone, Haima, Vamco, Yutu, Sarika**	**Goni, Phanfone, Haima, Vamco**	**Goni**
TP	6	6	6	6
FP	4	3	3	3
FN	30	30	30	30
Precision	0.60	0.67	0.67	0.67
Recall	0.17	0.17	0.17	0.17
F1 Score	0.26	0.27	0.27	0.27

All the results presented here, will now be discussed, and interpreted in the following chapter.

5 DISCUSSIONS

As highlighted in Chapter 1, there are multiple choices and uncertainties while designing an IbF model, which need to be quantified and communicated well for effective scaling up of early action processes. This research attempted to answer these questions by comparing two different approaches to the trigger model, and by developing a prototype interactive decision support tool for its evaluation. In this chapter, the results for each identified research objective and question will be critically reflected upon. It will also signify the implication of the results in an operational context.

5.1 Objectives and Questions

The main objective of the research was to evaluate how accurately two trigger models predict damage, and which parameters affect the accuracy. The objectives were achieved by first adapting an elementary trigger model of Bangladesh in the context of the Philippines for typhoon early action. The variables relating to hazard, vulnerability and exposure were generated in the process. This adapted model and the currently operational statistical IbF model were then tested on a historical event, typhoon Kammuri, for their impact predictions. Finally, the two models were evaluated for their performance across changes in different parameters and visualized through an interactive decision support tool. The research questions associated with these objectives are individually discussed below.

5.1.1 Model parameters

Research objective 1: To compare the input parameters of the two IbF trigger models

R.Q. 1.1: What is the difference in the way hazard forecast and hazard history are integrated in the two trigger models and converted to the required spatial and temporal scale?

Trigger threshold

For the Philippines, the trigger threshold using the elementary technique was identified to be 165 km/hr for activation, and pre-activation would be when PAGASA raises signal no.2. In terms of the lead time, the pre-activation is expected to reach at least 72 hours before the expected landfall, while the activation based on the windspeed threshold needs to have a minimum lead time of 30 hours. As opposed to the current T-EAP, this trigger model allows a two-stage activation, giving time to prepare necessary support in the early action. The 30 hours lead time also increases the likelihood of getting an accurate forecast closer to the event. It remains an open question whether this lead time is adequate for performing the necessary actions. Bangladesh is a good example of implementing early actions in a short lead time, but which is mostly community centered. In contrast, for activities such as shelter strengthening, 30 hours of lead time before landfall carries the risk of very high wind, making the implementation difficult. Hence the applicability of such triggers essentially depends on the choices of actions.

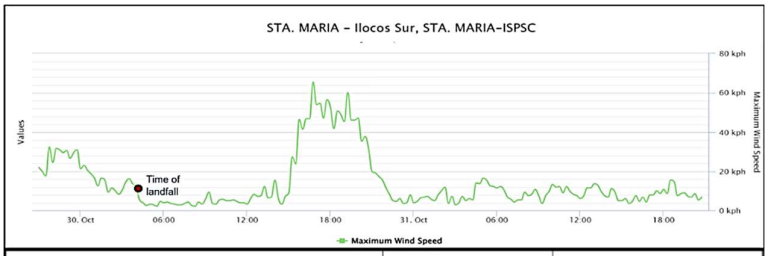
In the elementary model, the trigger threshold is based on a windspeed value instead of damage. This means it only utilizes a single forecast value to generate implementation map and trigger actions. This also raises a concern of whether one windspeed justifies the trigger level for all regions with different vulnerabilities. For example, if the typhoon is making landfall in Mindanao or the Central Philippines, then the wind speed of 165 km/hr will likely have already resulted in higher impact than in Metro Luzon. Conversely, in the statistical model, the predictions are made every 6 hours and the triggers are based on impact. This gives space for performance improvement over the lead time during the same event with more accurate forecasts. In future, the values of both approaches must be weighed against each other, while making the choices of trigger.

Windspeed estimation

The elementary modelling method estimates the windspeed for each municipality based on an equation which was generated for the Philippines during this study. The relation was formulated using two parameters: fraction of windspeed observed and the distance of municipality from landfall. In comparison, this curve was found to be notably different from the one operational for C-EAP. What also needs to be looked at is that observations made for Bangladesh were only based on data from 20 AWS for a single event, most of which were within 100 km from landfall. In contrast, for this research, more than 40 weather stations were included that covered unique spectra of distances and event types. To increase certainty in the estimation, more observations covering several events can be tested for Bangladesh in the future revisions of the C-EAP.

The reduction curve for the Philippines showed a strong initial decay and had more variations in data points over the distance. This can be indicative of differences in elevation and roughness over the region, as compared to Bangladesh which is mainly flat. The strength of a tropical storm in terms of its central pressure is very dependent on the influx of warm air above the sea and cool and drier air above land, as previously discussed in section 2.1.1. Many studies have looked into the interaction between the waves induced by a storm (DeMaria et al., 2006; Kaplan and Demaria, 1995), highlighting the importance of topography to be factored in, while generalizing the behaviour of a storm in a specific location. As Bangladesh is a landmass and the Philippines is a cluster of islands, the way a storm progresses across the two regions will differ. Even within the Philippines, storms over larger islands such as North Luzon could act different compared to smaller islands, such as in the central region. Therefore, such simplification in storm behavior based on limited parameters can be a source of inaccuracies. This further implies that adaptation of trigger models in a new context needs to factor in physical differences such as topography, along with all the risk parameters.

Likewise, there is also a concern of using distance from landfall instead of the actual track for constructing this curve. By definition, a landfall occurs when the centre of the storm reaches the land (NHC, n.d.). With an average radius of 100 to 250 km from the centre to the outermost edge, the storm has already reached some areas hours before the landfall happens or will pass some regions hours later. A good example is typhoon Kammuri that had four landfalls over 13 hours, in which the storm had travelled more than 500 kms, maintaining a windspeed of 165 km/hr (which is the trigger in this scenario). Figure 27 shows an example four AWS looked at in relation to typhoon Yutu (2018), which passed over a relatively large landmass. A clear interpretation can be made that a station is likely to experience maximum windspeed at a time different from the landfall when the storm is closer to that area. This underlines the fact that a storm's track is an essential parameter in wind estimation, and needs to be taken into account for further improvement of the operational elementary impact model.



Name of the station	Maximum windspeed recorded around landfall time (km/hr)	Maximum windspeed recorded in the day of the storm (km/hr)
QUIRINO STATE COLLEGE, DIFFUN	10.7	23.3
STA. MARIA - Ilocos Sur, STA. MARIA-ISPSC	5.4	65.52
BRGY. SANTIAGO, NAUJAN, oriental mindoro	0.5	32.76
SAMOKI, BONTOC, CAR	2.5	15.6

Figure 27 Windspeed recorded by 4 different AWS during typhoon Yutu, around its landfall time versus the maximum recorded within 12 hours from landfall

A good example can be taken from the statistical trigger model, which uses a parametric windfield model based on Holland (2008), for the estimation of various wind profiles at the municipality level. Even though based on a similar concept of wind decay over the distance, this modelling method is able to capture more characteristics of the storm through its intensity, track and central location. Additionally, interpolation techniques such as the parametric modelling are highly beneficial when ground-based stations are not available. A similar approach can be tested for the elementary model, even with limited hazard history data. In a recent study, Fakhruddin et al., (2022) assess ways to create synthetic cyclone tracks for 1000 years to generate wind attributes, using which more reliable results can be generated. At the same time, the accuracy of the parametric windfield model can also be validated against weather station data, to evaluate its performance. Both models at the moment do not take into consideration the terrain effect on the wind fields. Hence, these insights can be integrated for continuous improvement in the hazard components of impact forecasting models.

R.Q. 1.2: What is the difference in how the models incorporate physical and social vulnerability?

To establish the relation of physical vulnerability for elementary modelling, multiple damage curves were obtained based on windspeed reduction and impact history. These curves, used for impact prediction, were adapted to four geographical areas and two construction mechanisms to make them more localized. It was observed that the curve's fit shifted when the number of event variables used for its construction was reduced. Besides, the curves also deviated from the normal expected trend of a damage function for cases with less data such as Mindanao. In that reference, several limitations were identified in using this approach for impact predictions. Firstly, total damage in a municipality is correlated only to the windspeed based on the time and distance from landfall even though a stronger wind can be experienced when the storm travels closer to it. At the same time, damage during a tropical storm potentially also results from a storm surge or other related events, such as during Haiyan. Moreover, the possible inaccuracy in the reduction factor estimation due to the reasons pointed out while answering RQ1.1 will also propagate to the damage curve. Finally, as pointed out in a study by Andres Diaz Loaiza et al., (2021), topographic data play the most important role when generating damage curves for coastal area, which is not well accounted in this method.

The social vulnerability and coping capacity in elementary model were incorporated through an index generated at each municipal level. In this trigger model, the consideration of vulnerability index for each municipality allows their prioritization for intervention. Hence, in case of limited institutional capacity for early action, areas with higher vulnerability, such as in Mindanao or Central Philippines, will be given first priority, also considering the magnitude of impact predicted for those areas. For this research, the indicators related to poverty, dependent population, evacuation centre and house type were used for ranking the vulnerability. However, there is a flexibility in selection of indicators and weights, adapting it to fit the implementation purpose. For example, an agriculture-based indicator can be relevant when early action includes crop harvesting. However, since these weights are user-defined, it also introduces a great amount of bias, as their determination is not straightforward and open to debate. The subjectivity of these choices made during the modelling process should be considered and communicated well with all the relevant stakeholders.

The statistical method uses classifier algorithms to predict the impact based on a total of 36 variables that include observed weather, topographic features, social and physical vulnerability and damage records from 37 events. Therefore, compared to the elementary model, several variables draw the relation of impact to physical and social vulnerabilities. This method has a better ability to incorporate the complexities of the event since storm surges, landslide and rainfall records are also included. Consequently, this also means that the method is quite data demanding and requires a higher level of technical expertise. In areas with limited capacity in those terms, the implementation of a statistical modelling approach can be challenging.

