Geospatial Impact Analytics of Hydrometeorological Hazards: A Study on Urban and Suburban Floods in Sri Lanka using Online Textual Data

S. M. Dassanayake
Department of Decision Sciences, Faculty of Business, University of Moratuwa.
sandund@uom.lk

I. Mahakalanda
Department of Decision Sciences, Faculty of Business, University of Moratuwa.
indram@uom.lk

D. M. R. Sanjula B. Dissanayake
Department of Decision Sciences, Faculty of Business, University of Moratuwa.
dissanayakedmrsb.19@uom.lk

R. M. Pasan I. Gunathunga
Department of Decision Sciences, Faculty of Business, University of Moratuwa.
gunathungarmpi.19@uom.lk

N. K. G. Thilina P. Jayasooriya
Department of Decision Sciences, Faculty of Business, University of Moratuwa.
jayasooriyankgtp.19@uom.lk

R. M. Prageeth S. Rathnayake
Department of Decision Sciences, Faculty of Business, University of Moratuwa.
rathnayakarmps.19@uom.lk

C. L. Jayawardena
Department of Earth Resources Engineering, University of Moratuwa, Sri Lanka.
chulanthaj@uom.lk

Raj Prasanna
Joint Centre for Disaster Research, Massey University, Wellington, New Zealand.
R.Prasanna@massey.ac.nz

ABSTRACT

Urban and suburban communities in tropical countries like Sri Lanka typically experience hydrometeorological hazards that substantially damage property and lives. Although accurate forecasts of weather events are available, the decision-makers often fail to mitigate the actual impact of these forecasts alone. The adverse impacts experienced by the community and reported by news and online media complement this fact. The forecast-impact disparity underpins the scope for holistically linking the forecast data with actual impact. This paper presents a work-in-progress study that develops a geospatial analytics framework using online textual data for assessing the spatiotemporal impact of the hydrometeorological hazards in disaster hot spots. The preliminary findings show prospects for extending the study to impact-focused visualization and forecasting that capture the community's and decision makers' attention for better interventions. For example, these include the degree of disaster response, planning and scheduling critical infrastructure and estimating damages, compensations and insurance claims.

Keywords

Impact-based warnings, geospatial impact, textual data, urban flood.

INTRODUCTION

Hydrometeorological hazards such as floods, landslides, and mudslides caused by extreme weather events, such as hurricanes, tornadoes, and tropical monsoons, inflict significant damage on property and human lives. Sri Lanka, a tropical island nation close to the Bay of Bengal, frequently experiences the adverse consequences of hydrometeorological hazards. The Disaster Management Center of Sri Lanka reported that approximately 1.98 million Sri Lankans were affected by hydrometeorological hazards (mainly floods) from 2009 to 2018 (Basnayake et al., 2019). Moreover, Dias et al. (2018) employed a flood damage forecast model to show that a flood event in a 140 square km urban area, with average inundation depths ranging from 0.48 m to 1.28 m (and outliers up to 5.8 m), could inflict damages that range from USD 37 to 549 million. Many factors, including anthropogenic activities such as wetland reduction, land reclamation, and unsustainable development projects, can cause the impact severity of flood events to significantly differ from the predictions (Wagenaar et al., 2019). Investing in prevention, mitigation, and preparedness strategies that counter these causes can significantly mitigate the impact and severity of the hazards while reducing the inflicted economic damages by 70% (Shreve and Kelman, 2014). However, in developing countries such as Sri Lanka, the problems in regulatory structure, lack of funding, deficiencies in necessary laws and regulations, and lack of required resources and skills hinder the implementation of holistic measures for disaster resilience (Ginige et al., 2010; Jayawardena, 2015). A more pragmatic approach should enable a swift decision-making process for households, authorities, and government entities to minimize the impact severity of hydrometeorological hazards. Such an approach could use an input-output variable-based impact forecasting system, which requires lesser capital investments and resources.

