ORIGINAL ARTICLE

CIWEM Chartered Institution of Water and Environmental Management   Journal of Flood Risk Management   WILEY

Integration of catastrophe and entropy theories for flood risk mapping in peninsular Malaysia

Ghaith Falah Ziarh[1] | Md Asaduzzaman[2] | Ashraf Dewan[3] |
Mohamed Salem Nashwan[4] 🆔 | Shamsuddin Shahid[1] 🆔

[1]Department of Water and Environmental Engineering, School of Civil Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, Johor Bahru, Malaysia

[2]Department of Engineering, School of Digital, Technologies and Arts, Staffordshire University, Stoke-on-Trent, UK

[3]School of Earth and Planetary Sciences, Curtin University, Bentley, Perth, Australia

[4]Department of Construction and Building Engineering, College of Engineering and Technology, Arab Academy for Science, Technology and Maritime Transport (AASTMT), Cairo, Egypt

Correspondence
Shamsuddin Shahid, Department of Water and Environmental Engineering, School of Civil Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, Johor Bahru, Malaysia.
Email: sshahid@utm.my

Funding information
Staffordshire University, Grant/Award Number: QR-GCRF

Abstract

A major challenge in flood mapping using multi-criteria decision analysis (MCDA) is the selection of the flood risk factors and the estimation of their relative importance. A novel MCDA method through the integration of two state-of-the-art MCDA methods based on catastrophe and entropy theory is proposed for mapping flood risk in the Peninsular Malaysia, an area very susceptible to flooding events, is presented. A literature review was undertaken which identified the various socioeconomic, physical and environmental factors which can influence flood vulnerability and risk. A set of variables was selected using an importance index which was developed based on a questionnaire survey. Population density, percentage of vulnerable people, household income, local economy, percentage of foreign nationals, elevation and forest cover were all deemed highly relevant in mapping flood risk and determining the zones of maximum vulnerability. Spatial integration of factors using the proposed MCDA revealed that coastal regions are highly vulnerable to floods when compared to inland locations. Flood risk maps indicate that the northeastern coastal region of Malaysia is at greatest risk of flooding. The prediction capability of the integrated method was found to be 0.93, which suggests good accuracy of the proposed method in flood risk mapping.

KEYWORDS
catastrophe theory, composite vulnerability, entropy theory, flood risk, importance index, multi-criteria decision-making analysis

1 | INTRODUCTION

Extreme rainfall-driven events such as flood or droughts have increased in severity and frequency in many regions as a result (Nashwan, Ismail, & Ahmed, 2019a; Sediqi et al., 2019). In many parts of the world, floods have had devastating impacts in terms of loss of life and property (Dewan, 2013a; Nashwan, Shahid, & Wang, 2019b; Pour,

Bin Harun, & Shahid, 2014; Yaseen et al., 2019). To ameliorate the risks associated with flooding events on the development of a region, it is important to define the spatial distribution of at-risk locations, particularly in the context of ongoing climate change (Pour, Wahab, Shahid, Asaduzzaman, & Dewan, 2020; Santos & Reis, 2018).

Tropical regions tend to be more susceptible to changes in climate (Noor, Ismail, Shahid, Nashwan, &

Ullah, 2019; Rahman et al., 2019; Shahid et al., 2017), and thus, more susceptible to climate variability (Mishra & Liu, 2014). Recent studies have reported the adverse consequences on the societies and economies of tropical regions arising from the increasing frequency and severity of weather extremes (Noor et al., 2019; Noor, Ismail, Chung, Shahid, & Sung, 2018; Sa'adi, shahid, Ismail, Chung, & Wang, 2017; Shahid et al., 2016; Wong, Yusop, & Ismail, 2018). Malaysia, located in the tropics, is one country which has experienced climate and hydrological extremes in recent years (Khan et al., 2019; Mayowa et al., 2015; Sa'adi et al., 2017). The impacts of extreme rainfall and monsoonal rain-driven floods are increasingly evident in this region (Nashwan, Ismail, & Ahmed, 2018a). The flood in December 2014, affected thousands of people (Shahid et al., 2017) and resulted in huge economic loss to the country. As a result of these type of events, the development of flood management processes has been proposed to mitigate the negative impacts on people and the economy (Salarpour, Yusop, Yusof, Shahid, & Jajarmizadeh, 2013).

Extreme rainfall is generally considered to be the major driver of flooding in Malaysia. Various physical and socioeconomic factors, however, can also amplify the impact of these flood events. Alias et al. (2019) identified forest cover, elevation, and population density as having a great influence on the spatial variability of flood impacts. In assessing the flood risk of a region, it is important that physical and social factors be considered (Rahman et al., 2019).

Numerous studies on flood vulnerability and risk mapping have been conducted in recent years (Chen et al., 2014; Dano et al., 2019; Dewan & Yamaguchi, 2008; Elsheikh, Ouerghi, & Elhag, 2015; Feloni, Mousadis, & Baltas, 2019; Jato-Espino, Lobo, & Ascorbe-Salcedo, 2019; Matori, Lawal, Yusof, Hashim, & Balogun, 2014; Nigussie & Altunkaynak, 2019; Pradhan & Youssef, 2011). In general, different factors can be considered in a multi-criteria decision analysis (MCDA) system. The major challenges in developing flood risk maps using MCDA are involved in the selection of indicators and the weighting of the factors according to their importance in defining flood risk. A large number of physical, environmental and socio-economic factors are typically responsible for shaping the vulnerability of an area (Cutter et al., 2008; Dewan, 2013a; Dewan, 2013b; Rahman et al., 2019). Many criteria for the selection of indicators have been proposed in the literature, including their availability, measurability, practicality, relevance, and degree of responsiveness and sensitivity (Alamgir et al., 2019; Nashwan, Shahid, Chung, Ahmed, & Song, 2018b; Yli-Viikari et al., 2007). The selection criteria for indicators should be based on the specific study area characteristics and the research questions to be solved.

Various knowledge-based and data-driven MCDA methods have been proposed for the mapping of risk associated with natural hazards such as flooding (Dewan, 2013a; Dewan, 2013b). In a knowledge-driven method, the perceived influence of factors on flood susceptibility is based on the opinion of the decision-makers. Therefore the outcomes of a knowledge-based MCDA are always prone to be biased due to personal preferences (Nashwan et al., 2018b, Ahmed et al., 2015). This limits their applicability in many cases, particularly in regards to risk mitigation decision-making. The data-driven method attempts to overcome this drawback by assigning weightings to the different factors based on the properties of data itself (Ahmed, Shahid, & Nawaz, 2018; Alamgir et al., 2019). For this reason, the data-driven MCDA approach is often preferred for flood risk mapping.