As seen for both methods, the damage is generalized at municipality scale, which means the values in relation to construction mechanism or local topography will also be averaged. Moving forward, even the damage curves or model inputs can be localized at intervention scale, such as the municipality. This would mean a finer resolution of damage data would be required for individual household types. This calls for advocacy in local

government for sharing the household level damage information, at least in regions that have a well implemented data collection and sharing process such as the Philippines. For data scarce regions, the use of satellite based damage assessment can be useful, which have shown good accuracy in assessing multiple dimensions of impact (Hoque et al., 2016). There are still gaps for it to be utilized in an operational level and this can be reduced only if the procedure includes the development of standards with the involvement of all responsible institutions (Kerle, 2010). This can also stand true with local level impact information, which comes from various sources and has differences in definition of impact level.

5.2.2 Model performance

Research Objective 2: To validate the performance of the two trigger models

R.Q. 2.1: What is the impact of lead time on the performance of the two IbF models in predicting household damage?

From the prediction results at different lead times, it was found that even though both models reached the triggers 3 days before the event, their performances were considerably low. The two models, statistical and elementary, had the best prediction accuracy at 60 and 72 hours before landfall, respectively. In contrast to what is generally expected, both models did not show an improvement in performance with forecasts closer to the event. This can be attributed to the fact that Kammuri intensified in the last hours before landfall and deviated from its forecasted track (Figure 18). This means that the conclusion on overall performance of the model over lead time cannot be generalized based on a single event. But at the same time, tropical storms are complex phenomena, and these tools need to be robust enough to capture the unforeseen circumstances, which does not stand true for both the models in this study. While ideally accurate predictions for RI storms are desired, and there are several ongoing studies taking place (Miyamoto and Takemi, 2015; Wang and Zhou, 2007), we are not quite there yet. And it is also true that its frequency can only be expected to increase in the near future (Masson-Delmotte et al., 2021). In that case, adapting our strategies and modelling approaches to possible deflection is deemed necessary. Operationally this could be achieved in multiple way, by either increasing the implementation area or having a flexible lead time. Additionally, the models must also be tested on realistic scenarios such as in this research, but with a broader number of events with unique physical characteristics.

R.Q. 2.2: What is the impact of choices made in trigger thresholds and municipality prioritization, on the performance of damage prediction?

The choices of thresholds used had notable effects on the overall accuracy of both models. However, the elementary model showed a better ability to predict higher damage threshold, which was not true in the case of the statistical model. The lower prediction accuracy can potentially be linked with the class imbalance in the model inputs. Out of 9103 past damage records, only 472 municipalities have damage of more than 10%, and 229 exceed damage of 25%. In a real world applications, datasets will typically be imbalanced and bias correction techniques such as the one described by Wagenaar et al., (2021) can be used for eliminating this issue.

The performance of the elementary model reduced when only 5 municipalities were prioritized out of all predicted. The threshold for selection here becomes very relevant for the practicalities of intervention. In the case of Kammuri, a total of 36 municipalities had exceeded the damage threshold. So, even if our models had a 100 % accuracy, there may not be a capacity for intervention in these many potentially affected areas. Given that, assigning the vulnerability index values accurately also becomes very crucial in elementary modelling methods.

Currently most of the EAPs, including the two studied in this research, use a single trigger value. As highlighted in GRC, (2022), agencies are exploring ways contextualize the triggers for various seasons and socioeconomic categories, and this has to be done based on cumulative of multiple events instead of one extreme event. In some instances, for the development of EAP, the local community also helps in identifying the trigger levels based on their experience. For that, an evaluation of different trigger models done in this research can be expanded to find customized trigger values for each scenario.

R.Q. 2.3: What is the sensitivity of the two trigger models to changes in one of the predictors (in terms of hazard and vulnerability)?

In terms of forecast accuracy, an increase in the margin of error in forecasted landfall location increased the overall accuracy, but also lowered the precision score in the elementary modelling. However, the sensitivity of accuracy to the landfall location was much higher than to the intensity of windspeed. This implies that for accurately modelling the impact in this approach, it is more important to get the location of landfall correct, compared to its windspeed. And this is also true considering the state-of-art in forecast skill which over the year have improved more in modelling storm track than its intensity (Heming et al., 2019). But at the same time, the results suggest that if this model were to be operationalized in the Philippines, it tends to over-estimate for municipalities close to the landfall. From a visual interpretation of the prediction maps in Figure 25, it can be deduced that the predictions of elementary model are constricted around the landfall location while the statistical model predicts well across the entire track showing similarities with observed damage. Hence, the distance from landfall, which is the most important parameter in elementary model, is not providing the most correct estimation. In a real storm event, there could be municipalities that do not necessarily feel the first landfall and still experience the typhoon across its track. The elementary modelling approach can lead to lesser preparedness in these areas.

The effect of the availability of historical damage data on the elementary model was tested through multiple damage curves constructed based on different number of events. It was noted that the prediction accuracy did not improve on increasing the number of historical events. Results suggest that it is not only the quantity of historical data, but also the nature of events in use that affects the model performance. For example, removing typhoon Haiyan from the damage curve showed an improvement in the model performance, which is an exceptional event with a large proportion of damage attributed to storm surge (Lagmay et al., 2015). Hence, a linear relation of windspeed to damage is not applicable in that context. Moreover, since the central region had considerably larger amount of damage data (Annex: 4) and typhoon Kammuri also affected the same region, in this case study, the accuracy was not affected even when using single event damage data. Consequently, the resulting damage curves show that if prediction was made in regions with limited damage data, such as Mindanao, there is a higher uncertainty in the modelled results. Going forward, the suitability of these methods for data scarce areas within a region must be tested as well. It also means that, when employing historical dataset in modelling methods, there needs to be a good understanding of specific characteristics of the event.

For the statistical model, the sensitivity of the data quality in terms of timeliness of demographic variables was tested. The performance of the model did not show any change when 10 variables (total number of houses with roof and wall type) were improved with the statistics closer to the event date. There could be multiple possible explanations for this behaviour of results. The variables correlating to hazard information, such as wind speed and typhoon track, potentially have more weight in assisting the model in making an impact estimate. Similarly, it is also important to consider the scale of intervention which may also play a role here, and if the IbF was done at household level instead of municipality, the change in exposure would have more impact. As already shown in Figure 19, there is an increase in absolute value of population over the years, but relatively the four regions show a similar trend in the growth. Even so, the growth rate of population, is evidence of changing exposure, which must be factored into the impact models while defining triggers based on percentage instead of absolute values. There are already ongoing pilots on using absolute damage data and associated probabilities for triggering actions for typhoons in the same study area (OCHA, 2021b). In that context, a similar statistical impact model

is deployed by 510, which uses a regression instead of the classification algorithm tested in this research, to identify the highest impacted municipality based on probabilistic forecasts.

The damage history data coming from official sources showed multiple issues relating to duplication and inconsistency, along with discrepancies when compared to the existing model input. Even though the difference found may appear negligible, it is still 18% of the total damage data. It should also be noted that the corrections were made only for the single event, Kammuri, and did not affect the model performance. But a similar issue is imminent if recalculations are done for all historical typhoons used to train the model. This further emphasizes the importance of unique identifiers such as P-codes to be adapted even at the local level, which is the source of many datasets we use in our models. Not only for accuracy purposes, but using standardized codes instead of names, can help overcome the challenges of processing geographic information.

The test of data quality for the model in this research was limited to the timeliness of the demographic variables (population and household). Other dynamic variables, for example the assumptions made in averaged values of physical vulnerabilities over the year, could also have an effect. For instance, Typhoon Goni which was followed by Typhoon Vamco affected common areas in two provinces of the Philippines within 10 days (OCHA, 2020b). This means that houses already damaged by Goni and people in temporary shelters are not taken into consideration by the model. This opens the floor for further exploration of changes in exposure or physical and social vulnerabilities, and how prominent their effect is on the model performance.

R.Q. 2.4: What would be the most adequate combination of threshold and lead time for each model to trigger early action during the event?

From the results, it was seen that, using both trigger models, the threshold of FbA activation was reached at 72 hours before the predicted landfall. To carry out shelter strengthening, early harvesting of crops and livestock evacuation, the time frame of 72 hours have already been identified to be sufficient, and is the minimum time required for the operational EAP in the Philippines. If we consider the early action C-EAP adapted in this research, the activities mainly carried out are evacuation and distribution of food and water. This means that for both scenarios 72 hours would be more than enough with a lot of safety margin to carry out the aforementioned early actions.

When considering FbA, humanitarian agencies have to make a choice in two aspects, as termed by Lopez et al., (2020); (1) prevented event maximization, where the cost of preparedness is expected to justify what would be spent in response, hence there is no constraints in budget, and (2) expense minimization, where the goal is to minimize the impact but also considering the cost effectiveness. The selection of the optimal thresholds and impact model for this research can be looked at in the following two contexts in reference to typhoon Kammuri:

Prevented event maximization:

A recall score is more relevant when the goal is to maximize impact reduction without missing any municipality that could possibly exceed a threshold. Thus, in hindsight for typhoon Kammuri, the elementary model would have been the better choice for triggering action in terms of maximizing impact reduction, as it attained a higher recall score. Along with that, the prioritization done through vulnerability ranking is also relevant in terms of designing early actions, in such a way that even if the event does not materialize, those who need it the most benefit from it. Meanwhile, a damage threshold of 10 % with an error margin of 200 km in landfall location and 20 km/hr in windspeed, gives the highest number of correctly identified municipalities. The study was not able to re-run the statistical model, for which the predictions were limited to municipalities within 100km of the forecasted track. Based on visual interpretation, it can be deduced that an increase in this distance, would have reduced the effect of track change on the model results. The determination of most optimal distance can be a relevant test for future works.

Expense Minimization:

From a financial perspective, an F1 score is an ideal metric of evaluation to have a good balance in correctly identifying damage (recall), but also not overpredicting (precision). In that scenario, for the elementary model a 25 % damage threshold is seen as suitable. In terms of error margin, 50 km in landfall location and 20 km/hr in windspeed are recommended for optimal accuracy. Moreover, increasing the damage threshold also showed to have limited the number of municipalities for intervention, without compromising on the model performance in the case of the elementary model. However, the theoretical reliability of the statistical model is higher than the elementary model, since the latter makes less validated assumptions like in windspeed estimation. This is reinforced by the elementary model's inadequate performance at 24 hours (F1 score = 0.02), showing that it performs inconsistently. Conversely, the statistical model that uses machine learning, holistically considers more characteristics of the typhoon, and the model validation results allow a more objective measurement for financial risk.

5.3.3 Model explainability

Research Objective 3: To compare the explainability of the models in terms of their transparency and post-hoc interpretation

R.Q. 3.1: How can the results of forecasted damage by the two models be interpreted in terms of their parameters and prediction algorithm used?