Figure 1 shows a macro-level aggregation of the textual data on contemporary studies focusing on the hazards in Sri Lanka for the period 2011-2022. The bibliographic search conducted on the Scopus database (https://www.scopus.com/search) using the search term (TITLE-ABS-KEY (“hazards" AND "Sri Lanka")) yielded 494 results in this area. As shown in Figures 1a and b, the studies have focused on three fronts: disaster risk reduction (red colour code), understanding the natural disaster phenomena (blue colour), and socio-economic impact of disasters (green colour code). The studies scrutinize the approaches to modelling the impact of natural hazards using physical or computational approaches. Moreover, Figure 1c shows that the links between studies have minimal scope for the impact-based forecasting of meteorological hazard events. The application of hydrological models such as the Synthesis and Reservoir Regulation (SSARR) hydrologic model, the Hydrologic Modeling System (HEC-HMS), the North American Mesoscale Model (NAM) and the Non-linear Cascade Model has been a common trend to systematically forecast the physical damages in most countries similar to Sri Lanka (Sivakumar, 2015). For instance, this practice has been commonly employed in countries like India, Thailand, Kampodiy, Vietnam and China in forecasting the flood events induced by rivers in inter-state or inter-country regions (Sivakumar, 2015). The impact of the hydrometeorological hazards on the river basins and urban and sub-urban regions in Sri Lanka can be seen in the studies (Figure 1d). The notion has shifted towards using social media and natural hazard effects on communities in more recent studies (i.e., 2017-2018). More flooded sites were found in central urban districts, generally residential or industrial areas, roads, and other traffic-related places for hot spots affected by the disaster.

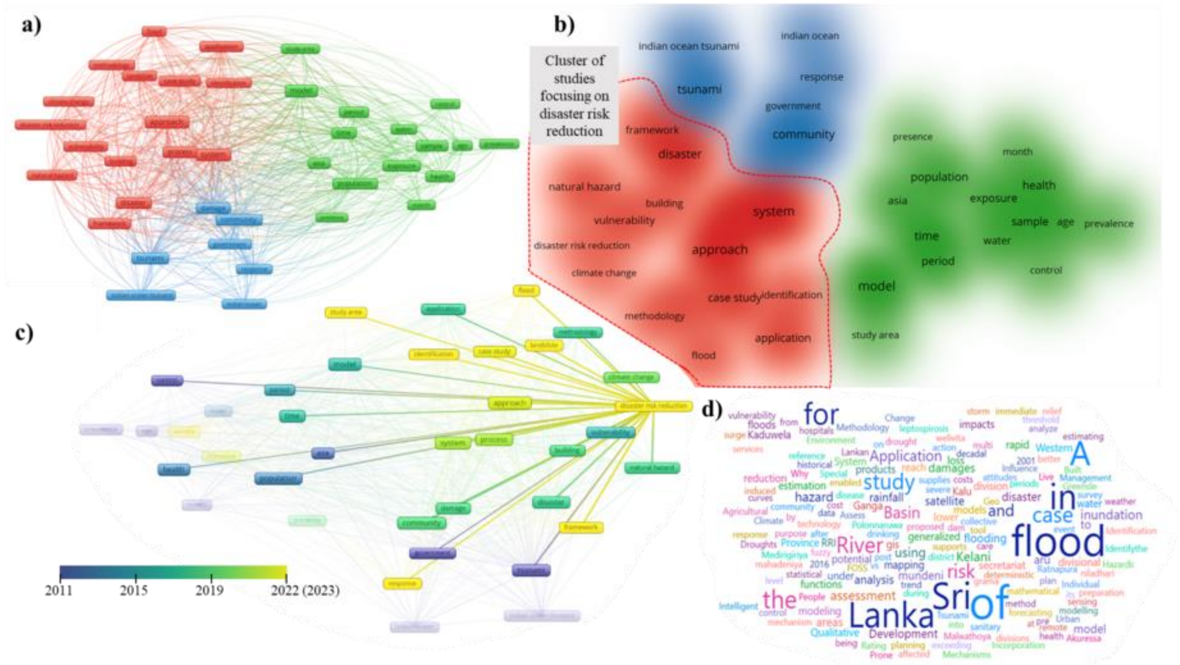

Figure 1. Textual data analysis of the studies pertinent to hazards in Sri Lanka published on Scopus indexed journals from 2011 to 2022: a) network of bibliographic cluster analysis of 494 articles [*Search term:* TITLE-ABS-KEY (“hazards" AND "Sri Lanka")]; b) high frequency words in respective study clusters (colour code: red - disaster risk reduction, blue - understanding natural disaster phenomena, green - socio-economic impact of disasters); c) key focus areas linked to disaster risk reduction with the year of publication; d) word cloud generated from the abstracts of the articles related to hydrometeorological hazards