Catastrophe and entropy theories are two such MCDA methods which have been found to be highly effective in the modelling and mapping of different natural hazards (Agarwal et al., 2016; Ahmed et al., 2015; He-Hua, Jian-Jun, Xiao-Yan, & Ye, 2018; Nashwan & Shahid, 2019b; Singh, Jha, & Chowdary, 2018; Zhou, Ma, Chen, Wu, & Luo, 2018). Catastrophe theory was developed to characterise discontinuous dynamic systems where changes are abrupt. It is a subjective method that estimates factors' importance based on internal structure of different factors, and thus assists in avoiding human bias in the decision-making process. Entropy is a measure of uncertainty of a random variable. It can be used to evaluate how the controlling factors influence the outcome; e.g., how different socio-economic factors govern flood susceptibility. Therefore, it can also be used in a similar manner to catastrophe theory in assigning weights without the input of expert opinion. Both methods, however, have inherent advantages and disadvantages when determining the indicators. The weight assigned to an indicator by catastrophe theory is often influenced by the number of groups into which the indicator has been classified (Cui & Singh, 2015). Therefore, indicator weights are partly influenced by human judgement (Al-Abadi, Shahid, & Al-Ali, 2016). This shortcoming can be avoided using an entropy-based weighting method (Castillo, Castelli, & Entekhabi, 2015; Tang & Wang, 2013). Integration of these two theories can provide some robustness to the weighting approach when assessing the risks associated with flood events.

A data-driven MCDA approach, integrating both catastrophe and entropy theories, is proposed in this study in order to provide an unbiased evaluation of the spatial pattern of flood risk in Peninsular Malaysia. The study has considered flood risk as a system - consisting of

different subsystems which can be evaluated using indicators. Entropy theory was used to assign the ranks of the indicators of the sub-systems while the catastrophe theory was used to assign weights to the different subsystems. The methodology proposed in this study can be employed for systematic evaluation of flood vulnerability factors, and to evaluate natural hazard risk in other regions.

2 | METHODS AND MATERIALS

2.1 | Geography and climate of peninsular Malaysia

Peninsular Malaysia (latitude 1.20° - 6.40°N; longitude 99.35° - 104.20°E) encompasses a land area of 130,598 km^2 (Figure 1). The topography consists of an irregular, inland mountainous region surrounded by shorelines, notably around the Peninsula. It is situated within a tropical climatic zone with year-round high temperature and humidity. The daily average temperature in the Peninsula varies between 21 and 32° C. Rainfall of the region is controlled by the interaction between two monsoonal systems and the heterogeneous land and sea surfaces. Most rainfall occurs during the two monsoonal seasons, the northeast monsoon between November and February, and the southwest monsoon between May and August (Muhammad et al., 2019). The northeast monsoon is the more intense of the two systems. Extreme rainfall events can often occur in consecutive days, leading to severe flooding, particularly in the west of Peninsular Malaysia (Nashwan, Ismail, & Ahmed, 2019a).

2.2 | Geospatial data

Secondary data was obtained and used in this study. District level socio-economic data of Peninsular Malaysia was collected from the Statistical Yearbooks of Malaysia (DOSM, 2018). The flood hazard map prepared by the Department of Irrigation and Drainage (DID) in 2016 using long-term historical inundation data, was collected and digitised. Maps of various physical factors related to flood vulnerability (such as forest cover) were generated from existing land use maps of 2018. An elevation map
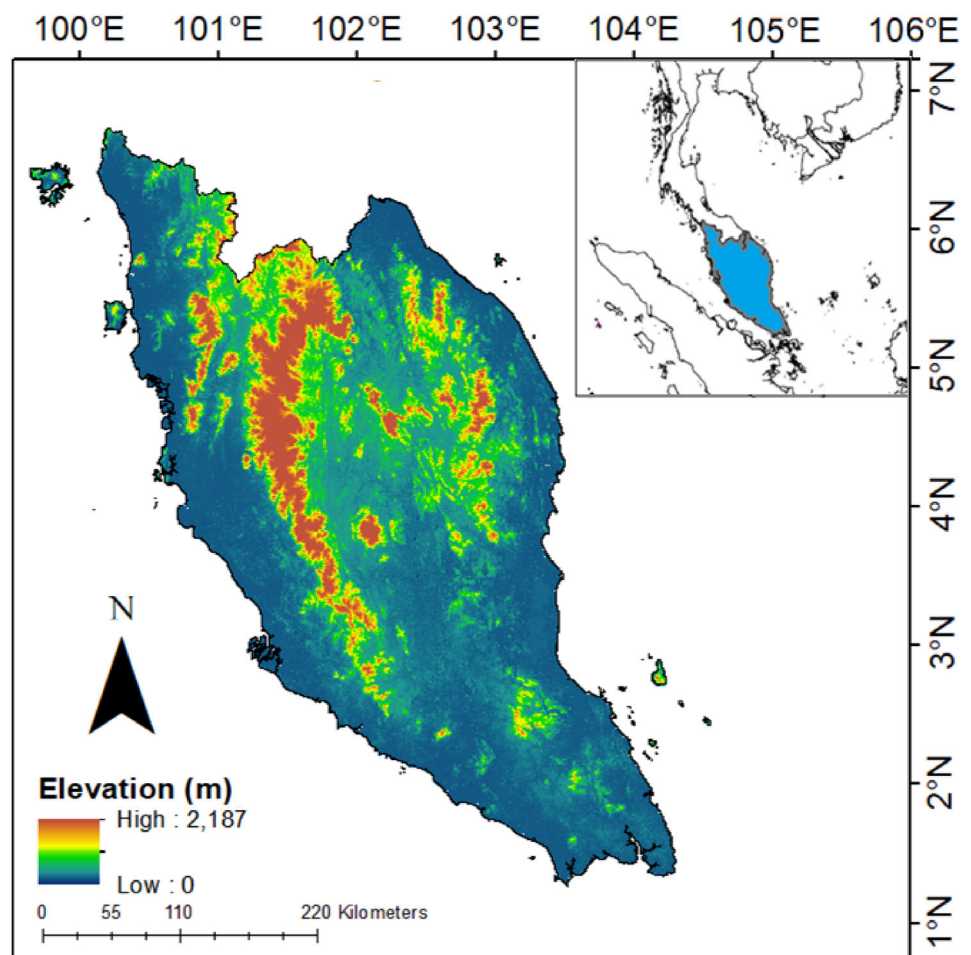

FIGURE 1 The geographical location and topography of peninsular Malaysia

was produced from an Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) digital elevation model (DEM) (https://earthexplorer.usgs.gov/).