As pointed out early on by Breiman, (2001), to make better predictions, generally, a complex model is required against interpretable functions. The results, however, suggest this might not be entirely true. The elementary modelling had a linear assumption concerning impact, based on calculated windspeed with minimal variables involved. The accuracy results can easily be traced back to the damage curve or reduction factor curve to identify the possible source of over or underestimation. In fact, the entire model can be summarized within a single equation combining the damage curve and windspeed reduction value. However, the interpretability is relatively lower with several variables involved in the statistical modelling. Especially since this is not a rule-based algorithm, they do not operate under a simple 'cause and effect approach'. There are, however, multiple proven methods such as the partial dependence plots (Friedman, 2001) and individual conditional expectations (Goldstein et al., 2015), to better visualize the role of each variables in such machine learning algorithms, moving us closer in decoding the black box. But at the same time, it should be taken into consideration that these plots should be interpretable also by audiences other than domain experts.

Among the two models, the elementary technique was observed to be easily reproducible, considering a smaller number of forecasted and observed dataset. This also meant post-hoc evaluation, in this case, was more straightforward and required only few storm parameters from the historical forecast. However, in the case of a statistical model, the formulation of the model but also its post-event validation requires an extension preparation of the dataset since the forecast information are not officially stored after 180 days.

Transparency can also be looked at in terms of access of codes and dataset of the models to allow the reviewing process (European Commission, 2021). In that context, the process of the elementary model, even though well described, is limited to a written document (BDRCS et al., 2021), and the access to datasets is not provided. This raises concern of ambiguity in the modelling process while adapting this model. On the other hand, the statistical model along with the written description (PRC et al., 2020) is also well documented in the GitHub repository[9] along with all the datasets, making it accessible for users to validate.

[9] https://github.com/rodekruis

R.Q. 3.2: What is a suitable method to evaluate the choices and uncertainties in the models through an interactive platform?

Interpretation of a model is often limited to the developers or scientific experts, which in fact should also be for the intended users. This adds more reasons to explore ways to not only evaluate the trigger models, but also make it accessible. An interactive dashboard was created to establish transparency in the modelling approaches, and intends to serve the purpose of communicating two things in an IbF model: (1) The uncertainties in different parameters, which may not be within our control. For example, the skill of the forecast provider or the credibility of damage data used for model construction. Meanwhile, (2) there are certain choices taken by the decision-makers, such as at what time to trigger action, safety margins for forecast error or indicators for priority ranking. Both of these elements are incorporated into the decision support tool.

One potential use of the presented tool is in the development and training process of EAP. When trigger models are designed, a detailed skill testing is done (IFRC, 2020a), where historical forecast is compared to historical observation to assess how often a trigger would reach and chances of 'acting in vain'. This process involves funding agencies, meteorological offices, government stakeholders, National Red Cross Society members, in short, individuals from different backgrounds and interests. The effectiveness of the process can be increased through a dynamic interactive platform that is standardized across all projects. Furthermore, a recent study also highlighted how user centered designs must be strongly advocated in communicating risk (Twomlow et al., 2022), and identifying the gaps would be the logical first step. The study by van den Homberg et al., (2020) found that when implementation maps for Kammuri were provided 3 days before the event, the end-users failed to comprehend the role of forecast uncertainty in the impact results. In those events, the forecast roles can be highlighted better with a similar interactive tool that can be used during the training phase by the service providers. It can also potentially incorporate scenario what-if analysis including future projections and climate change scenarios and its effect on the model performance.

The concept behind this platform can also be employed for an effective post event evaluation. There are many portals existing across various countries for real-time information dissemination during crisis, with forecasted parameters[10] [11] [12]. But these tools do not currently incorporate features to compare the predictions against the observed event and how it changed over time. The interactive platform developed in this study fits very well with the principles of the evaluation framework for predictive analysis in humanitarian decision making (OCHA, 2021c). It can be an effective means in technical review for benchmarking the tools and testing their robustness and usability. But the process of validation, as done in this research, might not be applicable to all hazards. For example, in heat waves where the early action ideally reduces the impact that was predicted, a straightforward validation technique cannot be employed, and further studies are required for generation of common evaluation metrics.

As highlighted above, the purpose of model interpretations may vary to either improve the models or to justify the results. These use-cases must be taken into consideration while further designing similar tools along with measuring its effectiveness on achieving the purpose.

[10] https://www.510.global/impact-based-forecasting-system/
[11] https://bipadportal.gov.np/
[12] https://startnetwork.org/dynamic-risk-monitors

6 CONCLUSION AND RECOMMENDATIONS

Section 6.1 lists the limitations of this research and provides recommendations for further work. Finally, section 6.2 provides a concluding remark from the author's perspective.

6.1 Limitations

As already highlighted in previous chapters, several assumptions were made during this study that was done with reasonable justification and considering the objectives. Along with that, several improvements and additional tests could be done in future work. The section below lists down some key limitations along with recommendations to improve it, where relevant.

Datasets and modelling:

- One of the biggest limitations in this research was that the evaluation of the two models was done using a single event. Based on a single case, one can only deduce how actionable the impact-based forecasts are in that scenario, and how well they represented changes in hazard forecast. But it is not possible to draw conclusions on the overall suitability/performance of one model over the other. For this, a more elaborate statistical analysis with many more historical events needs to be done.
- Wind speed observations taken from AWS had to be manually extracted, limiting the efficiency and the quantity of the derived dataset. Provision of downloadable data format such as .csv would have allowed a faster retrieval and lowered the chances of inaccuracies. A better estimation of windspeed reduction would have been possible if the landfall windspeed was also taken from the ground station to calculate the reduction factor. Unfortunately, this was not possible due to the unavailability of operational stations very close to landfall at that time.
- The household damage obtained from NDRRMC was used for the calculation of accuracy metrics, with an assumption of being the ground truth. However, since these data are collected at the Barangay level (smallest administrative unit) and then aggregated to the municipality, there might be a possible discrepancy in definitions of 'totally damaged house'. These factors of data quality have not been considered in this research.
- In preparation of risk information (reduction factor curve, damage curve and vulnerability index) for elementary modelling, the considerations were largely based on the C-EAP. The elementary model can also be improved by testing the applicability of parametric modelling, instead of using a reduction factor curve to incorporate typhoon track parameter. Finally, better documentation of this model (example: GitHub) in addition to the written report, can support future researchers and practitioners to adapt and validate it quickly.
- Some experiments, such as the wind simulation testing, were not possible in this research and assumptions were made based on literature, which may have compromised the quality of obtained damage curves. For future research, expert consultation can be taken to assign vulnerability class and weights to have a better reliability in the results.
- Due to limitations in data preparation, the operational statistical model could not be tested for performance evaluation, as the forecast of past typhoons in the ECMWF database is stored only for the past 180 days. The predictions used for accuracy metrics were made during the event in 2019, when the model had a coarser resolution and used a deterministic forecast. In the future, there can be more exploration of ways on increasing the feasibility of such post-hoc evaluation.
- Furthermore, the effect of data quality on the statistical model performance was limited to the timeliness of variables relating to population and household, while testing other dynamic variables is open to further research.
- The parametric windfield model used in the statistical approach, was considered more reliable due to the inclusion of a greater number of wind parameters. However, testing the validity of this is also an

open possibility. There are several weather station datasets available for the Philippines, as utilized in this research for reduction curve generation. These observations can be used to cross-validate the windfields that have been estimated at each municipality level to increase the reliability of this approach.

- Forecast lead time when the predictions were made, differed for the two models since the information was dependent on separate sources. Due to the same reason, a prediction of statistical model using the actual windspeed of Kammuri (0 hours lead time) was also not available. A common frequency of predictions in terms of lead time, would have allowed more objective measurement in their performances.
- The impact history information used for both the models was based on damage data obtained from the national level disaster authority. While preparing these datasets, and in order to link them, the utility of unique identifier p-code was recognized. There were discrepancies identified in the model input which was corrected, but only for the event Kammuri. The inaccuracies shed light on the importance of revisiting these datasets to examine their correctness. At the same time, in organizational level, advocacies for standardization of local level datasets are needed.

Interactive portal:

- Currently, the interactive platform was built to answer the research questions, and is based largely on subjective choices. Moving forward, it must be co-created considering what the service providers are able to and willing to share, and also a requirement analysis of the users. There is a potential of expanding its use with addition of several other variables. For example, the forecast uncertainties in the statistical modelling method require more complex parameters and are not included as a part of this decision portal, which can be explored moving forward. Along with that, a dynamic choice of weights and indicators for the vulnerability index can be incorporated to allow case-specific adjustments and testing. Similarly, the prototype only includes the list of multiple typhoon events and its actual damage map, and the effect of lead time was also tested but not visualized. These can also be useful additions to evaluate the performance across various events and lead time scenarios, to be able to draw better conclusions.
- The statistical model was not explicitly tested on variable importance to correlate the behaviour of predictions. In future work, the inclusion of graphics such as the partial dependence plots can be utilized to understand the role of each variable better and dissect the black box which can also be visualization portal.

6.2 Conclusion

The contribution of this research is twofold: (1) benchmarking of two very different approaches of IbF based on their damage prediction performance on a historical typhoon event in the Philippines, and (2) designing a suitable decision-support tool to evaluate and communicate the choices and uncertainties involved in an impact forecasting model. One model used only a few hazard and impact history parameters (elementary modelling), while the other used a machine learning algorithm trained on a substantially larger number of variables (statistical modelling). The study did not only perform a theoretical assessment of these models, but by testing them on a real case, multiple insights were generated for its operational relevance. The results showed that, in hindsight for typhoon Kammuri, the elementary model would have given slightly better performance than the statistical model. This implies that, for this specific case, complex was not the better choice. But at the same time, the results also varied generally on changing parameters associated with lead time, trigger threshold, forecast uncertainties and input data, and neither model produced satisfactory results. More importantly, the unprecedented change in forecasted information highly affected the model performance. This further highlights the importance of testing model robustness across multiple events and scenarios, deepening the integration of such tests in the IbF. This study also demonstrated how, while developing or validating these trigger models, a thorough understanding of the local context in FbA approaches as well as the characteristics of historical events are needed, and this was achieved through interaction with several experts in the domain.

