Multi-hazard Risk Assessment by Integrating Machine Learning and GIS

Surassawadee Phoompanich [*1], Stuart Barr [†1] and Rachel Gaulton [‡1]

[1]School of Engineering, Newcastle University

January 30, 2019

Summary

The frequent hydro-meteorological hazards experienced by Thailand cause significant infrastructure impact and enormous economic and human cost. Existing approaches for individual and multiple hazard assessment are limited due to uncertainties associated with input variables and the fact that often there is only limited knowledge of the spatiotemporal interactions of hazards. To address these issues we employ a GIS-based Naïve Bayes and Bayesian Network approach for computing the probability of a hazard and the spatiotemporal joint probability of hazards and subsequent impacts. An integrated spatiotemporal risk map is generated by considering vulnerability, capacity and exposure to identify risk areas in order to support the national policies on reduction of risk from natural disasters. The results revealed that the pattern of individual and multi-hazard risks is dependent primarily on terrain characteristics and seasonality. Low lying areas were at-risk from floods and droughts while mountainous areas are exposed to floods, which may potentially trigger landslides in the rainy season, and are also prone to forest fires in summer season. The results highlighted that the greatest multi-hazard risk of floods and droughts was in low lying areas of Thailand with an area of 327,956 km^2, or 63% of the total area.

KEYWORDS: Machine learning, Naïve Bayes, Bayesian Network, Multi-hazard risk assessment

1. Introduction

Natural hazards are inevitable and have caused considerable damage to people and their surroundings. Mitigation of hazardous impacts by identifying risk areas was mentioned as important in the Twelfth National and Social Development Plan (2017-2021) of Thailand (NESDB, 2017). Investigation of risk areas exposed to either individual or multiple hazards, and the relationships between hazards and vulnerabilities is required to support this plan.

The existing risk assessment approaches often used in Thailand are not frequently updated and are difficult to implement. The uncertainties from input variables and subjective evaluation by expert appraisals usually affect the results. Machine learning, a subset of artificial intelligence, has been shown to be a potentially promising approach for addressing complex natural hazard problems (Tehrany *et al.*, 2014; Vogel *et al.*, 2014). Machine learning algorithms, namely Naïve Bayes (NB) and Bayesian Network (BN) based on Bayes' theorem rely on prior knowledge and learning from known input data for predicting risk. These algorithms have been used to assess floods (Liu *et al.*, 2017), landslides (Tsangaratos and Ilia, 2016), and forest fires (Dlamini, 2010; Zwirglmaier *et al.*, 2013). However, their use for assessing the spatiotemporal risk associated with multiple hazards is less well developed. In this paper we present the results of a study to employ NB and BN to produce the first fine-spatial scale multi-hazard risk assessment for Thailand.

While several open-source packages for NB and BN implementation exist, such as Scikit-learn, Bayes Net Toolbox for Matlab and Bayespy, they are unable to handle large volumes of geospatial and Earth Observed data required for a national scale spatiotemporal risk assessment. Therefore, this study

[*] s.phoompanich2@newcastle.ac.uk
[†] stuart.barr@newcastle.ac.uk
[‡] rachel.gaulton@newcastle.ac.uk

developed an integrated Python-based GIS approach where NB and the BN was employed for an assessment of spatial probability of individual and multiple hazards. A directed acyclic graph (DAG) that represents the relations between hazards and vulnerabilities was constructed and the conditional probability table (CPT) of a node indicating probabilities of the nodes in the DAG was analysed to describe the causalities between hazards and vulnerabilities. The results were then analysed together with vulnerability, capacity and exposure assessment to generate an integrated spatiotemporal risk assessment for Thailand.