3 | METHODOLOGY

3.1 | Importance index for identification of flood vulnerability factors

Flood vulnerability factors were first identified through a literature survey. The identified factors were carefully checked, and those found relevant to the study area were selected. A questionnaire survey was conducted among the various stakeholders in order to rank the factors according to their importance in defining flood vulnerability in Peninsular Malaysia. A non-random judgmental procedure was followed to select the samples from academics, disaster management experts, local councillors and people involved in emergency management (e.g., relief and rescue operations). The judgmental sampling was conducted as only a limited number of people possess knowledge on different factors, affecting flood vulnerability. A total of 50 samples was selected. Attention was given to maintain homogeneity in different groups. The age of the respondent was between 36 and 54 years with a median age of 44 years. A structured questionnaire was distributed to selected samples. All of them voluntarily participated in the survey and returned their responses within a week. Respondents were asked to rank the factors on a scale of 1 to 3. A wider scale (e.g., 1 to 5) does not significantly change the ranking of factor, but does have the potential to confuse the respondents, so a scale of 1 to 3 was used. The responses were used to rank the factors using the Importance Index (Lim & Alum, 1995) as defined below:

$$\text{Importance Index} = \frac{3n_1 + 2n_2 + 1n_3}{3(n_1 + n_2 + n_3)} \quad (1)$$

where n_1, n_2 and n_3 are the total responses of 1, 2 and 3 provided by the respondents during the questionnaire survey. The values of the importance index range from 0 to 1, where 1 indicates highly important, and 0 indicates no importance.

The factors are usually ranked according to the importance index, and the top-ranked factors are considered to be the most useful factors to include in the development of flood vulnerability maps. Usually, the first few top-ranked factors are selected, but the determination of these tends to be subjective, and can be biased to a modeller's preference. This potential lack of objectivity can be overcome through the use of a data classification method known as Jenks optimization (Jenks, 1967). This method uses an importance index to classify derived values according to their variance. The classification is derived in such a way that variance in the importance index within a class is minimised, but among the classes is maximised. Factors that were ranked top by the Jenks optimization tool were considered for inclusion in the flood vulnerability mapping work.

3.2 | Integration of catastrophe and entropy theories

Using catastrophe theory, flood vulnerability is considered to comprise of several subsystems, each of which can be evaluated based on one or more criteria or indicators. The values of all indicators are first normalised between 0 and 1, with 1 indicating high vulnerability to flooding (e.g., high population density indicates high vulnerability to flood). Equation (2) is used for low vulnerability (e.g., more forest cover indicates less vulnerability to flooding) and Equation (3) is employed for high vulnerability. The normalisation formulae (Wang, Liu, Zhang, & Chen, 2011) are given by:

$$Y = \begin{cases} 1 & 0 \leq X \leq a_1 \\ (a_2 - X)/(a_2 - a_1) & a_1 < X < a_2 \\ 0 & X \geq a_2 \end{cases} \quad (2)$$

$$Y = \begin{cases} 1 & X \geq a_1 \\ (X - a_1)/(a_2 - a_1) & a_1 < X < a_2 \\ 0 & 0 \leq X \leq a_1 \end{cases} \quad (3)$$

where X is the indicator, a_1 and a_2 are minimum and maximum values of the indicator. The catastrophe fuzzy membership functions are then used to assign ranking to each indicator. This helps in removing incompatibility issues between the initial indicator values (Ahmed et al., 2015; Wang et al., 2011). There are seven catastrophe models that can be used for the estimation of catastrophe fuzzy membership functions depending on the number of indicators of a subsystem. The catastrophe models and the formula used for estimation of membership function or rating of each indicator are shown in Table 1 in which a represents state variable and u, v, w, x are control variables. The state variable is related to control variables based on different catastrophe models. a_i represents catastrophe fuzzy membership function of the control variable, i, where i can be u, v, w or x depending on the model. Details of the models and estimation of catastrophe fuzzy membership functions can be found in Wang et al. (2012).

TABLE 1 The catastrophe models and the formula used for the estimation of fuzzy membership functions

Catastrophe model	Control variable	Normalisation formula
Cusp	2	$a_u = u^{0.5}$ and $a_v = v^{0.33}$
Swallowtail	3	$a_u = u^{0.5}$, $a_v = v^{0.33}$ and $a_w = w^{0.25}$
Butterfly	4	$a_u = u^{0.5}$, $a_v = v^{0.33}$, $a_w = w^{0.25}$ and $a_x = x^{0.20}$
Wigwam	5	$a_u = u^{0.5}$, $a_v = v^{0.33}$, $a_w = w^{0.25}$, $a_x = x^{0.20}$ and $a_y = y^{0.17}$

The entropy method is used to estimate the weight of each subsystem. If the number of a subsystem is m, and the number of indicators of a subsystem is n, the matrix of Eigenvalue, Y can be estimated using the normalised values of the indicators as (Chen & Li, 2010):

$$Y = \begin{bmatrix} y_{11} & y_{21} & \cdot & y_{1n} \\ y_{21} & y_{22} & \cdot & y_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ y_{m1} & y_{m2} & \cdot & y_{mn} \end{bmatrix} \quad (4)$$

The matrix is used to calculate the ratio index as:

$$P_{ij} = \frac{y_{ij}}{\sum\limits_{i-1}^{m} y_{ij}} \quad (5)$$

The ratio index is then used to estimate the entropy (Amiri et al. 2014) as follows:

$$e_j = -\frac{1}{\ln m} \sum_{i=1}^{m} P_{ij} \ln P_{ij} \quad (6)$$

The weight of each subsystem is finally calculated using Equation (7) as shown below:

$$w_j = \frac{1 - e_j}{\sum\limits_{j=1}^{n} (1 - e_j)} \quad (7)$$

3.3 | Computation of flood vulnerability and risk

Socio-economic and physical factors were subsequently integrated to estimate flood vulnerability index (FVI) (Balica, Douben, & Wright, 2009) using the following equation:

$$FVI = (F1_w F1_r + F2_w F2_r + ... + FN_w FN_r)/N, \quad (8)$$

where F represents flood vulnerability factor, N is the number of factors, w indicates the weights of the factors and r is the rank of different values.

3.4 | Assessment of model performance

The performance of the proposed method was evaluated using the receiver operating characteristic (ROC) curve. The ROC considers only two classes (A and B) to validate the model. Two values can have a maximum of four possible outcomes. If a method can identify a flood zone correctly, it is considered as true positive (TP) otherwise false positive (FP). Similarly, if the method fails to locate a flood zone correctly, it is considered as a true negative (TN) otherwise false negative (FN). In the ROC curve (Huang & Ling, 2005), TP is drawn against FP and then the area under the curve (AUC) is estimated to define model accuracy as:

$$\text{Accuracy} = \frac{\sum \text{TP} + \sum \text{TN}}{\text{Total Population}}. \quad (9)$$

The AUC in ROC curve is widely used in evaluating the performance of a classification model. It provides a measure of model capability for identifying different classes. The AUC in ROC provides different measures of model performance such as its sensitivity (TP/TP + FN), specificity (TN/TN + FP) and false alarm ratio (FP/TN + FP). The AUC-ROC is considered to be a composite metric for the reliability estimation of different properties of the classification model performance (Huang & Ling, 2005).