Another major take away for IbF practitioners is a call towards accessibility in the evaluation process to move towards transparent decision making. With an expansion in data-driven decision-making for humanitarian support, an increasing number of tools is being operationalized. With that comes the challenge of effectively communicating their complexities to the end users and allowing them to conduct scenario "what-if" analyses. The resulting interactive platform of this research was only an attempt to explore its benefit and present a prototype. Going forward, its operationalization can make this evaluation of decision tools more systematic for funding agencies and the local actors, for an evidence-based understanding of IbF. The end goal is not always to find the best performing model, but to find the one that is best suited for the implementation plan, and such evaluation tools can accelerate this process.

As we are dealing with gaps in data collection and quality, gaps also exist in how effectively this information can be incorporated for making informed decisions. Hence, moving forward, we are responsible to make these processes more open source, but also easily interpretable and usable by other practitioners and policy makers.

LIST OF REFERENCES

510, 2020. IBF: TYPHOON TRIGGER WARNING MODEL – 510 GLOBAL [WWW Document]. URL https://www.510.global/ibf-typhoon-trigger-warning-model/ (accessed 10.30.21).

510, 2019. AUTOMATED IMPACT MAP TYPHOON KAMMURI – 510 GLOBAL [WWW Document]. URL https://www.510.global/automated-impact-map-sent-120hrs-before-typhoon-kammuri-arrives/ (accessed 11.6.21).

510 Global, 2022. Community Risk Dashboard - Philippines [WWW Document]. URL https://dashboard.510.global/#!/community_risk (accessed 4.9.22).

ADRC, 2018. THE PHILIPPINE DISASTER RISK REDUCTION AND MANAGEMENT SYSTEM [WWW Document]. URL https://www.adrc.asia/countryreport/PHL/2018/Philippines_CR2018B.pdf (accessed 10.28.21).

Alexander, D., 2018. Natural Disasters. Nat. Disasters 1–632. https://doi.org/10.4324/9781315859149

Alfieri, L., Burek, P., Dutra, E., Krzeminski, B., Muraro, D., Thielen, J., Pappenberger, F., 2013. GloFAS-global ensemble streamflow forecasting and flood early warning. Hydrol. Earth Syst. Sci. 17, 1161–1175. https://doi.org/10.5194/HESS-17-1161-2013

Andres Diaz Loaiza, M., David Bricker, J., Meynadier, R., Duong, T., Ranasinghe, R., Nicolaas Jonkman, S., 2021. Development of damage curves for buildings near La Rochelle during Storm Xynthia based on insurance claims and hydrodynamic simulations. https://doi.org/10.5194/nhess-2021-161

Anticipation Hub, 2020. Anticipation Hub: Forecast-based Financing (FbF) in the Philippines [WWW Document]. URL https://www.anticipation-hub.org/experience/anticipatory-action-in-the-world/philippines/forecast-based-financing-fbf-in-the-philippines/ (accessed 6.19.21).

Anticipation Hub, n.d. Anticipatory action around the world [WWW Document]. URL https://www.anticipation-hub.org/ (accessed 6.15.22).

ASTI-DOST, 2022. Philsensor [WWW Document]. URL https://philsensors.asti.dost.gov.ph/ (accessed 4.13.22).

Bank, W., 2005. Natural Disaster Risk Management in the Philippines : Enhancing Poverty Alleviation Through Disaster Reduction [WWW Document]. URL https://openknowledge.worldbank.org/handle/10986/8748 (accessed 3.27.22).

Basconcillo, J., Moon, I.J., 2021. Recent increase in the occurrences of Christmas typhoons in the Western North Pacific. Sci. Reports 2021 111 11, 1–10. https://doi.org/10.1038/s41598-021-86814-x

BDRCS, GRC, IFRC, Climate Centre, 2021. Forecast-based Financing / Action Early Action Protocol Cyclone Bangladesh.

Bierens, S., Boersma, K., Homberg, M.J.C. van den, 2020. The Legitimacy, Accountability, and Ownership of an Impact-Based Forecasting Model in Disaster Governance. Polit. Gov. 8, 445–455. https://doi.org/10.17645/PAG.V8I4.3161

Bischiniotis, K., de Moel, H., van den Homberg, M., Couasnon, A., Aerts, J., Guimarães Nobre, G., Zsoter, E., van den Hurk, B., 2020. A framework for comparing permanent and forecast-based flood risk-reduction strategies. Sci. Total Environ. 720, 137572. https://doi.org/10.1016/J.SCITOTENV.2020.137572

BMD, 2014. Signals | Bangladesh Meteorological Department [WWW Document]. URL

http://live4.bmd.gov.bd/p/Signals/ (accessed 4.7.22).

Breiman, L., 2001. Statistical modeling: The two cultures. Stat. Sci. https://doi.org/10.1214/ss/1009213726

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 2017. Classification and regression trees. Classif. Regres. Trees 1–358. https://doi.org/10.1201/9781315139470/CLASSIFICATION-REGRESSION-TREES-LEO-BREIMAN-JEROME-FRIEDMAN-RICHARD-OLSHEN-CHARLES-STONE

Build Change, 2014. Post-Disaster Reconnaissance Report Damage Assessment and Housing and Markets Survey 2013 Bohol Earthquake and Typhoon Yolanda [WWW Document]. URL https://buildchange-web.s3.amazonaws.com/resources/pdfs/Build Change_Philippines Reconnaissance Report.pdf (accessed 2.4.22).

Casati, B., Wilson, L.J., Stephenson, D.B., Nurmi, P., Ghelli, A., Pocernich, M., Damrath, U., Ebert, E.E., Brown, B.G., Mason, S., 2008. Forecast verification: current status and future directions. Meteorol. Appl. 15, 3–18. https://doi.org/10.1002/MET.52

Chang, W., Cheng, J., Alllaire, J., Xie, Y., McPherson, J., 2020. shiny: Web Application Framework for R.

Chen, R., Zhang, W., Wang, X., 2020. Machine Learning in Tropical Cyclone Forecast Modeling: A Review. Atmos. 2020, Vol. 11, Page 676 11, 676. https://doi.org/10.3390/ATMOS11070676

Coughlan De Perez, E., Van Den Hurk, B., Van Aalst, M.K., Jongman, B., Klose, T., Suarez, P., 2015. Forecast-based financing: An approach for catalyzing humanitarian action based on extreme weather and climate forecasts. Nat. Hazards Earth Syst. Sci. 15, 895–904. https://doi.org/10.5194/nhess-15-895-2015

De Perez, E.C., Van Den Hurk, B., Van Aalst, M.K., Amuron, I., Bamanya, D., Hauser, T., Jongma, B., Lopez, A., Mason, S., De Suarez, J.M., Pappenberger, F., Rueth, A., Stephens, E., Suarez, P., Wagemaker, J., Zsoter, E., 2016. Action-based flood forecasting for triggering humanitarian action. Hydrol. Earth Syst. Sci. 20, 3549–3560. https://doi.org/10.5194/hess-20-3549-2016

DeMaria, M., Knaff, J.A., Kaplan, J., 2006. On the Decay of Tropical Cyclone Winds Crossing Narrow Landmasses. J. Appl. Meteorol. Climatol. 45, 491–499. https://doi.org/10.1175/JAM2351.1

DOST-ASTI, 2018. Ensuring the Accuracy and Quality of Data from DOST-ASTI's Automated Weather Stations [WWW Document]. URL https://asti.dost.gov.ph/communications/angsurian/2018/volume-1/issue-1/ensuring-the-accuracy-and-quality-of-data-from-dost-astis-automated-weather-stations/ (accessed 2.17.22).

DOST-ASTI, n.d. DOST-ASTI | Advanced Science and Technology Institute [WWW Document]. URL https://asti.dost.gov.ph/ (accessed 3.27.22).

EM-DAT, 2022. The International Disasters Database [WWW Document]. URL https://public.emdat.be/data (accessed 10.28.21).

Espada, R., 2018. Return period and Pareto analyses of 45 years of tropical cyclone data (1970–2014) in the Philippines. Appl. Geogr. https://doi.org/10.1016/j.apgeog.2018.04.018

European Commission, 2021. LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS [WWW Document]. URL https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206 (accessed 6.12.22).

Fakhruddin, B., Kintada, K., Hassan, Q., 2022. Understanding hazards: Probabilistic cyclone modelling for disaster risk to the Eastern Coast in Bangladesh. Prog. Disaster Sci. 13, 100216.

https://doi.org/10.1016/J.PDISAS.2022.100216

FAO, 2020. Bangladesh 2020 severe monsoon floods [WWW Document]. URL http://www.fao.org/3/cb0458en/CB0458EN.pdf (accessed 6.13.21).

Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. Ann. Stat. https://doi.org/10.1214/aos/1013203451

Gevaert, C.M., Carman, M., Rosman, B., Georgiadou, Y., Soden, R., 2021. Fairness and accountability of AI in disaster risk management: Opportunities and challenges. Patterns 2. https://doi.org/10.1016/J.PATTER.2021.100363

Goldstein, A., Kapelner, A., Bleich, J., Pitkin, E., 2015. Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation. J. Comput. Graph. Stat. 24, 44–65. https://doi.org/10.1080/10618600.2014.907095

Goodman, B., Flaxman, S., 2016. European Union regulations on algorithmic decision-making and a "right to explanation."

GOVPH, n.d. PAGASA [WWW Document]. URL https://www.pagasa.dost.gov.ph/index.php (accessed 5.15.22).

GRC, 2022. Considering socioeconomic parameters in triggers for anticipatory action - Anticipation Hub [WWW Document]. URL https://www.anticipation-hub.org/news/considering-socioeconomic-parameters-in-triggers-for-anticipatory-action (accessed 6.4.22).