2. Data sets and methods

2.1 Data sets

Data used in this study were comprised of geospatial data and remote sensing images. Past hazard events including flood inundation (GISTDA, 2016), landslide scars and flash flood-prone area (DMR, 2011; DMR, 2015; DMR, 2016), drought events derived from a combined satellite index of Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Normalized Difference Water Index (NDWI), Normalized Difference Moisture Index (NDMI), and Normalized Difference Drought Index (NDDI), and monthly burnt areas (LPDAAC, 2017) were used to indicate the presence of a hazard. Geospatial data of geology, soil drainage, streams and rivers, groundwater, roads, irrigation zone, land use, relative humidity, including precipitation from GSMaP, SRTM DEM 90m., and land surface temperature derived from MODIS as potential contributing factors were compiled. Attributes associated with socio-economic (DOPA, 2017; NSO, 2016), physical (NSO, 2012), and environmental (RFD, 2016; RFD, 2017) vulnerabilities, and capacity based on preparation for disaster response of local communities and authorities (DDPM, 2018), and emergency services for hazard response (MOPH, 2016; NIEM, 2018) were generated at the province administrative level for vulnerability and capacity assessment. Satellite images of the DMSP-OLS night-time light data (NGDC, 2018), LandScan 2017 dataset (ORNL, 2017), global GDP (Ghosh *et al.*, 2010), and assets and public infrastructure, including agricultural and forest lands, were used for exposure assessment. All data were manipulated into a geodatabase, with a project coordinate system, WGS84/UTM zone 47N.

2.2 Methods

The conceptual framework of the study is presented in Figure 1. As the NB and the BN models available were for discrete variables, continuous data were converted to categorical data, while for certain discrete data re-quantisation (binning) was undertaken to provide categorical groups. Sample points were used to extract multiple values from multiple rasters of past hazard events and their potential trigger variables. Extracted values were used to compute the probability of occurrence of an individual hazard on the basis of a trigger variable was derived using Bayes' theorem (Equation 1).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1}$$

where, $P(A)$ is the prior probability of the event A, $P(B)$ is prior probability of the event B, $P(B|A)$ is the probability of B given A, and $P(A|B)$ is the probability of A given B

For the multi-hazard assessment, the DAG demonstrating the qualitative relations between hazards and vulnerabilities was developed through expert knowledge by intensive and extensive one-to-one interviews with 24 experts in Thailand and from an analysis of past events. The joint probability distribution was calculated using Equation 2.

$$P(X) = \prod_{i=1}^{n} P(X_i|C_i) \tag{2}$$

where, $P(X)$ represents the joint probability distribution (JPD) of the nodes in the DAG, X_i presents both the potential variable and its corresponding node, C_i are the parents of X_i, and n represents hazards and their potential factors in the BN

69 social, economic, physical, and environment vulnerability factors and 21 capacity indicators, indicating the strength or resource available to adapt or exhibit robustness to a particular vulnerability were considered. These were normalized and combined to produce a composite index map of vulnerability and capacity. All spatial hazard assessments were analysed with regards to vulnerability, capacity, and exposure in order to generate an integrated spatiotemporal risk assessment.