4 | RESULTS AND DISCUSSION

4.1 | Identification of flood vulnerability factors

A large number of factors related to flood vulnerability in Peninsular Malaysia have been documented in various studies (Alias et al., 2019; Dano et al., 2019; Elsheikh et al., 2015). A total of 19 factors were identified from the literature review based on availability, measurability and sensitivity. These factors are given in Table 2. A questionnaire survey was conducted among the stakeholders (50 individuals in total) to rank these 19 variables based on their importance and relevance to local conditions. The responses were then used to estimate their importance, and an index was developed. The estimated

TABLE 2 Flood vulnerability factors identified using the importance index

No.	Factor	Response			Total	Importance index	Rank
		1	2	3			
1	Population density	50	0	0	50	1.0000	1
2	Percentage of vulnerable people	45	4	1	50	0.9600	2
3	Land elevation	44	4	2	50	0.9467	3
4	Gini coefficient (inequity in wealth)	43	6	1	50	0.9467	4
5	Percentage of foreign population	42	7	1	50	0.9400	5
6	House-hold income	40	10	0	50	0.9333	6
7	Forest cover	42	5	3	50	0.9267	7
8	Education	39	10	1	50	0.9200	8
9	Accessibility of facilities	38	11	1	50	0.9133	10
10	Male to female ratio	36	13	1	50	0.9000	13
11	Local knowledge on flood zones	33	16	1	50	0.8800	15
12	Crop cultivated area	33	11	6	50	0.8467	18
13	Urbanisation ratio (urban to non-urban area in a district)	34	9	7	50	0.8467	19
14	Diversity in income	37	3	10	50	0.8467	20
15	Percentage people involve in trading	35	6	9	50	0.8400	21
16	Evacuation training	31	9	10	50	0.8067	25
17	Poverty rate	23	21	6	50	0.7800	29
18	Communication facilities	30	6	14	50	0.7733	31
19	Safety index	17	9	24	50	0.6200	36

importance and the rank of each factor, based on the importance index, are given in Table 2.

The ranking of different factors was classified using the Jenks optimization method (Table 3). The first column of Table 3 shows the class and the second column exhibits factors belonging to that class. The seven top factors identified are: (1) population density (2) percentage of vulnerable people; (3) elevation; (4) Gini coefficient; (5) percentage of foreign nationals; (6) household income; and (7) forest cover. These factors were considered important, in the Malaysian context, for delineating flood risk and carried forward for further study.

4.2 | Assessment of flood vulnerability

4.2.1 | Spatial distribution of flood vulnerability

Flood vulnerability factor data identified for Peninsular Malaysia was used to prepare flood maps. The spatial distribution of the factors identified as of most importance in assessing flood vulnerability in Peninsular Malaysia is presented in Figure 2. The values of population density,

TABLE 3 Clustering of flood vulnerability factors according to the importance index

Class	Factors
1	(1) population density (2) percentage of vulnerable people; (3) elevation; (4) Gini coefficient; (5) percentage of foreign nationals; (6) household income; (7) forest cover
2	(8) education; (9) accessibility to facilities (10) male to female ratio; (11) local knowledge regarding floods; (12) crop cultivated area; (13) urbanisation ratio
3	(14) diversity in income; (15) percentage of trading people; (16) evacuation training; (17) poverty rate; (18) communication facilities; (19) safety index

percentage of vulnerable people to the total population, Gini coefficient, household income and percentage of foreign population to the total population were divided into five classes using the natural break algorithms. Most of the lands of Peninsular Malaysia are situated at an elevation of 200 m so elevation data below 200 m is classified into four equal classes when preparing the elevation map.

Figure 2 shows the high variability in population density which characterises Peninsular Malaysia. Higher
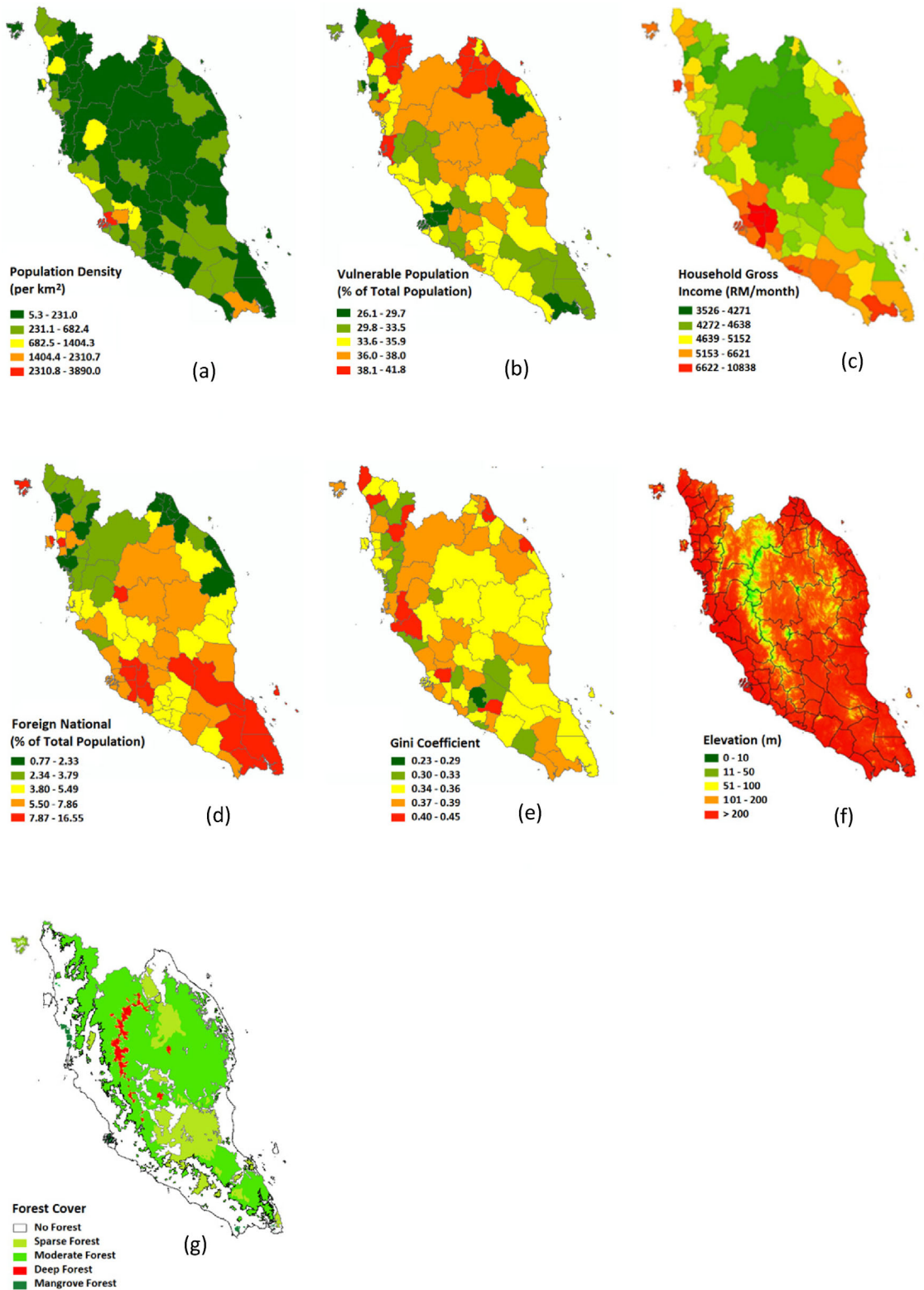
ZIARH ET AL.