Greco, S., Ishizaka, A., Tasiou, M., Torrisi, G., 2018. On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness. Soc. Indic. Res. 2018 1411 141, 61–94. https://doi.org/10.1007/S11205-017-1832-9

Grimes, A., Mercer, A.E., 2016. Diagnosing Tropical Cyclone Rapid Intensification Through Rotated Principal Component Analysis of Synoptic-Scale Diagnostic Fields. Recent Dev. Trop. Cyclone Dyn. Predict. Detect. https://doi.org/10.5772/63988

Gros, C., Bailey, M., Schwager, S., Hassan, A., Zingg, R., Uddin, M.M., Shahjahan, M., Islam, H., Lux, S., Jaime, C., Coughlan de Perez, E., 2019. Household-level effects of providing forecast-based cash in anticipation of extreme weather events: Quasi-experimental evidence from humanitarian interventions in the 2017 floods in Bangladesh. Int. J. Disaster Risk Reduct. 41, 101275. https://doi.org/10.1016/J.IJDRR.2019.101275

Heming, J.T., Prates, F., Bender, M.A., Bowyer, R., Cangialosi, J., Caroff, P., Coleman, T., Doyle, J.D., Dube, A., Faure, G., Fraser, J., Howell, B.C., Igarashi, Y., McTaggart-Cowan, R., Mohapatra, M., Moskaitis, J.R., Murtha, J., Rivett, R., Sharma, M., Short, C.J., Singh, A.A., Tallapragada, V., Titley, H.A., Xiao, Y., 2019. Review of Recent Progress in Tropical Cyclone Track Forecasting and Expression of Uncertainties. Trop. Cyclone Res. Rev. 8, 181–218. https://doi.org/10.1016/J.TCRR.2020.01.001

Holland, G., 2008. A Revised Hurricane Pressure–Wind Model. Mon. Weather Rev. 136, 3432–3445. https://doi.org/10.1175/2008MWR2395.1

Hoque, M.A.A., Phinn, S., Roelfsema, C., Childs, I., 2016. Assessing tropical cyclone impacts using object-based moderate spatial resolution image analysis: a case study in Bangladesh. Int. J. Remote Sens. 37, 5320–5343. https://doi.org/10.1080/01431161.2016.1239286

IFRC, 2022a. Disaster Relief Emergency Fund (DREF) | IFRC [WWW Document]. URL https://www.ifrc.org/disaster-relief-emergency-fund-dref (accessed 11.7.21).

IFRC, 2022b. FbA Dashboard | Forecast-based Action by the DREF [WWW Document]. URL

https://public.tableau.com/app/profile/ifrcgo/viz/FbADashboard/Dashboard1 (accessed 11.1.21).

IFRC, 2020a. FbF Practitioners Manual, Chapter 4,Develop Early Action Protocol [WWW Document]. URL https://manual.forecast-based-financing.org/en/chapter/develop-early-action-protocol/ (accessed 6.22.21).

IFRC, 2020b. DREF Final Report Philippines: Typhoon Kammuri [WWW Document]. URL https://reliefweb.int/sites/reliefweb.int/files/resources/MDRPH037dfr.pdf (accessed 11.6.21).

IFRC, 2017. Philippines: Tropical Storm Tembin Emergency Plan of Action (EPoA) DREF n° MDRPH026 - Philippines | ReliefWeb [WWW Document]. URL https://reliefweb.int/report/philippines/philippines-tropical-storm-tembin-emergency-plan-action-epoa-dref-n-mdrph026 (accessed 6.6.22).

ISDR, 2005. Building the Resilience of Nations and Communities to Disasters [WWW Document]. URL www.unisdr.org/wcdr

ISDR, 2004. Terminology: Basic terms of disaster risk reduction [WWW Document]. URL https://www.unisdr.org/2004/wcdr-dialogue/terminology.htm (accessed 11.1.21).

Jan, K., Caravani, A., 2017. Financing Disaster Risk Reduction: A 20 year story of international aid - - Research reports and studies | Enhanced Reader [WWW Document]. URL chrome-extension://dagcmkpagjlhakfdhnbomgmjdpkdklff/enhanced-reader.html?pdf=https%3A%2F%2Fwww.humanitarianresponse.info%2Fsites%2Fwww.humanitarianresponse.info%2Ffiles%2Fdocuments%2Ffiles%2Fkhm_0801_gfdr_financingdrr.pdf (accessed 5.24.21).

Kaplan, J., Demaria, M., 1995. A simple empirical model for predicting the decay of tropical cyclone winds after landfall. J. Appl. Meteorol. https://doi.org/10.1175/1520-0450(1995)034<2499:ASEMFP>2.0.CO;2

Kaplan, J., DeMaria, M., 2003. Large-scale characteristics of rapidly intensifying tropical cyclones in the North Atlantic basin. Weather Forecast. https://doi.org/10.1175/1520-0434(2003)018<1093:LCORIT>2.0.CO;2

Katz, R.W., Lazo, J.K., 2012. Economic Value of Weather and Climate Forecasts. Oxford Handb. Econ. Forecast. https://doi.org/10.1093/OXFORDHB/9780195398649.013.0021

Kerle, N., 2010. Satellite-based damage mapping following the 2006 Indonesia earthquake-How accurate was it? Int. J. Appl. Earth Obs. Geoinf. 12, 466–476. https://doi.org/10.1016/j.jag.2010.07.004

Kirchhoff, C.J., Lemos, M.C., Dessai, S., 2013. Actionable Knowledge for Environmental Decision Making: Broadening the Usability of Climate Science. http://dx.doi.org/10.1146/annurev-environ-022112-112828 38, 393–414. https://doi.org/10.1146/ANNUREV-ENVIRON-022112-112828

Knapp, K.R., Kruk, M.C., Levinson, D.H., Diamond, H.J., Neumann, C.J., 2010. The International Best Track Archive for Climate Stewardship (IBTrACS): Unifying Tropical Cyclone Data. Bull. Am. Meteorol. Soc. 91, 363–376. https://doi.org/10.1175/2009BAMS2755.1

Lagmay, A.M., Kerle, N., 2015. Storm-surge models helped for Hagupit. Nat. 2015 5197544 519, 414–414. https://doi.org/10.1038/519414b

Lagmay, A.M.F., Agaton, R.P., Bahala, M.A.C., Briones, J.B.L.T., Cabacaba, K.M.C., Caro, C.V.C., Dasallas, L.L., Gonzalo, L.A.L., Ladiero, C.N., Lapidez, J.P., Mungcal, M.T.F., Puno, J.V.R., Ramos, M.M.A.C., Santiago, J., Suarez, J.K., Tablazon, J.P., 2015. Devastating storm surges of Typhoon Haiyan. Int. J. Disaster Risk Reduct. 11, 1–12. https://doi.org/10.1016/J.IJDRR.2014.10.006

Langmay, A.M., Racoma, B.A., Aracan, K.A., Alconis-Ayco, J., Saddi, I.L., 2017. Disseminating near-real-time hazards information and flood maps in the Philippines through Web-GIS. J. Environ. Sci. 59, 13–23. https://doi.org/10.1016/J.JES.2017.03.014

LawPhil, 2009. Formation of super-regions. Executive Order No. 561 [WWW Document]. URL https://www.lawphil.net/executive/execord/eo2006/eo_561_2006.html (accessed 4.12.22).

Lopez, A., Coughlan de Perez, E., Bazo, J., Suarez, P., van den Hurk, B., van Aalst, M., 2020. Bridging forecast verification and humanitarian decisions: A valuation approach for setting up action-oriented early warnings. Weather Clim. Extrem. 27, 100167. https://doi.org/10.1016/j.wace.2018.03.006

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M.I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J.B.R., Maycock, T.K., Waterfield, T., Yelekçi, O., Yu, R., Zhou, B., (eds.), 2021. IPCC: Climate Change 2021: The Physical Science Basis. Cambridge Univ. Press. Press.

McInerny, G.J., Chen, M., Freeman, R., Gavaghan, D., Meyer, M., Rowland, F., Spiegelhalter, D.J., Stefaner, M., Tessarolo, G., Hortal, J., 2014. Information visualisation for science and policy: engaging users and avoiding bias. Trends Ecol. Evol. 29, 148–157. https://doi.org/10.1016/J.TREE.2014.01.003

Mechler, R., 2005. Cost-benefit Analysis of Natural Disaster Risk Management in Developing Countries Manual [WWW Document]. URL http://pure.iiasa.ac.at/id/eprint/7320/1/Cost-B-A.pdf (accessed 10.5.21).

Mittelstadt, B., Russell, C., Wachter, S., 2019. Explaining Explanations in AI. https://doi.org/10.1145/3287560.3287574

Miyamoto, Y., Takemi, T., 2015. A Triggering Mechanism for Rapid Intensification of Tropical Cyclones. J. Atmos. Sci. 72, 2666–2681. https://doi.org/10.1175/JAS-D-14-0193.1

Molteni, F., Buizza, R., Palmer, T.N., Petroliagis, T., 1996. The ECMWF Ensemble Prediction System: Methodology and validation. Q. J. R. Meteorol. Soc. 122, 73–119. https://doi.org/10.1002/QJ.49712252905

Montgomery, M.T., Farrell, B.F., 1993. Tropical cyclone formation. J. Atmos. Sci. https://doi.org/10.1175/1520-0469(1993)050<0285:TCF>2.0.CO;2

Mylne, K.R., 2002. Decision-making from probability forecasts based on forecast value. Meteorol. Appl. 9, 307–315. https://doi.org/10.1017/S1350482702003043

NDRRMC, 2019. NDRMCC Update: Situational Report No. 19 regarding response and effects of Typhoon "TISOY" [WWW Document]. URL https://reliefweb.int/sites/reliefweb.int/files/resources/SitRep_No_19_re_Response_Actions_and_Effects_of_TY_Tisoy_I_N_Kammuri_as_of_13Dec2019_6PM.pdf (accessed 11.6.21).

NDRRMC, n.d. National Disaster Risk Reduction and Management Council [WWW Document]. URL https://ndrrmc.gov.ph/index.php (accessed 11.6.21).

NHC, n.d. Definition of the NHC Track Forecast Cone [WWW Document]. URL https://www.nhc.noaa.gov/aboutcone.shtml (accessed 10.24.21a).

NHC, n.d. Glossary of NHC Terms [WWW Document]. URL https://www.nhc.noaa.gov/aboutgloss.shtml (accessed 2.10.22b).

OCHA, 2021a. Anticipatory Action Framework-30 May 2021 Bangladesh Monsoon Floods [WWW

Document]. URL https://reliefweb.int/sites/reliefweb.int/files/resources/Final - Bangladesh AA Framework.pdf

OCHA, 2021b. Philippines: CERF Pilot Anticipatory Action for 2021-2022 (As of 28 August 2021) - Philippines | ReliefWeb [WWW Document]. URL https://reliefweb.int/report/philippines/philippines-cerf-pilot-anticipatory-action-2021-2022-28-august-2021 (accessed 6.8.22).