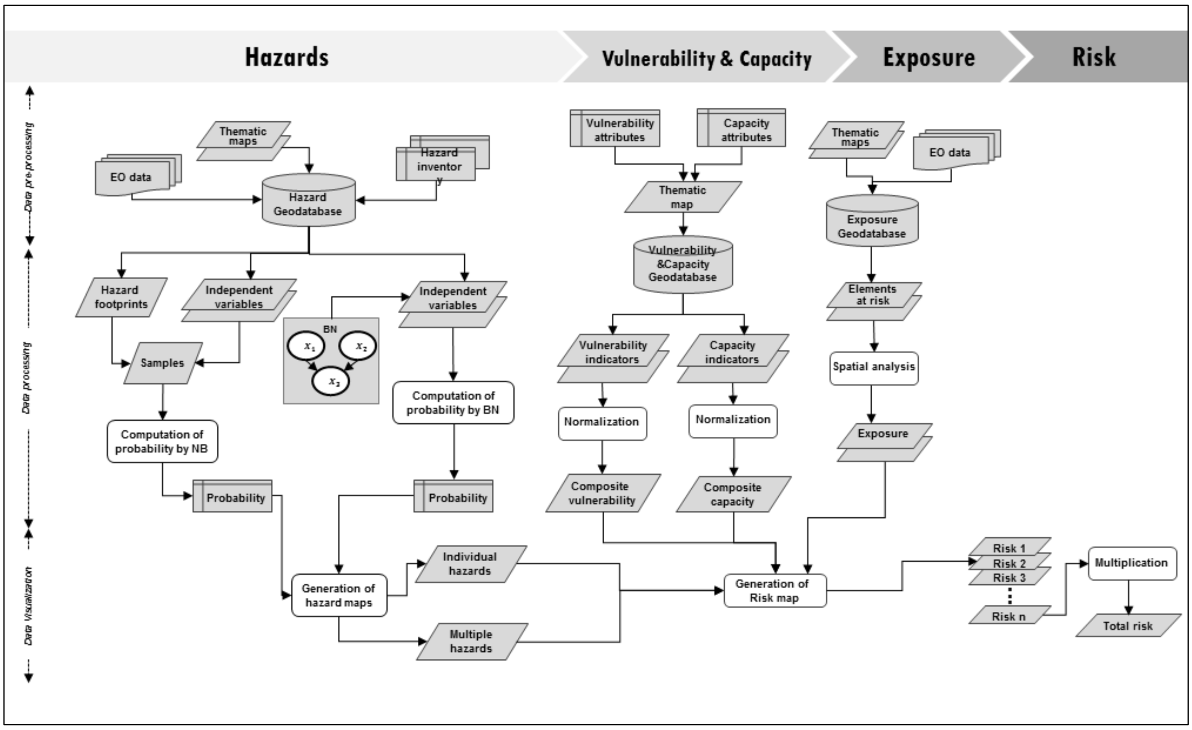

Figure 1 Conceptual framework for a national scale multi-hazard risk assessment

3. Results and Discussion

Figure 2 presents the composite vulnerability index (A), the composite capacity index (B), and the spatiotemporal individual risk assessment (C). The highest overall vulnerability was in Bangkok (100%), followed by Nakhon Ratchasima (82%), and Chiang Mai (78%) (Table 1). These provinces are the main cities and the business centres of each region with overcrowded population (e.g., the density of population and housing unit in Bangkok were 3,624 and 1,841 per sq.km) and public vulnerable infrastructure. The highest capacity was found for Nakhon Ratchasima (100%), Chiang Mai (85%), and Ubon Ratchathani (81%), respectively. These provinces had a high number of local trainees on emergency response, and large emergency and medical resources for disaster response i.e. volunteers, health personnel, hospitals and ambulances. Based on the past hazard events and their potential factors, the high-risk areas with over 20% annual probability of floods and droughts were in lowland areas with an area of approximately 76,089 and 293,456 km^2, respectively. Landslides and forest fires with probability of over 30% covered an area of 129,989 and 150,610 km^2, and occurred high mountain and the plateaus in northern and western regions. The pattern of individual hazard risk is dependent on terrain characteristics and a seasonality. Low land areas were at-risk to floods during the rainy season and droughts in summer, while high mountains exposed to floods might trigger landslides in rainy season, and forest fires usually occurred in summer.