CIWEM Chartered Institution of Water and Environmental Management | Journal of Flood Risk Management—**WILEY** | **7 of 13**

FIGURE 2 Spatial distribution of: (a) population density; (b) vulnerable population; (c) household income; (d) foreign national; (e) Gini coefficient; (f) elevation; and (g) forest land

population densities were observed in Klang and Johor Bahru where the major urban centres are located, while districts in the central part of the area (covered mainly by mountains and forest) have a low population density. The percentage of vulnerable population, relative to the total population, was higher in the northern districts, and lower in the districts where the major urban centres are located. Household income is higher in the districts with large

TABLE 4 Ratings of different classes of flood vulnerability subsystems using the catastrophe theory (initial and normalised values of different classes of the indicators are also shown)

Subsystem	Indicator range	Initial value	Normalised value	Rating
Population density (person/sq.km)	5.3–231.0	118.2	0.00	0.00
	231.1–682.4	456.8	0.11	0.49
	682.5–1,404.3	1,043.4	0.31	0.75
	1,404.4–2,310.7	1857.6	0.58	0.90
	2,310.8–3,890.0	3,100.4	1.00	1.00
Vulnerable population (% of total population)	26.1–29.7	27.90	0.00	0.00
	29.8–33.5	31.65	0.31	0.68
	33.6–35.9	34.75	0.57	0.87
	36.0–38.0	37.00	0.76	0.95
	38.1–41.8	39.95	1.00	1.00
Household income (RM/month)	3,525–4,271	3,898.0	0.00	0.00
	4,273–4,638	4,455.5	0.12	0.49
	4,639–5,152	4,895.5	0.21	0.67
	5,153–6,621	5,887.0	0.41	0.84
	6,622–10,838	8,730.0	1.00	1.00
Foreign nationals (% of total population)	0.77–2.33	1.55	0.00	0.00
	2.34–3.79	3.07	0.14	0.53
	3.80–5.49	4.65	0.29	0.73
	5.50–7.86	6.68	0.48	0.86
	7.87–16.55	12.21	1.00	1.00
Gini coefficient	0.23–0.29	0.260	0.00	0.00
	0.30–0.33	0.315	0.33	0.70
	0.34–0.36	0.350	0.55	0.86
	0.37–0.39	0.380	0.73	0.94
	0.40–0.45	0.425	1.00	1.00
Elevation (m)	0–10	5.0	0.00	0.00
	11–50	25.5	0.08	0.44
	51–100	75.5	0.29	0.73
	101–200	150.5	0.59	0.90
	201–300	250.5	1.00	1.00
Forest cover	Mangrove forest	0.00	0.00	0.00
	Dense forest	1.00	0.25	0.63
	Moderate forest	2.00	0.50	0.84
	Sparse forest	3.00	0.75	0.94
	No forest	4.00	1.00	1.00

cities and lower in the central regions. A higher percentage of foreign nationals to the total population was observed in the districts surrounding major cities. Most of the existing industries are located in districts where large numbers of foreign workers are also located. The districts with a high Gini coefficient were found to be irregularly distributed over Peninsular Malaysia. Along the coast, land is generally below 50 m, gradually increasing to 200 m and above in the central part of the country. About 65% land of Peninsular Malaysia is covered by forest. In this study, general forest is classified as sparse, moderately dense to dense, with mangrove forest separately classified.

TABLE 5 The weights of different flood vulnerability subsystems obtained using entropy theory

Subsystem	Initial value	Ratio (P_{ij})	Entropy (e_j)	Weight
Population density (person/sq.km)	118.2	0.02	−0.07	0.30001
	456.8	0.07	−0.19	
	1,043.4	0.16	−0.29	
	1857.6	0.28	−0.36	
	3,100.4	0.47	−0.35	
Vulnerable population (% of total population)	27.90	0.16	−0.30	0.006495
	31.65	0.18	−0.31	
	34.75	0.20	−0.32	
	37.00	0.22	−0.33	
	39.95	0.23	−0.34	
Household income (RM/month)	3,898.0	0.14	−0.28	0.037554
	4,455.5	0.16	−0.29	
	4,895.5	0.18	−0.31	
	5,887.0	0.21	−0.33	
	8,730.0	0.31	−0.36	
Foreign nationals (% of total population)	1.55	0.06	−0.16	0.178647
	3.07	0.11	−0.24	
	4.65	0.17	−0.30	
	6.68	0.24	−0.34	
	12.21	0.43	−0.36	
Gini coefficient	0.260	0.15	−0.28	0.011540
	0.315	0.18	−0.31	
	0.350	0.20	−0.32	
	0.380	0.22	−0.33	
	0.425	0.25	−0.34	
Elevation (m)	5.0	0.01	−0.05	0.362660
	25.5	0.05	−0.15	
	75.5	0.15	−0.28	
	150.5	0.30	−0.36	
	250.5	0.49	−0.35	
Forest cover	0.00	0.07	−0.18	0.103094
	1.00	0.13	−0.27	
	2.00	0.20	−0.32	
	3.00	0.27	−0.35	
	4.00	0.33	−0.37	