OCHA, 2021c. PREDICTIVE ANALYTICS IN HUMANITARIAN RESPONSE [WWW Document]. URL https://data.humdata.org/dataset/2048a947-5714-4220-905b-e662cbcd14c8/resource/76e488d9-b69d-41bd-927c-116d633bac7b/download/peer-review-framework-2020.pdf (accessed 6.22.21).

OCHA, 2021d. Common Operational Datasets - Humanitarian Response Intelligence [WWW Document]. URL https://storymaps.arcgis.com/stories/dcf6135fc0e943a9b77823bb069e2578 (accessed 3.30.22).

OCHA, 2020a. Anticipatory action in Bangladesh before peak monsoon flooding – The Centre for Humanitarian Data [WWW Document]. URL https://centre.humdata.org/anticipatory-action-in-bangladesh-before-peak-monsoon-flooding/ (accessed 6.19.21).

OCHA, 2020b. Philippines: Typhoon Vamco (Ulysses) and Super Typhoon Goni (Rolly) Snapshot (As of 14 November 2020) - Philippines | ReliefWeb [WWW Document]. URL https://reliefweb.int/report/philippines/philippines-typhoon-vamco-ulysses-and-super-typhoon-goni-rolly-snapshot-14 (accessed 5.19.22).

OXFAM, 2021. Groups provide pre-disaster financial aid to more than 2,600 families in E. Samar before Typhoon Odette hits | Oxfam Philippines [WWW Document]. Press Release. URL https://philippines.oxfam.org/latest/press-release/groups-provide-pre-disaster-financial-aid-more-2600-families-e-samar-typhoon (accessed 4.23.22).

Pacheco, B.M., Hernandez Jr, J.Y., E.A.J., T., Castro, P.P.., Germar, F.J., Ignacio, U.P., 2014. Development of Vulnerability Curves of Key Building Types to Different Hazards in the Philippines [WWW Document]. URL https://dokumen.tips/reader/f/development-of-vulnerability-curves-of-key-building-types-to-different-hazards (accessed 2.9.22).

PAGASA, 2022. Modified Tropical Cyclone Warning System - PAGASA [WWW Document]. URL https://bagong.pagasa.dost.gov.ph/index.php/press-release/108 (accessed 4.7.22).

Powell, M.D., Houston, S.H., Reinhold, T.A., 1996. Hurricane Andrew's landfall in South Florida. Part I: Standardizing measurements for documentation of surface wind fields. Weather Forecast. https://doi.org/10.1175/1520-0434(1996)011<0304:HALISF>2.0.CO;2

Powell, M.D., Murillo, S., Dodge, P., Uhlhorn, E., Gamache, J., Cardone, V., Cox, A., Otero, S., Carrasco, N., Annane, B., Fleur, R.S., 2010. Reconstruction of Hurricane Katrina's wind fields for storm surge and wave hindcasting. Ocean Eng. 37, 26–36. https://doi.org/10.1016/J.OCEANENG.2009.08.014

PRC, IFRC, GRC, FRC, Climate Centre, 2020. TYPHOON EARLY ACTION PROTOCOL [WWW Document]. URL https://www.forecast-based-financing.org/wp-content/uploads/2020/09/Typhoon-Pocket-EAP.pdf (accessed 6.23.21).

PSA, 2021. 2020 Census of Population and Housing (2020 CPH) Population Counts Declared Official by the President | Philippine Statistics Authority [WWW Document]. URL https://psa.gov.ph/content/2020-census-population-and-housing-2020-cph-population-counts-declared-official-president (accessed 4.11.22).

PSA, n.d. Philippine Statistics Authority [WWW Document]. URL https://psa.gov.ph/ (accessed 3.1.22).

Radford, A.M., 1994. Forecasting the movement of tropical cyclones at the Met. Office. Meteorol. Appl. 1, 355–363. https://doi.org/10.1002/MET.5060010406

Red Cross Red Crescent Center, 2020. The Future of Forecasts: Impact-based Forecasting for Early Action - International Federation of Red Cross and Red Crescent Societies [WWW Document]. URL https://media.ifrc.org/ifrc/document/future-forecasts-impact-based-forecasting-early-action/ (accessed 5.24.21).

Roy, C., Sarkar, S.K., Åberg, J., Kovordanyi, R., 2015. The current cyclone early warning system in Bangladesh: Providers' and receivers' views. Int. J. Disaster Risk Reduct. 12, 285–299. https://doi.org/10.1016/J.IJDRR.2015.02.004

Saha, M., Ahsan, M.N., Shit, P., Saha, M.K., Khan, N.A., 2014. Changing Profile of Cyclones in the Context of Climate Change and Adaptation Strategies in Bangladesh 7, 1–12.

Santos, G.D.C., 2021. 2020 tropical cyclones in the Philippines: A review. Trop. Cyclone Res. Rev. 10, 191–199. https://doi.org/10.1016/J.TCRR.2021.09.003

Shreve, C.M., Kelman, I., 2014. Does mitigation save? Reviewing cost-benefit analyses of disaster risk reduction. Int. J. Disaster Risk Reduct. 10, 213–235. https://doi.org/10.1016/J.IJDRR.2014.08.004

Tallapragada, V., 2016. Overview of the NOAA/NCEP operational hurricane weather research and forecast (HWRF) modelling system. Adv. Numer. Model. Data Assim. Tech. Trop. Cyclone Predict. 51–106. https://doi.org/10.5822/978-94-024-0896-6_3

Tanner, T., Gray, B., Guigma, K., Iqbal, J., Levine, S., Macleod, D., Nahar, K., Rejve, K., Cabot Venton, C., 2019. Scaling up early action Lessons, challenges and future potential in Bangladesh.

The Philippines Humanitarian Country Team, 2020. SUPER TYPHOON GONI (ROLLY) AND TYPHOON VAMCO (ULYSSES) PHILIPPINES HUMANITARIAN NEEDS AND PRIORITIES [WWW Document]. URL https://reliefweb.int/sites/reliefweb.int/files/resources/PHL-TyphoonGoniVamco-HumNeedsPriorities-Revision-201126.pdf (accessed 10.24.21).

Thornes, J.E., Stephenson, D.B., 2001. How to judge the quality and value of weather forecast products. Meteorol. Appl. 8, 307–314. https://doi.org/10.1017/S1350482701003061

Tóth, Z., Zhu, Y., 2003. Probability and Ensemble Forecasts [WWW Document]. URL https://www.researchgate.net/publication/266335540 (accessed 10.13.21).

Twomlow, A., Grainger, S., Cieslik, K., Paul, J.D., Buytaert, W., 2022. A user-centred design framework for disaster risk visualisation. Int. J. Disaster Risk Reduct. 77, 103067. https://doi.org/10.1016/J.IJDRR.2022.103067

U.S Weather Bureau, 1961. Meteorological Satellite Laboratory Report Manuscript [WWW Document]. URL https://ntrs.nasa.gov/api/citations/19620001116/downloads/19620001116.pdf (accessed 10.23.21).

Valentijn, T., Margutti, J., van den Homberg, M., Laaksonen, J., 2020. Multi-hazard and spatial transferability of a CNN for automated building damage assessment. Remote Sens. 12, 1–29. https://doi.org/10.3390/RS12172839

van den Homberg, M., Monné, R., Spruit, M., 2018. Bridging the information gap of disaster responders by optimizing data selection using cost and quality. Comput. Geosci. 120, 60–72. https://doi.org/10.1016/J.CAGEO.2018.06.002

van den Homberg, M.J.C., Gevaert, C.M., Georgiadou, Y., 2020. The changing face of accountability in humanitarianism: Using artificial intelligence for anticipatory action. Polit. Gov.

https://doi.org/10.17645/pag.v8i4.3158

Wagenaar, D., Hermawan, T., van den Homberg, M.J.C., Aerts, J.C.J.H., Kreibich, H., de Moel, H., Bouwer, L.M., 2021. Improved Transferability of Data-Driven Damage Models Through Sample Selection Bias Correction. Risk Anal. 41, 37–55. https://doi.org/10.1111/risa.13575

Wang, B., Zhou, X., 2007. Climate variation and prediction of rapid intensification in tropical cyclones in the western North Pacific. Meteorol. Atmos. Phys. 2007 991 99, 1–16. https://doi.org/10.1007/S00703-006-0238-Z

WFP, 2016. WFP El Niño 2015-2016 PREPAREDNESS AND RESPONSE [WWW Document]. URL https://documents.wfp.org/stellent/groups/public/documents/ep/wfp282093.pdf (accessed 10.25.21).

Wieringa, J., 1992. Updating the Davenport roughness classification. J. Wind Eng. Ind. Aerodyn. 41, 357–368. https://doi.org/10.1016/0167-6105(92)90434-C

Wilkinson, E., Weingärtner, L., Choularton, R., Bailey, M., Todd, M., Kniveton, D., Cabot Venton, C., 2018. Forecasting hazards, averting disasters Implementing forecast-based early action at scale [WWW Document]. URL https://cdn.odi.org/media/documents/12104.pdf (accessed 10.25.21).

Willitts-King, B., Weingärtner, L., Pichon, F., Spencer, A., 2020. Risk-informed approaches to humanitarian funding Using risk finance tools to strengthen resilience.

WMO, 2021a. Weather-related disasters increase over past 50 years, causing more damage but fewer deaths | World Meteorological Organization [WWW Document]. URL https://public.wmo.int/en/media/press-release/weather-related-disasters-increase-over-past-50-years-causing-more-damage-fewer (accessed 10.26.21).

WMO, 2021b. Tropical Cyclone Programme | World Meteorological Organization [WWW Document]. URL https://public.wmo.int/en/programmes/tropical-cyclone-programme (accessed 10.23.21).

WMO, 2020. Tropical Cyclones | World Meteorological Organization [WWW Document]. URL https://public.wmo.int/en/our-mandate/focus-areas/natural-hazards-and-disaster-risk-reduction/tropical-cyclones (accessed 10.23.21).

WMO, 2015. WMO Guidelines on Multi-hazard Impact-based Forecast and Warning Services [WWW Document]. URL https://library.wmo.int/doc_num.php?explnum_id=7901 (accessed 10.17.21).