Although the predictive capability of hydrometeorological models has significantly improved with time, the disasters continue to result in severe adverse impacts on the communities (Weyrich et al., 2018). The evidence suggests that people tend to overlook the hydrometeorological warnings since they have little understanding of the forecast with the potential damages (Weyrich et al., 2018). The recent literature advocates including information about hazards, impacts, and 'what to do' in a warning message (e.g., Singkran 2017; Potter et al. 2018; Robbins and Titley, 2018). Potter et al. (2018) indicate that impact-based warnings may be more effective than phenomenon-based warnings in influencing the recipient's perception of the hazardous event (their sense of threat, concern, and understanding of the potential impacts). Robbins and Titley (2018) show that a symbol and polygon-based summary layer give an at-a-glance view of likely high-impact weather for the week ahead, together with the gridded daily probability forecasts would make the forecast message stronger. For instance, shifting from a passive response that relies mainly on structural measures and emergency responses to a progressive response that emphasizes non-structural measures and participatory collaboration among government agencies and stakeholders (people, public, and private agencies in the affected areas) have improved the flood risk management of Thailand (Singkran, 2017). Word frequency analysis and text mining from social media have substantial applications in impact forecasting and modelling the dynamics of hurricane evacuation decisions (Jayasingh et al., 2016; Fang et al., 2019; Dou et al., 2021; Roy and Hasan, 2021). The significant advantage of these attempts is the rapid modelling capability of temporal variation of precipitation distribution and hazard intensity using real-time data. Yeo et al. (2018) showed that rapid network assessment and manual content analysis could expedite the information extraction from the megabytes of digitally available data. The recent developments in these areas have focused on rapid and accurate disaster response strategies using the fusion of multi-source multimodal data through robust deep learning techniques (Algiriyage et al., 2022). Nevertheless, ensuring the accuracy of online information and expanding the application for all phases of disaster management pose substantial challenges for future research.

This work-in-progress paper shows the potential of developing a geospatial analytics model that provides both visualizations and predicts the impacts of hydrometeorological hazards by extracting both online and offline weather impact data for Sri Lanka. Moreover, to counter the absence of online data in rural Sri Lanka due to infrastructure issues and a relatively low number of technology users, this study explores the possibility of collecting primary textual data reported from various stakeholders. The collation of data from such sources expands the sample size and the number of attributes or predictive variables to analyze. The working paper employs online impact reports to extract the textual data and mine the location information. The geospatial distribution of the hazard events is visually analyzed using the mined data employing open-source geographical information system software (i.e., QGIS). The prospective steps of the study would develop supervised learning classification models to predict the impacts using the insights taken from visual analytics. The study will produce a recommender system that can provide descriptive, predictive and scenario analytics for the decision makers to make informed decisions almost in real-time.

PROPOSED METHODOLOGY

Figure 2 shows a holistic framework of the work-in-progress methodology used in this study. It contains the following phases: data acquisition, preprocessing, development of analytics models, visualization analytics and impact predictions. Data acquisition tools extract weather impact data from online newspaper articles, social media, and other online sources. In the data preprocessing stage, insights into hydrometeorological hazards are derived using descriptive statistic measures. Here, the data summarizations provide basic information about the characteristics of the variables in the data set, such as the mean number of deaths, and injured, where the highest number of events occurred, and the relationships between the variables.