4.2.2 | Spatial distribution of composite flood vulnerability

The ranks of different features of each flood vulnerability subsystem were estimated using the catastrophe theory. The results are presented in Table 4. The values of each indicator were classified using the Junk's optimization technique. The classified values of the indicators are given in the second column of Table 4. The mean value of each class was used as the initial value for the estimation of rank of each class, and the mean values were first normalised using Equations (2) and (3). The mean and the normalised values of the different classes of the indicators are given in the third and fourth column of Table 4. The catastrophe functions were then used for the estimation of ranks of each class. The Wigwam model (Table 2) was used for this
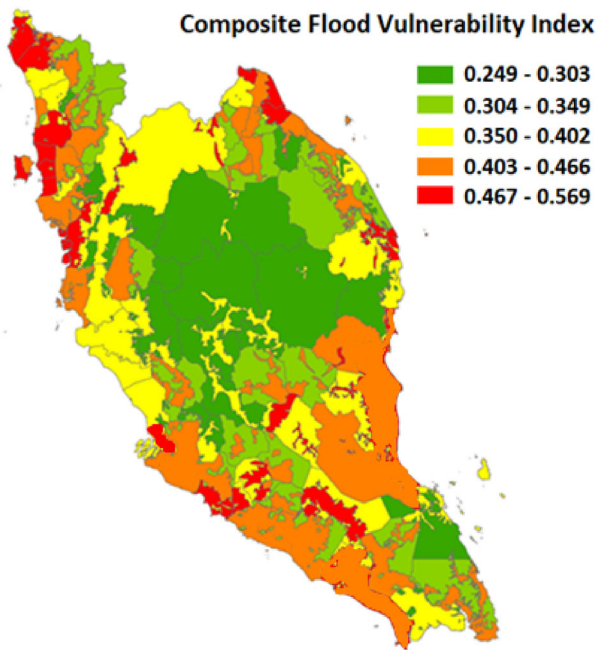
FIGURE 3 Composite flood vulnerability map of peninsular Malaysia

FIGURE 4 Flood hazard map (source: Department of Irrigation and Drainage, Malaysia)

purpose as there were five classes for each indicator. As an example, the normalised values of the population density classes were 0.0, 0.11, 0.31, 0.58 and 1.00. Application of the catastrophe fuzzy functions Wigwam model estimated the ratings from the normalised values as $a_u = u^{0.5} = 0.0^{0.5} = 0.0$, $a_v = v^{0.33} = 0.11^{0.33} = 0.49$, $a_w = w^{0.25} = 0.31^{0.25} = 0.75$, $a_x = x^{0.20} = 0.58^{0.20} = 0.90$ and $a_y = y^{0.17} = 1.0^{0.17} = 1.0$. Similarly, the ratings of all other classes of the other indicators were estimated and are presented in the last column of Table 4.

Next, the weight of each subsystem was estimated using the entropy theory (Table 5). The initial values of different classes of the indicators were used to estimate the ratio of each value using Equation 5. Each value was divided by the sum of all initial values of an indicator. Then the entropy of each class was estimated using Equation 6, the results of which are presented in the fourth column of Table 5. Finally, the weight of each subsystem was estimated using Equation 7. These results are presented in the last column of Table 5.

The estimated weight of different subsystems, and rating of different classes of the indicator of the subsystems, were used in Equation 8 to estimate flood vulnerability. The vulnerability values ranged from 0.249 to 0.569. The values were then classified into five categories using Junk's optimization method. Figure 3 shows the resulting flood vulnerability map of Peninsular Malaysia.

The map (Figure 3) shows that coastal areas of Peninsular Malaysia are the most vulnerable, while the central areas are the least susceptible to floods. The highest
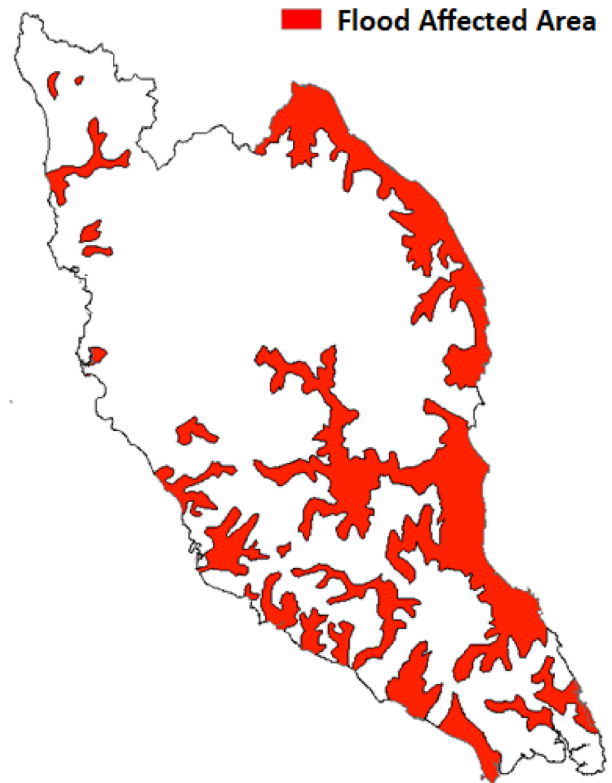
vulnerability was observed along the coastal regions in the northwest, extreme northeast, central-east and southwest of the Peninsula. The rest of the coastal regions exhibited flood vulnerability scores in the range 0.402–0.466.

4.3 | Flood hazard map

Figure 4 shows a flood hazard map of the area. Recent floods that occurred in Peninsular Malaysia in 1965, 1967, 1971, 1973, 1983, 1988, 1993, 1998, 2000, 2007, 2008, 2009, 2010 and 2014 are taken into account. The map was developed based on inundation areas during historical floods, estimated through ground observation and satellite images. The result indicates that floods mostly occur on the coastal plains. The flood hazard vulnerability is higher in the south, especially on the southeast and southwest coastal areas, while flood vulnerability is lower in the Centre of the Peninsula as these areas are generally not exposed to floods.

4.4 | Flood risk estimation

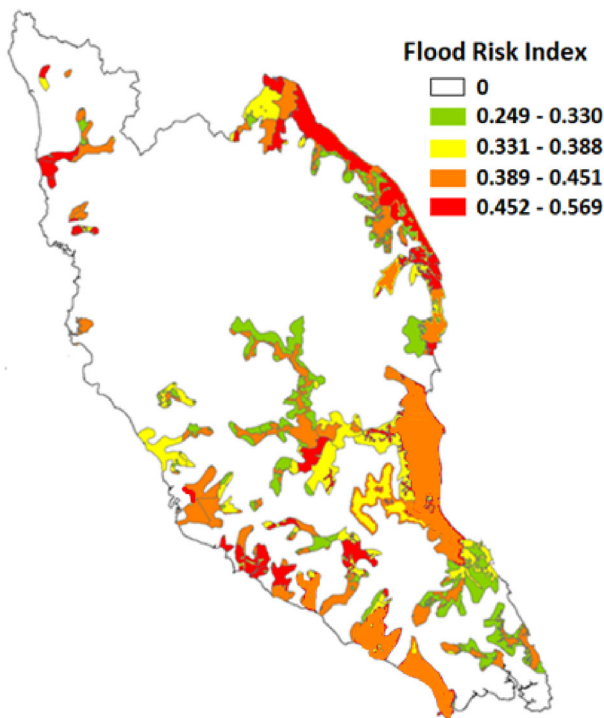Figure 5 shows spatial distribution of flood risk. This map has been developed by multiplying the flood hazard

FIGURE 5 Spatial distribution of flood risk

TABLE 6 The performance of integrated catastrophe-entropy method in mapping flood risk

Flood risk	True positive	False positive
1	34	45
2	15	18
3	6	9
4	5	6
5	3	5
Sum	63	83
AUC	**0.93**	

4.5 | Validation of flood risk map

The general perception regarding flood risk is that devastating flood events occur more frequently, that is, the risk is higher, in the northeast districts of Malaysia. Some districts in the southeast, however, are also severely affected by frequent floods. The AUC ROC tool was used to evaluate whether the flood risk map, developed using the integrated method proposed in this study, was able to map the flood risk zones of Peninsular Malaysia. The results are shown in Table 6 and indicate that the prediction capability of the integrated method is 0.93, reinforcing the suggestion that integrated use of both the catastrophe and entropy methods can provide the locations of flood risk with good accuracy.