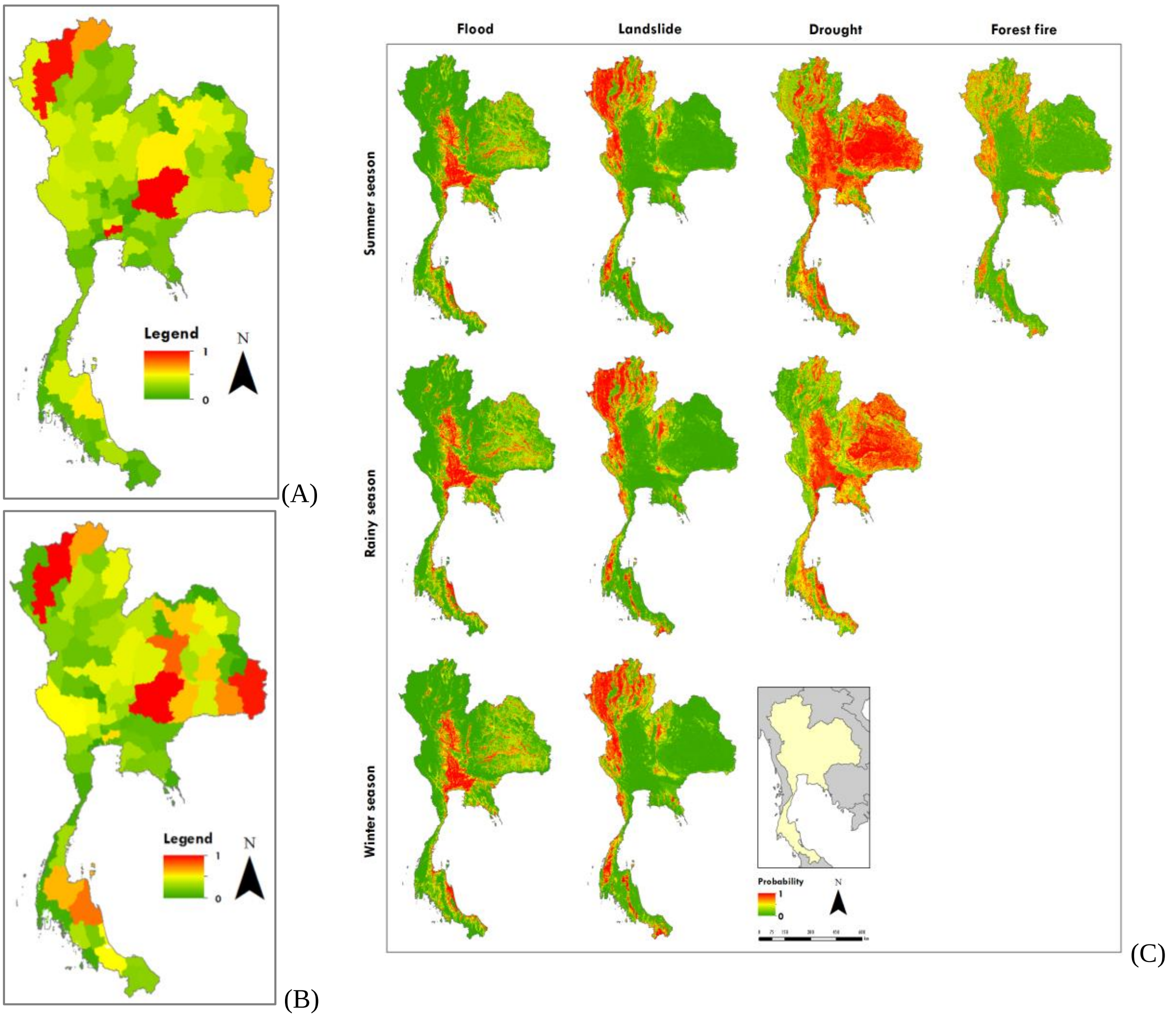

Figure 2 Maps of composite vulnerability index (A), composite capacity index (B), and individual hazard risk assessment (C) that present the value ranging between 0 (0%) and 1 (100%)

Table 1 The 10 highest vulnerability and capacity provinces.

No.	Provinces	Vulnerability	No.	Provinces	Capacity
1	Bangkok	1.000000	1	Nakhon Ratchasima	1.000000
2	Nakhon Ratchasima	0.815931	2	Chiang Mai	0.851867
3	Chiang Mai	0.788123	3	Ubon Ratchathani	0.810381
4	Chiang Rai	0.576008	4	Khon Kaen	0.698894
5	Ubon Ratchathani	0.499628	5	Nakhon Si Thammarat	0.669200
6	Nakhon Si Thammarat	0.463167	6	Sisaket	0.626519
7	Khon Kaen	0.459412	7	Chiang Rai	0.593365
8	Chaiyaphum	0.456699	8	Surat Thani	0.568373
9	Udon Thani	0.454883	9	Burirum	0.545904
10	Phitsanulok	0.395242	10	Udon Thani	0.544205