Moreover, data transformations, such as min-max and log transformations, are applied to address the skewness of scale-dependent multi-modal data. The development of the analytics model includes feeding the data to a geospatial database and visually analyzing the impacts. The impact prediction model will be developed using a machine learning model that feeds the output dashboard (developed using Microsoft Power Bi) to visualize and better reveal the impact of the hazard. As such, the optimum machine learning model calibration would be defined based on the insight derived from the preliminary analytics phases.

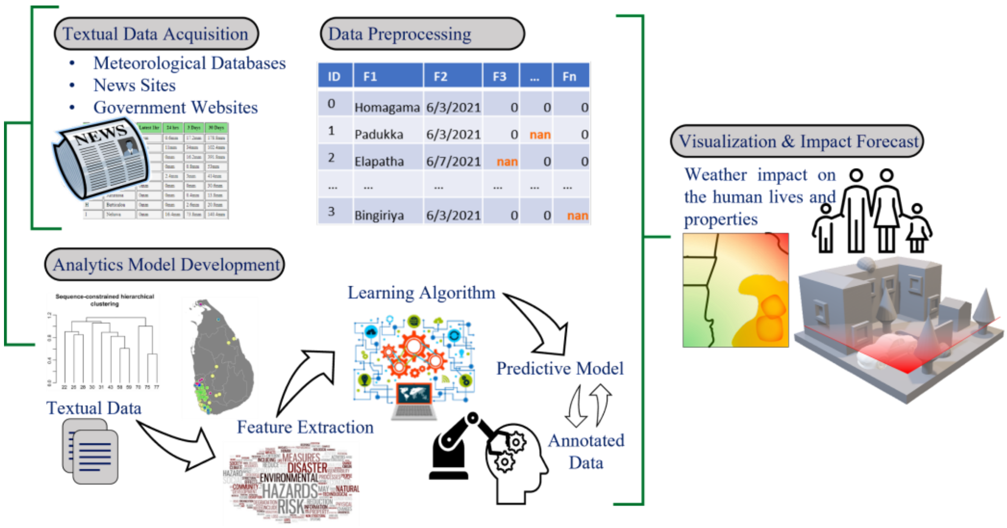

Figure 2. Proposed integrated framework for forecasting flood impact using online textual data

PRELIMINARY WORK

Textual Data Acquisition

The weather forecast data for Sri Lanka from the year 2016 to 2022 was taken from the website of the Department of Meteorology Sri Lanka (meto.gov.lk). The impact data was collected from a disaster inventory system called DesInventar (desinventar.net). The DesInventar database was created in 1994 by the Network for Social Studies in Disaster Prevention in South America (Montgomery and Pinchoff, 2017). DesInventar offers an updated version of the local disaster index within the guidelines of the disaster risk and management indicators program with the aid of the Inter-American Development Bank for the Americas (Marulanda et al., 2010; Edirisinghe and Maduranga, 2021). Also, DesInventar has found applications in Economic Analysts and Sectoral decision-makers since it can identify both the potential effects of extreme events and the persistent accumulation of small local disasters (Marulanda et al., 2010). Moreover, the reported information has been manually fact-checked using local news websites.

Data Pre-processing

Text-based data acquired using web data mining contains repeated information with incomplete information on geolocations. Typically, the dataset contains missing values since online news sources report the incidents differently. In some instances, the reported locations were ambiguous: the central city was reported instead of the administrative division. Thus, the dataset was manually checked to identify the significant discrepancies. The missing values were replaced by the highest frequency of the reported data obtained from the meteorology department of Sri Lanka (meto.gov.lk). Text analytics and matching algorithms were used to check for discrepancies in the English spellings of the local town names using a lexicon of the towns and villages. Table 1 contains the ordered principal feature classes pre-processed for the descriptive and geospatial visual analytics.