5 | CONCLUSION

A data-driven MCDA approach through the integration of both catastrophe and entropy theories is proposed in this work which can provide an unbiased assessment of flood risk distribution in Peninsular Malaysia. This method can also be used for systematic assessment of the factors relevant to flood vulnerability and risk zone delineation. Seven major factors were accountable for flood risk. These were population density, percentage of vulnerable people, household income, and economy of the region, percentage of foreign nationals, elevation and forest cover. Using the proposed MCDA technique, this study revealed that coastal regions of Peninsular Malaysia are highly vulnerable to floods than inland locations. The highest flood risk was observed on the northwest coast. The efficiency of the proposed method was assessed using the AUC-ROC tool which indicated an accuracy of 0.93. The spatial variability of flood susceptible zones, and the factors that influence it, can be used to develop measures necessary for reducing future flood risk in Malaysia. The methodological framework necessary for the reliable mapping of flood risk proposed in this study, can be applied elsewhere. Despite the

map data with the flood vulnerability map data. The flood risk values were divided into five classes using Junk's optimization technique to derive the flood risk map. The values range from 0 to 0.569, with the lower value indicating low flood risk and higher values indicating higher risk. Figure 5 shows the highest flood risk zone (0.452–0.569) is located in the northeast coastal region. A large part of the southeast and southwest coasts also has a high flood risk rating (0.389–0.451). In general, it appears that locations with a high flood risk are more prevalent on the east coast than on the west coast.

There is a general conception that flood risk is much higher in the northeast coastal region (Nashwan, Shahid, et al., 2018b; Pradhan & Youssef, 2011). It is also highly devastating when flood occurs in urbanised areas along the central-western coast. The flood risk map generated in this study was found to match very well with the conception. Comparison of the flood risk map with flood vulnerability factors revealed causes of high flood risk in the northeast and central-western coasts. In the northeast, a high ratio of vulnerable people, low household income and greater inequality in wealth distribution (e.g., high Gini coefficient) are major factors contributing to flood risk. On the other hand, high population density together with a high ratio of foreign nationals is associated with high flood risk in the central-western coastal region of Peninsular Malaysia.

method was used to determine flood risk locations, there is opportunity to improve this work. For example, sensitivity of flood vulnerability factors can be evaluated to understand their relative importance. In addition, accuracy of the maps generated in this study depends on the quality and resolution of data. Hence, the effect of uncertainty in determining flood vulnerability and risk can be estimated in a future work. Besides, data of smaller administrative unit (sub-district or council) can be used for mapping flood risk areas accurately.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Mohamed Salem Nashwan https://orcid.org/0000-0003-4007-5878
Shamsuddin Shahid https://orcid.org/0000-0001-9621-6452

REFERENCES

Agarwal, A., Maheswaran, R., Sehgal, V., Khosa, R., Sivakumar, B., & Bernhofer, C. (2016). Hydrologic regionalization using wavelet-based multiscale entropy method. *Journal of Hydrology*, *538*, 22–32.

Ahmed, K., Shahid, S., Bin Harun, S., Ismail, T., Nawaz, N., & Shamsudin, S. (2015). Assessment of groundwater potential zones in an arid region based on catastrophe theory. *Earth Science Informatics*, *8*, 539–549.

Ahmed, K., Shahid, S., & Nawaz, N. (2018). Impacts of climate variability and change on seasonal drought characteristics of Pakistan. *Atmospheric Research*, *214*, 364–374.

Al-Abadi, A. M., Shahid, S., & Al-Ali, A. K. (2016). A GIS-based integration of catastrophe theory and analytical hierarchy process for mapping flood susceptibility: A case study of Teeb area, Southern Iraq. *Environmental Earth Sciences*, *75*, 687.

Alamgir, M., Mohsenipour, M., Homsi, R., Wang, X., Shahid, S., Shiru, M. S., ... Yuzir, A. (2019). Parametric assessment of seasonal drought risk to crop production in Bangladesh. *Sustainability*, *11*, 1442.

Alias, N., Salim, N., Mat Taib, S., Mohd Yusof, M. B., Saari, R., Adli, W., ... Blenkinsop, S. (2019). Community responses on effective flood dissemination warnings - a case study of the December 2014 Kelantan flood, Malaysia. *Journal of Flood Risk Management*, *13*, e12552.

Balica, S. F., Douben, N., & Wright, N. G. (2009). Flood vulnerability indices at varying spatial scales. *Water Science and Technology*, *60*, 2571–2580.

Castillo, A., Castelli, F., & Entekhabi, D. (2015). An entropy-based measure of hydrologic complexity and its applications. *Water Resources Research*, *51*, 5145–5160.

Chen, T.-Y., & Li, C.-H. (2010). Determining objective weights with intuitionistic fuzzy entropy measures: A comparative analysis. *Information Sciences*, *180*, 4207–4222.

Chen, Y., Barrett, D., Liu, R., Gao, L., Zhou, M., Renzullo, L., Cuddy, S. & Emelyanova, I. (2014). A spatial framework for regional-scale flooding risk assessment. Paper presented at 7th International Congress on Environmental Modelling and Software, pp. 15–19.

Cui, H., & Singh, V. P. (2015). Configurational entropy theory for streamflow forecasting. *Journal of Hydrology*, *521*, 1–17.

Cutter, S. L., Barnes, L., Berry, M., Burton, C., Evans, E., Tate, E., & Webb, J. (2008). A place-based model for understanding community resilience to natural disasters. *Global Environmental Change*, *18*, 598–606.

Dano, U. L., Balogun, A.-L., Matori, A.-N., Wan Yusouf, K., Rimi Abubakar, I., Mohamed, S., ... Pradhan, B. (2019). Flood susceptibility mapping using GIS-based analytic network process: A case study of Perlis, Malaysia. *Water*, *11*, 615.

Dewan, A. (2013a). *Floods in a megacity: Geospatial techniques in assessing hazards, risk and vulnerability*. Netherlands: Springer.

Dewan, A. M. (2013b). *Hazards, risk, and vulnerability. Floods in a megacity geospatial techniques in assessing hazards, risk and vulnerability*. Dordrecht: Springer.