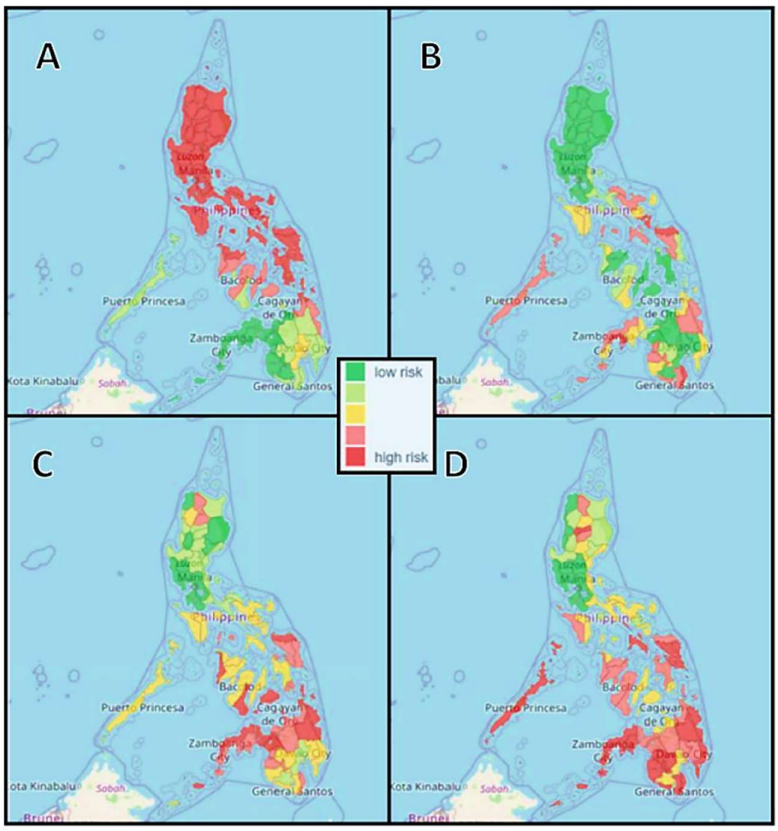
WMO, 2010. GUIDELINES FOR CONVERTING BETWEEN VARIOUS WIND AVERAGING PERIODS IN TROPICAL CYCLONE CONDITIONS.

WMO, 2008. Communicating Forecast Uncertainty for Service Providers | World Meteorological Organization [WWW Document]. URL https://public.wmo.int/en/bulletin/communicating-forecast-uncertainty-service-providers (accessed 6.9.22).

Yu, H., Chen, G., Brown, B., 2013. A New Verification Measure for Tropical Cyclone Track Forecasts and its Experimental Application. Trop. Cyclone Res. Rev. 2, 185–195. https://doi.org/10.6057/2013TCRR04.01

7 ANNEX

Annex 1: Typhoon exposure and vulnerability in the Philippines retrieved from the community risk assessment dashboard[13]



Map A	Cyclones exposure
Map B	Percentage of houses with a strong roof
Map C	Percentage of houses with strong wall type
Map D	Vulnerable population (that includes Child head of HH, Single head of HH, Disabled person, Solo Parent, older people)

Annex 2: Dataset

Table 15 Source of all the dataset and the models used during this research

Impact Model	Data Description	Source	Retrieved from
Statistical Modelling	Windspeed Estimation for statistical modelling was done using the existing model built by 510 that uses Holland 2008 parametric modelling	510 IBF Typhoon Model	github.com/rodekruis

[13] https://dashboard.510.global/#!/community_risk

	Impact prediction for statistical model was made using the existing impact model from 510	510 Typhoon Impact Model	
	Forecasted wind speed and track for historical events	ECMWF	
	Pre-disaster indicators used as explanatory and target variables in the impact model were used from the existing model input database	Multiple Sources	
	Population Statistics	Philippines Statistics Authority	psa.gov.ph/statistics/census/projected-population
Elementary modelling	Actual windspeed and landfall location for Kammuri	IBTrACKS	ibtracs.unca.edu/
	Forecasted wind speed and landfall location for Kammuri	Joint Typhoon Warning Center (JWTC)	Media reportings
	Wind observations from Automatic Weather Stations for 8 historical typhoons	Philsensors- Dost-Asti	https://philsensors.asti.dost.gov.ph/
	Indicators for composite weighted overlay to calculate the vulnerability index	Human Data Exchange by UN-OCHA	humdata.org
	Post-event household damage data for historical events to build the damage curve	National Disaster Risk Reduction and Management Council, Philippines	510 Database
Validation	Household damage data for typhoon Kammuri to validate the predictions of both the models	National Disaster Risk Reduction and Management Council, Philippines	https://ndrrmc.gov.ph/ Tisoy Situational Report No: 19

Annex 3: Data quality issues:

Place code, also known as P-code, which is being adopted in the humanitarian community, represents individual administrative regions with distinct values (OCHA, 2021d). P-codes allow a systematic linkage of data, avoiding any possibility of miscounting and making the validation process much faster. Currently, the administrative regions of the Philippines are also set up based on similar P-codes for three administrative levels (region, province, and municipality), as shown in Figure 28, creating a standard dataset for all the processes.

Figure 28 An example of a P-Code used to link all the datasets for the Philippines at the municipality level

Table 16 Examples of inconsistency in naming conventions of administrative boundaries in official data sources

NDRRMC Data	Standard Dataset with P-codes
CALABARZON	Region IV-A
SagÃƒÂ±ay	sagnay
City of naga	naga city

Table 17 Example of duplication in official names of administrative boundaries in official data sources

Name of Municipality	Region	Province
San Isidro	Region II	Isabela
San Isidro	Region III	Nueva Ecija
San Isidro	Region VII	Bohol
San Isidro	Region VIII	Leyte
San Isidro	Region VIII	Northern Samar
San Isidro	Region XI	Davao del Norte
San Isidro	Region XI	Davao Oriental
San Isidro	CAR	Abra
San Isidro	Region XIII	Surigao del Norte

Table 18 An example illustrating the discrepancy in the count of damage due to Kammuri when compared against the current input database for the model

Province	Municipality	Observed number of completely damages houses in the municipality (after manual correction)	Observed number of completely damages houses in the municipality (In the current model input)	Total Housing Units	Percentage of houses totally destroyed
Occidental Mindoro	San Jose	40	0	33208	0.12
Romblon	San Jose	274	0	2392	11.45
Camarines Sur	San Jose	74	0	8684	0.85
Northern Samar	San Jose	554	0	3907	14.18

Annex 4: Windspeed Reduction Factor

Figure 29 Example of four imagery with locations of wind observation stations, and the corresponding roughness coefficient (Zo) assigned to it

It can be observed that there are not very significant differences in the reduction factor before and after normalizing the observations. This might lie in the fact that the values are expressed as a fraction of landfall windspeed, instead of absolute number. However, it should be noted that the normalization factor used here is merely an estimate and are subject to deviation from the true value.

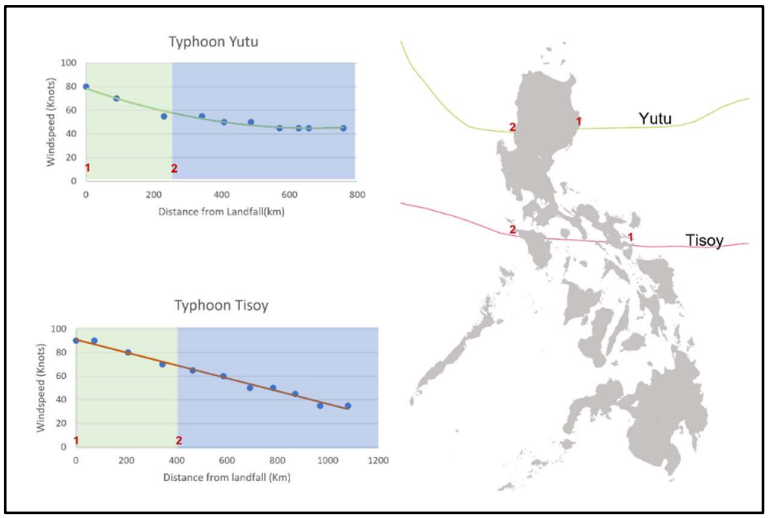


Figure 30 A representation of how wind speed decays over the distance from landfall based on windspeed recorded at different intervals of the track by WMO

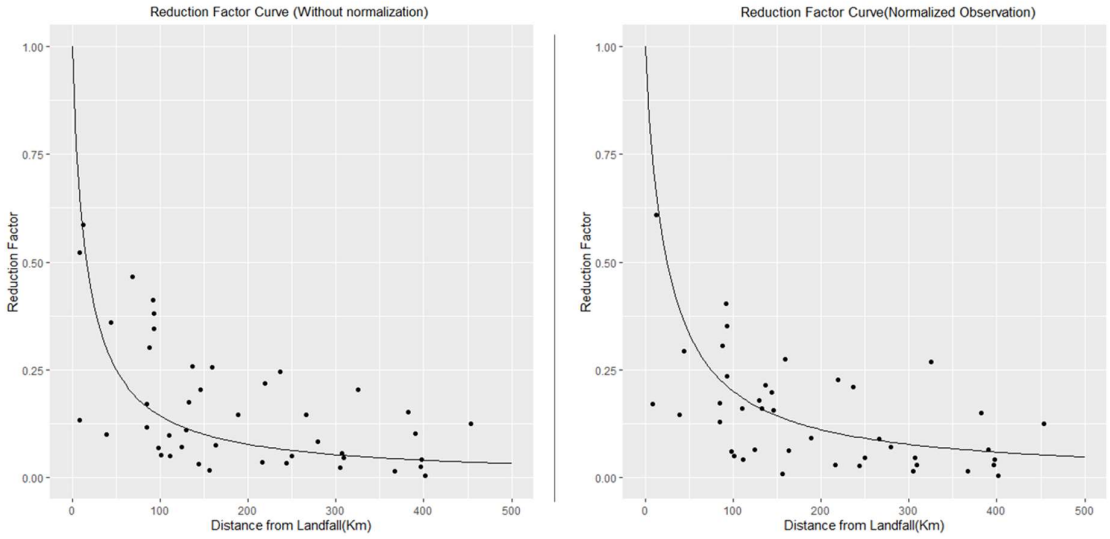


Figure 31 Windspeed reduction factor curve obtained for the Philippines with factor of reduced windspeed (observations -AWS) against the distance from landfall plotted without the normalization (left) and after normalizing (right)

Annex 5 Damage Curves

Table 19 List of all the events and the number of municipality if affected (at least one completely damage house) across 4 different regions.