Table 1. Feature Collection

Feature	Description
Event	Floods occurred due to hydrometeorological events
Division	Administrative area or city that is affected by the incident
Date	Date of occurrence of the flood event
Deaths	Number of fatalities recorded as directly caused by the flood event
Injured	Number of injured recorded as directly caused by the flood event
Missing	Number of individuals whose whereabouts is unknown since the disaster. This also includes people who are presumed dead, although there is no physical evidence
Houses Destroyed	Number of houses levelled, buried, collapsed, or damaged to the extent that they are no longer habitable
Houses Damaged	Number of houses with minor damage, not structural nor architectural, which may continue being lived in, although they may require some repair or cleaning
Affected Individuals	Number of individuals including all gender and age who were affected by the flood event
Families Affected	Number of families affected by the flood event
Rainfall (mm)	The precipitation height (in mm) reported by the selected media

Analytics Model Development

The impact of the disaster evolved with time and space, presenting scope for establishing a data-driven approach from the forecast and the reported impact data. Figure 3 shows the repeated events and locations and indicates the property damage. The number of houses damaged has been used as a measure for this because the community equally experiences the same impact. The figure shows that the impact can fluctuate in different magnitudes since the number of families affected had changed in orders of tenfold when the impact on the property was comparatively low. In the case of high property damage, the affected families were less. This shows the uncertainty associated with the impact predictions. A potential approach to solve this issue would be employing another layer of data - a data fusion technique to include the qualitative measure of the inundation. Also, a digital elevation model with a proper dataset on the building locations would enable correlating the actual impact. However, that requires multiple parameters such as the surface characteristics of the area, slope characteristics and the vegetation characteristics of the respective area to overlay the inundation map on the locations. The data well represent the impact of the 2017-2018 extreme weather events since a relatively uniform impact was recorded throughout Sri Lanka.

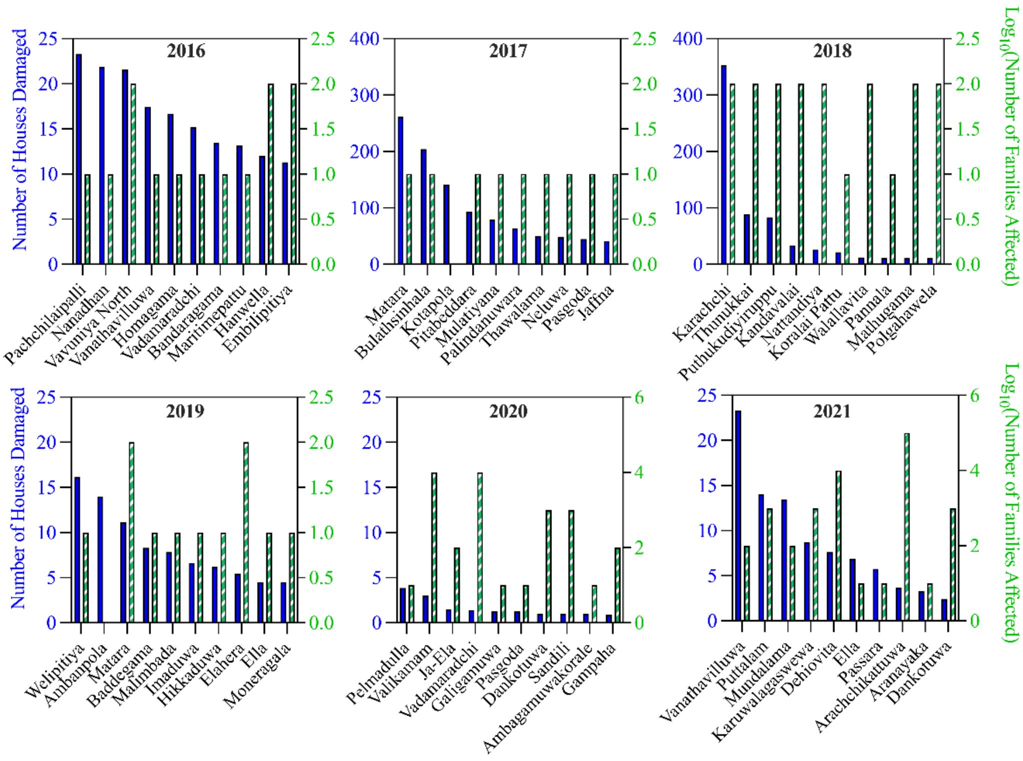

Figure 3. Top ten cities with the largest impact of the flood events from 2016 to 2021.