Figure 3 shows the spatiotemporal multi-hazard risk assessment and Table 2 shows the areas of multi-hazard risk in different periods. Most of the forest fire and drought risk areas over 10% probability were in northeastern and central regions, and some parts of mountains in the north and west of Thailand, covering an area of 116,577 km^2. Some areas in the southern region were also at-risk to multi-hazards in the wet period, covering 23,310 km^2. The transitional period showing the causal relations among forest fires, floods, and landslides denoted the same risk pattern as the wet period, except the north of Thailand, covering approximately 123,476 km^2. The overall multi-hazard risk representing the relationships of four hazards identified that approximately 95,452 km^2 of the country had a multi-hazard risk over 10% annual probability. This probability represented low spatiotemporal correlation among four hazards because of the climatic conditions and topographical characteristics. Between October and May, the southern region was influenced by the monsoons and tropical cyclones which contribute abundant rainfall over the region. This resulted in high humidity and high soil moisture that might cause floods and landslides, so human-caused fires were rare due to the difficulty of setting a fire in tropical evergreen forest or rain forest. Relation between forest fires and droughts was therefore low.

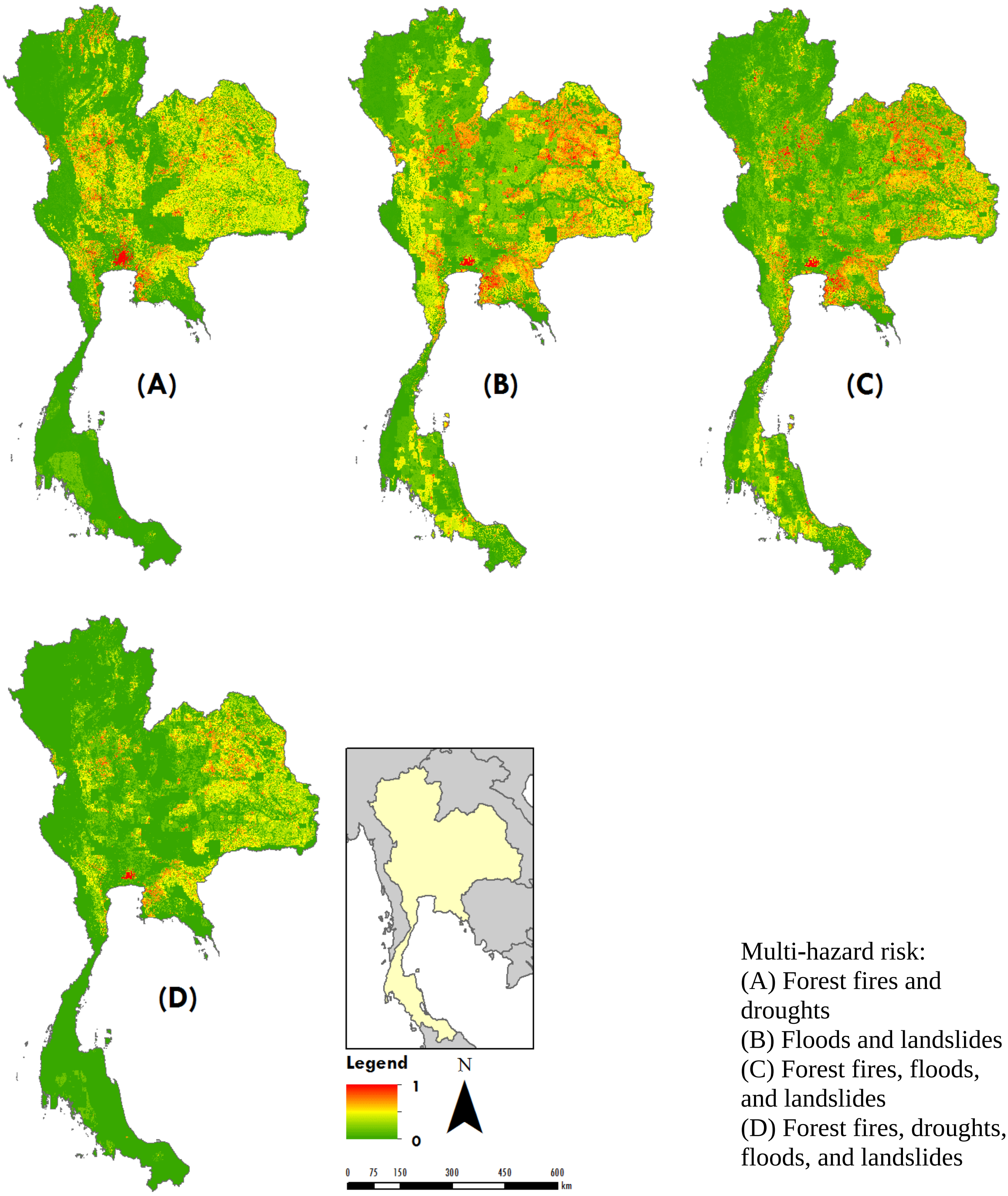

Figure 3 A map of spatiotemporal multi-hazard risk assessment presenting the probability of risk ranging between 0-100%

Table 2 The probability of multi-hazard risk assessment in different seasons

Probability	Multi-hazard risk areas (unit: sq.km.)			
	Forest fires & Droughts	**Floods & Landslides**	**Forest fires, floods, and landslides**	**Forest fires, droughts, floods, and landslides**
0%	1,641	1,655	1,651	1,642
0-10%	345,361	322,323	387,618	414,127
11-20%	116,577	100,902	55,174	36,985
21-30%	44,358	83,071	63,648	54,553
31-40%	2,308	4,192	3,841	3,199
41-50%	478	449	434	358
51-60%	265	174	174	165
61-70%	170	146	146	133
71-80%	55	50	50	50
81-90%	7	7	7	7
91-100%	2	2	2	2

4. Conclusions

This study has developed an integrated machine learning algorithm and a GIS-based approach for spatiotemporal multi-hazard risk assessment. Four natural hazards were investigated by Naïve Bayes while multi-hazard assessment and their causalities were analysed via a Bayesian Network. Learning from the past hazard events coupled with expert knowledge on hazards was employed. A large number of vulnerability- and capacity-based indicators were applied to determine the potential for individuals and communities to be harmed by hazards. Human and physical exposures were analysed by a GIS-based approach and the different elements exposed to different hazards were mapped. A probabilistic risk assessment, including vulnerability, capacity and exposure, was mapped.