	GONI (2020)	PHANFONE (2019)	HAIMA (2013)	VAMCO (2020)	YUTU (2018)	SARIKA (2016)	HAIYAN (2013)	Total affected municipalities
Central	74	113	1	40	2	0	298	528
Metro Luzon	44	10	8	74	0	0	24	160
North Luzon	37	5	292	110	70	57	2	573
Mindanao	1	1	0	0	0	0	19	21

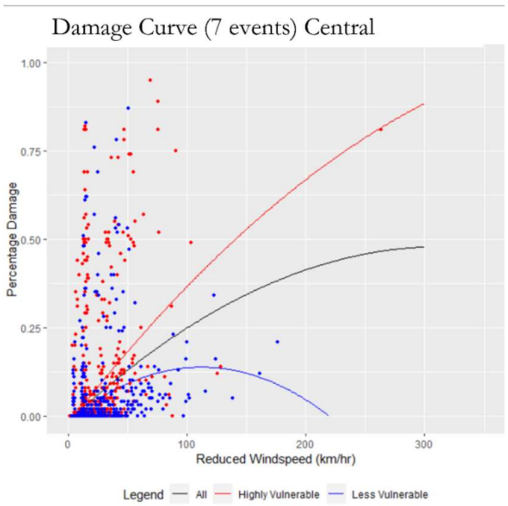
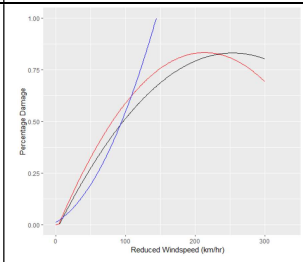
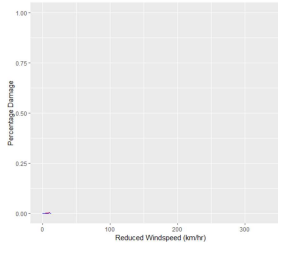
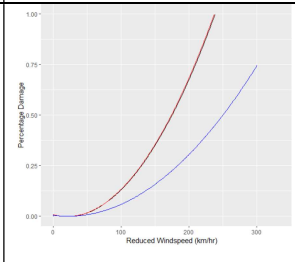
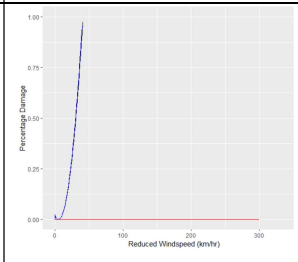
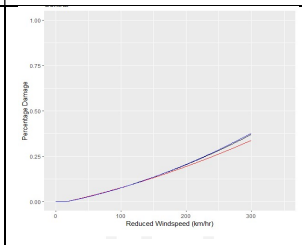
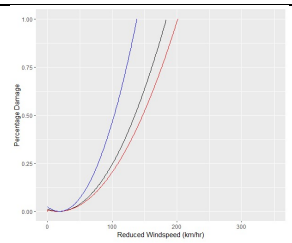
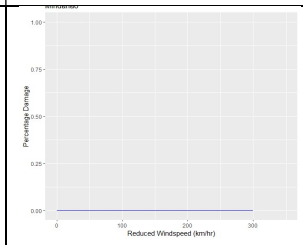
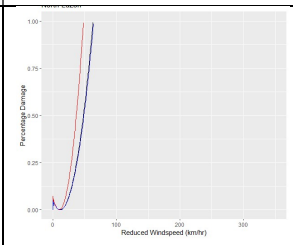


Figure 32 Data points of the damage curve for central region using 7 events.

Table 20 Damage curves for four super regions based on combination of different events which highlight the role of event characteristics in influencing the shape of the curve

	Central	Metro Luzon	Mindanao	North Luzon
1 Event (Haiyan)				
1 Event (Goni)				

Annex 6: Vulnerability index generated for different indicators

Poverty index: The index for the percentage of the population below the poverty line, shows higher values in southern areas of the country, the coastal parts of the Central region and some areas of North Luzon.

Dependent population index: In terms of dependent populations, a low number of municipalities can be identified with values above 80 %. Areas of Luzon and Central Philippines have a relatively higher index, but the difference is not very substantial.

House type index: Mindanao shows a very high proportion of houses built with light weight material and that are susceptible to wind, followed by the Central Philippines and coastal areas of Northern Luzon.

Evacuation centre accessibility index: Compared to other indices, the percentage of the population in terms of access to an evacuation cente is unevenly distributed amongst the municipalities and cannot be generalized as being high or low in a particular region. However, some islands in the central Philippines which are in the coastal extremes and susceptible to frequent typhoons, have 80 to 100 % of population without access to an evacuation centre, which is alarming.

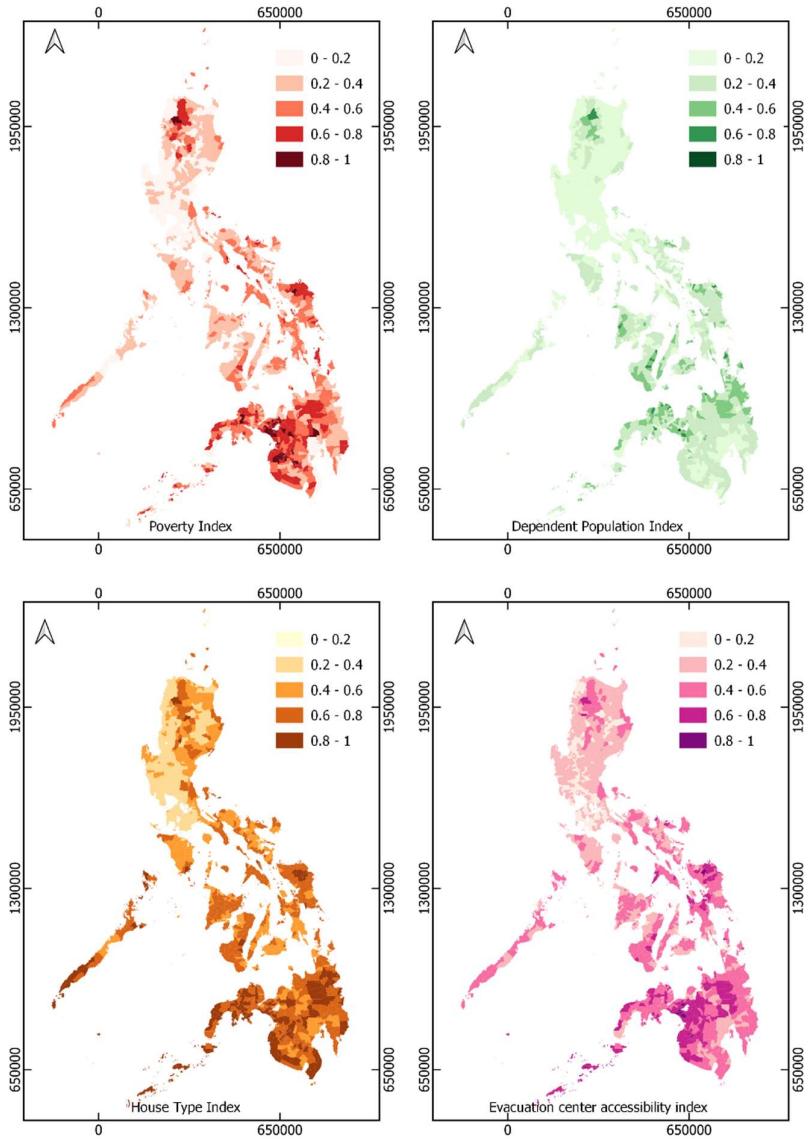


Figure 33 Index for 4 indicators of vulnerability generated at municipality level

Annex 7: Description of decision support tool

Variables	Description	Functional for:	Remarks
Landfall X, Landfall Y, Maximum sustained wind	This is the forecasted or actual parameters of the typhoon and can be tested for any event and lead time	Elementary method	The coordinated must be entered in projected coordinate system. Windspeed must be 1-min sustained.
Landfall location and windspeed error	This allows a selection of tolerance of error in forecast for re-calculation of reduced windspeed	Elementary method	
Actual Vs Predicted Map	Displays the landfall location and impact that was reported and what was predicted based on the variables selected.	Elementary and Statistical method	The two maps are mirrored so can be interacted to make area specific comparisons as shown in Figure 24. The cursor can also be hovered around each municipality to get more details such as damage percentage and windspeed.
Damage Curve	To display the damage curve which is being used to convert calculated windspeed into damage. Also allows selection of number of events for the curve construction and the area curve to display.	Elementary method	Based on the number of events, prediction will be recalculated but the selection of super region is only for display purpose
Accuracy metrics	The bar chart on the bottom left displays the accuracy of prediction when compared against the actual damage from the event.	Elementary and Statistical method	Can be downloaded as an image to allow comparison of multiple scenarios
Trigger threshold	Selection of damage percentage which will be used as the threshold for trigger	Elementary and Statistical method	This value cannot affect the prediction but will only change the calculation of accuracy metrics
Priority Municipality	Out of all the predictions, number of municipalities to prioritize based on vulnerability index	Elementary method	This value cannot affect the prediction but will only change the calculation of accuracy metrics
Select Map to Display	Selection of actual damage or vulnerability index calculated	-	The index is based on results obtained from section 4.2.4
Select a typhoon	Contains list of past typhoons in the area	-	Currently meant only for displaying the actual damage for the event. The prediction results from statistical modelling only include Kammuri.
Select a region	To select a region for prediction and display	Elementary method	This will reduce the processing time for faster predictions

Annex 8: Accuracy Metrics

Table 21 Accuracy metrics using elementary impact model, based on number of municipalities exceeding the trigger threshold of 10% damage, considering four different allowances of error in landfall location

Elementary Model				
Landfall location error	0 km	50 km	100 km	200 km
TP	6	3	8	16
FP	4	2	17	74
FN	30	33	28	20
Precision	0.60	0.60	0.32	0.18
Recall	0.08	0.17	0.22	0.44
F1 Score	0.15	0.26	0.26	0.25

Table 22 Accuracy metrics using elementary impact model, based on number of municipalities exceeding the trigger threshold of 10% damage, considering four different allowances of error in landfall windspeed

Landfall Windspeed Error	0 km/hr	20 km/hr	50 km/hr	70 km/hr
TP	3	4	4	4
TN	1595			
FP	2	5	5	5
FN	33	32	32	32
Precision	0.60	0.44	0.44	0.44
Recall	0.08	0.11	0.11	0.11
F1 Score	0.15	0.18	0.18	0.18