Figure 4 shows the overlay maps generated for spatial and temporal variation of the hazard incidents for the 2016-2018 period. A high collation of the points can be observed in southwest Sri Lanka (Figure 1a). This conforms to the prior knowledge since Sri Lanka experiences severe adverse impacts from the southwest - monsoon season (May - September). Also, the spatial segregation of the point cloud towards the northern part of Sri Lanka shows the impact of the northeast-monsoon Season (December - February), which generates rapid and extreme hurricane events due to the pressure fluctuations of the Bay of Bengal. The repeated nature of the hazard events along the Kelani valley and the Attanagalu Oya - a river in Gampaha District- (Figure 1b) demonstrates the occurrence of flash flood events near the urban and suburban regions in the western province after the southwest-monsoon season. From the visual analytics, it is clear that the disaster hotspots are geospatially distributed and concentrated near the river valleys (mainly the Kelani valley) and directionally distributed along the southwest to northeast direction.

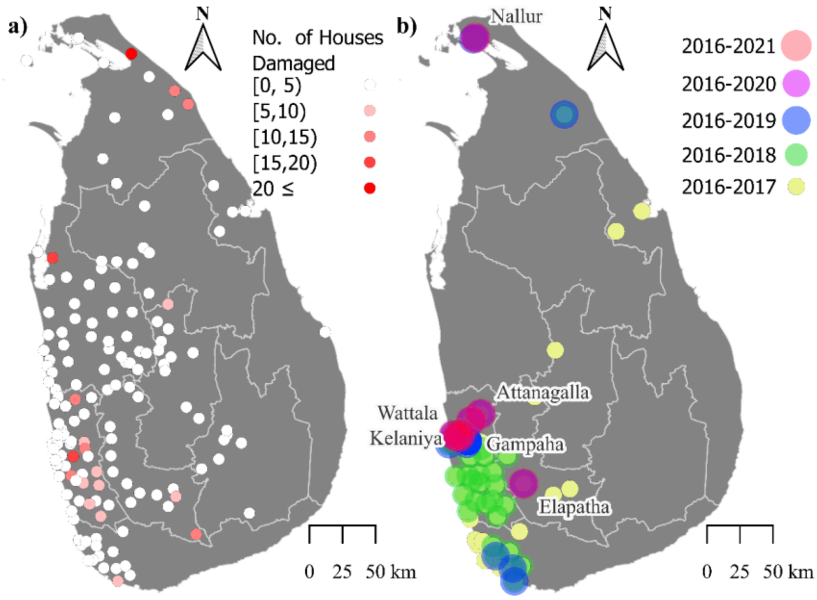

Figure 4. Geospatial distribution of the weather impacts: a) property damages in terms of the houses damaged; b) Repeated occurrences of the flood events from 2016 to 2021

CLOSING REMARKS AND FUTURE WORKS

The fusion of geospatial analytics and textual data analytics can effectively provide deeper insights into the impact of hydrometeorological hazards in urban and suburban regions of the island nation of Sri Lanka. An improved measure for the impact severity of these hazards derived from this analysis can be integrated into the decision-making process. The next stage of this study will classify the impact severity based on a five-scale rank: very low, low, moderate, high, and critical. This would help to compare the magnitude of the hydrometeorological intensity of the weather impact and the necessity of the corrective measures. Such an approach will enable the respective stakeholders and government entities to assess the expected level of intervention they should make during extreme hydrometeorological events. The preliminary results presented in this work-in-progress paper will be extended to visualize and predict the impact and severity of the events. To this end, future studies shall integrate cloud-based machine learning and deep learning approaches.

REFERENCES

Basnayake, A., Jayasinghe, L. and Weerathunga, S., 2019. Disaster management in Sri Lanka: A case study of administrative failures.

Dias, P., Arambepola, N.M.S.I., Weerasinghe, K., Weerasinghe, K.D.N., Wagenaar, D., Bouwer, L.M. and Gehrels, H., 2018. Development of damage functions for flood risk assessment in the city of Colombo (Sri Lanka). *Procedia engineering*, 212, 332-339.