The results revealed that the pattern of individual and multi-hazard risks is dependent on terrain characteristics and seasonality. Most of the low lands were at-risk to floods in rainy and droughts in summer seasons. In contrast, high mountains were occasionally exposed to floods that might trigger landslides in the rainy season and forest fires in summer season. The quantitative risk ranges between 0 (0%) and 1 (100%), presenting low and high risk probability instead of the qualitative risk (e.g. low, moderate, high). The spatiotemporal interactions of hazards were demonstrated in both qualitative structure of a directed acyclic graph and quantitative probability. Analysed probability of the multi-hazard risk assessment provide effective information on the spatial and seasonal pattern of multi-hazard risk. The potential factors to be considered for multi-hazard risk assessment were climatic and socio-economic conditions while contributing factors of topography, physical, and environment were a considerable increase of risk. Due to differences in the spatial and temporal pattern of the hazards, areas at-risk to multiple hazards are divided into two zones. Floods and droughts with approximately 382,361 km^2 were mostly in low lying areas while forest fires and landslides with 129,989 km^2 were found to have significantly stronger correlation in high mountains. This analysis provided understanding of the nature of hazard phenomena and the relationship between hazards and vulnerabilities in Thailand. These results were used to generate a full scale of natural hazard risk assessment framework to support the national policies on risk reduction and mitigation in Thailand.

References

DDPM (2018) *Number of Community-based Disaster Risk Management (CBDRM), One Tambon One Search and Rescue Team (OTOS), Mr. Warning System, and Civil Defence Volunteer*. Bangkok, Thailand: Department of Disaster Prevention and Mitigation (Accessed: 12 December 2018).

Dlamini, W.M. (2010) 'A Bayesian belief network analysis of factors influencing wildfire occurrence in Swaziland', *Environmental Modelling & Software,* 25(2), pp. 199-208.

DMR (2011) *A local scale landslide susceptibility mapping: Uttaradit province*. [Online]. Available at: http://www.dmr.go.th/ (Accessed: 29 November 2016).