Wagenaar, D.J., Dahm, R.J., Diermanse, F.L.M., Dias, W.P.S., Dissanayake, D.M.S.S., Vajja, H.P., Gehrels, J.C. and Bouwer, L.M. (2019) Evaluating adaptation measures for reducing flood risk: A case study in the city of Colombo, Sri Lanka. *International Journal of Disaster Risk Reduction*, 37, 101162.

Shreve, C.M. and Kelman, I., 2014. Does mitigation save? Reviewing cost-benefit analyses of disaster risk reduction. *International journal of disaster risk reduction*, 10, 213-235.

Ginige, K., Amaratunga, D. and Haigh, R. (2010) Developing capacities for disaster risk reduction in the built environment: capacity analysis in Sri Lanka. *International Journal of Strategic Property Management*, 14, 4, 287-303.

Jayawardena, A.W. (2015) Hydro-meteorological disasters: Causes, effects and mitigation measures with special reference to early warning with data driven approaches of forecasting. *Procedia IUTAM*, 17, 3-12.

Sivakumar, S.S. (2015) Flood mitigation strategies adopted in Sri Lanka a review. *International Journal of Scientific and Engineering Research*, 3, 6.

Weyrich, P., Scolobig, A., Bresch, D.N. and Patt, A. (2018) Effects of impact-based warnings and behavioral recommendations for extreme weather events. *Weather, climate, and society*, 10, 4, 781-796.

Singkran, N. (2017) Flood risk management in Thailand: Shifting from a passive to a progressive paradigm. *International journal of disaster risk reduction*, 25, 92-100.

Potter, S.H., Kreft, P.V., Milojev, P., Noble, C., Montz, B., Dhellemmes, A., Woods, R.J. and Gauden-Ing, S. (2018) The influence of impact-based severe weather warnings on risk perceptions and intended protective actions. *International journal of disaster risk reduction*, 30, 34-43.

Robbins, J.C. and Titley, H.A. (2018) Evaluating high-impact precipitation forecasts from the Met Office Global Hazard Map (GHM) using a global impact database. *Meteorological Applications*, 25, 4, 548-560.

Jayasingh, S.K., Mantri, J.K. and Gahan, P. (2016) Application of Text Mining for Faster Weather Forecasting. *International Journal of Computer Engineering and Information Technology*, 8, 11, 213-219.

Fang, J., Hu, J., Shi, X. and Zhao, L. (2019) Assessing disaster impacts and response using social media data in China: A case study of 2016 Wuhan rainstorm. *International journal of disaster risk reduction*, 34, 275-282.

Dou, M., Wang, Y., Gu, Y., Dong, S., Qiao, M. and Deng, Y. (2021) Disaster damage assessment based on fine-grained topics in social media. *Computers & Geosciences*, 156, 104893.

Roy, K.C. and Hasan, S. (2021) Modeling the dynamics of hurricane evacuation decisions from twitter data: An input output hidden markov modeling approach. *Transportation research part C: emerging technologies*, 123, 102976.

Yeo, J., Comfort, L. and Jung, K. (2018) Timely assessment of disaster and emergency response networks in the aftermath of superstorm Sandy, 2012. *Online Information Review*.

Algiriyage, N., Prasanna, R., Stock, K., Doyle, E.E. and Johnston, D. (2022) Multi-source Multimodal Data and Deep Learning for Disaster Response: A Systematic Review. *SN Computer Science*, 3, 1, 1-29.

Montgomery, M.R. and Pinchoff, J. (2017) Extreme-Event Disasters in Poor Countries: Lessons From Desinventar.

Marulanda, M.C., Cardona, O.D. and Barbat, A.H. (2010) Revealing the socioeconomic impact of small disasters in Colombia using the DesInventar database. *Disasters*, 34, 2, 552-570.

Edirisinghe, M. and Maduranga, U.G.D. (2021) Distribution of lightning accidents in Sri Lanka from 1974 to 2019 using the DesInventar database. *ISPRS International Journal of Geo-Information*, 10, 3, 117.