DMR (2015) *Landslide Event Records*. Available at: http://www.dmr.go.th/ (Accessed: 23 December 2015).

DMR (2016) *A local scale landslide susceptibility mapping: Phetchabun province*. [Online]. Available at: http://www.dmr.go.th/ (Accessed: 28 November 2016).

DOPA (2017) *Statistics of Demographic Data*. Available at: http://stat.bora.dopa.go.th/ (Accessed: 3 March 2017).

Ghosh, T., Elvidge, C., Sutton, P.C., Baugh, K.E., Powell, R. and Anderson, S. (2010) 'Shedding light on the global distribution of economic activity'.

GISTDA (2016) *Thailand Flood Monitoring System*. Available at: http://flood.gistda.or.th/ (Accessed: 22 May 2016).

Liu, R., Chen, Y., Wu, J., Gao, L., Barrett, D., Xu, T., Li, X., Li, L., Huang, C. and Yu, J. (2017) 'Integrating Entropy-Based Naïve Bayes and GIS for Spatial Evaluation of Flood Hazard', *Risk analysis*, 37(4), pp. 756-773.

LPDAAC, N. (2017) 'Combined MODIS. Version 6. '. 6 April 2018. NASA EOSDIS Land Processes DAAC, USGS Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota. Available at: https://lpdaac.usgs.gov/ (Accessed: 6 April 2018).

MOPH (2016) *Report on public health resource 2016*. Bangkok, Thailand: Ministry of Public Health (Accessed: 7 November 2018).

NESDB (2017) *The Twelfth The National Economic and Social Development Plan (2017-2021)*. [Online]. Available at: http://www.nesdb.go.th/ (Accessed: 22 November 2018).

NGDC (2018) 'Version 4 DMSP-OLS Nighttime Lights Time Series' Center, N.G.D. Available at: http://www.ngdc.noaa.gov/ (Accessed: 5 October 2018).

NIEM (2018) *Number of worker, operation, and emergency units in Emergency Medicine System* (Accessed: 19 November 2018).

NSO (2012) *The 2010 Population and Housing Census: Province level*. Bangkok, Thailand.

NSO (2016) *Statistical data on population & society*. Available at: http://web.nso.go.th/ (Accessed: 24 December 2017).

ORNL (2017) 'The LandScan 2017™ High Resolution global population' Laboratory, O.R.N. Available at: https://landscan.ornl.gov/ (Accessed: 10 November 2018).

RFD (2016) *Forest Fires in Thailand*. Available at: http://www.forest.go.th/ (Accessed: 29 January 2016).

RFD (2017) *Forest land assessment by province, and forest fire frequency and their burnt areas*. Available at: http://www.forest.go.th/ (Accessed: 29 May 2017).

Tehrany, M.S., Pradhan, B. and Jebur, M.N. (2014) 'Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS', *Journal of Hydrology*, 512, pp. 332-343.

Tsangaratos, P. and Ilia, I. (2016) 'Comparison of a logistic regression and Naïve Bayes classifier in landslide susceptibility assessments: The influence of models complexity and training dataset size', *Catena*, 145, pp. 164-179.

Vogel, K., Riggelsen, C., Korup, O. and Scherbaum, F. (2014) 'Bayesian network learning for natural hazard analyses', *Natural Hazards and Earth System Sciences*, 14(9), pp. 2605-2626.

Zwirglmaier, K., Papakosta, P. and Straub, D. (2013) *International Conference on Structural Safety & Reliability in 2013 (ICOSSAR 2013)*. New York, United States, June 16-20, 2013.

Biographies

Surassawadee Phoompanich is a Thailand government funded PhD student in the School of Engineering at Newcastle University. Since 2003 she has worked at Geo-Informatics and Space Technology Development Agency (GISTDA), Thailand as a Geo-informatics scientist. Her research interest is the use of geospatial engineering for spatiotemporal multi-hazard risk assessment. Stuart Barr is Professor of Geospatial Systems Engineering in the School of Engineering at Newcastle University. Rachel Gaulton is a lecturer in Remote Sensing in the School of Engineering at Newcastle University.