Dewan, A. M., & Yamaguchi, Y. (2008). Effect of land cover changes on flooding: Example from greater Dhaka of Bangladesh. *International Journal of Geoinformatics*, *4*, 11–20.

DOSM. (2018). *Annual report, 2018*. Putrajaya, Malaysia: Department of Statistics Malaysia.

Elsheikh, R. F. A., Ouerghi, S., & Elhag, A. R. (2015). Flood risk map based on GIS, and multi criteria techniques (case study Terengganu Malaysia). *Journal of Geographic Information System*, *7*, 348–357.

Feloni, E., Mousadis, I., & Baltas, E. (2019). Flood vulnerability assessment using a GIS-based multi-criteria approach—The case of Attica region. *Journal of Flood Risk Management*, *13*, e12563.

He-Hua, W., Jian-Jun, Z., Xiao-Yan, G., & Ye, C. (2018). A multiple criteria decision analysis method for alternative assessment results obeying a particular distribution and application. *Mathematical Problems in Engineering*, *2018*, 14.

Huang, J., & Ling, C. X. (2005). Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, *17*(3), 299–310.

Jato-Espino, D., Lobo, A., & Ascorbe-Salcedo, A. (2019). Urban flood risk mapping using an optimised additive weighting methodology based on open data. *Journal of Flood Risk Management*, *12*, e12533.

Jenks, G. F. (1967). The data model concept in statistical mapping. *International Yearbook of Cartography*, *7*, 186–190.

Khan, N., Pour, S. H., Shahid, S., Ismail, T., Ahmed, K., Chung, E.-S., ... Wang, X. (2019). Spatial distribution of secular trends in rainfall indices of peninsular Malaysia in the presence of long-term persistence. *Meteorological Applications*, *26*, 655–670.

Lim, E. C., & Alum, J. (1995). Construction productivity: Issues encountered by contractors in Singapore. *International Journal of Project Management*, *13*, 51–58.

Matori, A. N., Lawal, D. U., Yusof, K. W., Hashim, M. A., & Balogun, A.-L. (2014). Spatial analytic hierarchy process model for flood forecasting: An integrated approach. *IOP Conference Series: Earth and Environmental Science*, *20*, 012029.

Mayowa, O. O., Pour, S. H., Shahid, S., Mohsenipour, M., Harun, S. B., Heryansyah, A., & Ismail, T. (2015). Trends in rainfall and rainfall-related extremes in the east coast of peninsular Malaysia. *Journal of Earth System Science*, *124*, 1609–1622.

Mishra, A., & Liu, S. C. (2014). Changes in precipitation pattern and risk of drought over India in the context of global warming. *Journal of Geophysical Research: Atmospheres*, *119*, 7833–7841.

Muhammad, M. K. I., Nashwan, M. S., Shahid, S., Bin Ismail, T., Song, Y. H., & Chung, E. S. (2019). Evaluation of empirical reference evapotranspiration models using compromise programming: A case study of peninsular Malaysia. *Sustainability*, *11*, 4267.

Nashwan, M. S., Ismail, T., & Ahmed, K. (2018a). Flood susceptibility assessment in Kelantan river basin using copula. *International Journal of Engineering and Technology*, *7*, 584–590.

Nashwan, M. S., Ismail, T., & Ahmed, K. (2019a). Non-stationary analysis of extreme rainfall in peninsular Malaysia. *Journal of Sustainability Science and Management*, *14*, 17–34.

Nashwan, M. S., & Shahid, S. (2019b). Symmetrical uncertainty and random forest for the evaluation of gridded precipitation and temperature data. *Atmospheric Research*, *230*, 104632.

Nashwan, M. S., Shahid, S., Chung, E. S., Ahmed, K., & Song, Y. H. (2018b). Development of climate-based index for hydrologic hazard susceptibility. *Sustainability*, *10*, 2182.

Nashwan, M. S., Shahid, S., & Wang, X. J. (2019b). Assessment of satellite-based precipitation measurement products over the Hot Desert climate of Egypt. *Remote Sensing*, *11*, 555.

Nigussie, T. A., & Altunkaynak, A. (2019). Modeling the effect of urbanization on flood risk in Ayamama watershed, Istanbul, Turkey, using the MIKE 21 FM model. *Natural Hazards*, *99*, 1031–1047.

Noor, M., Ismail, T., Chung, E.-S., Shahid, S., & Sung, J. (2018). Uncertainty in rainfall intensity duration frequency curves of peninsular Malaysia under changing climate scenarios. *Water*, *10*, 1750.

Noor, M., Ismail, T., Shahid, S., Nashwan, M. S., & Ullah, S. (2019). Development of multi-model ensemble for projection of extreme rainfall events in peninsular Malaysia. *Hydrology Research*, *50*, 1772–1788.

Pour, S. H., Bin Harun, S., & Shahid, S. (2014). Genetic programming for the downscaling of extreme rainfall events on the east coast of peninsular Malaysia. *Atmosphere*, *5*, 914–936.

Pour, S. H., Wahab, A. K. A., Shahid, S., Asaduzzaman, M., & Dewan, A. (2020). Low impact development techniques to mitigate the impacts of climate-change-induced urban floods: Current trends, issues and challenges. *Sustainable Cities and Society*, *62*, 102373. https://doi.org/10.1016/j.scs.2020.102373

Pradhan, B., & Youssef, A. (2011). A 100-year maximum flood susceptibility mapping using integrated hydrological and hydrodynamic models: Kelantan River corridor, Malaysia. *Journal of Flood Risk Management*, *4*, 189–202.

Rahman, M., Ningsheng, C., Islam, M. M., Dewan, A., Iqbal, J., Washakh, R. M. A., & Shufeng, T. (2019). Flood susceptibility assessment in Bangladesh using machine learning and multi-criteria decision analysis. *Earth Systems and Environment*, *3*, 585–601.

Sa'adi, Z., shahid, S., Ismail, T., Chung, E.-S., & Wang, X.-J. (2017). Trends analysis of rainfall and rainfall extremes in Sarawak, Malaysia using modified Mann–Kendall test. *Meteorology and Atmospheric Physics*, *131*, 1–15.

Salarpour, M., Yusop, Z., Yusof, F., Shahid, S., & Jajarmizadeh, M. (2013). Flood frequency analysis based on t-copula for Johor river, Malaysia. *Journal of Applied Sciences*, *13*, 1021–1028.

Santos, P. P., & Reis, E. (2018). Assessment of stream flood susceptibility: A cross-analysis between model results and flood losses. *Journal of Flood Risk Management*, *11*, S1038–S1050.

Sediqi, M. N., Shiru, M. S., Nashwan, M. S., Ali, R., Abubaker, S., Wang, X., ... Manawi, S. M. A. (2019). Spatio-temporal pattern in the changes in availability and sustainability of water resources in Afghanistan. *Sustainability*, *11*, 5836.

Shahid, S., Pour, S. H., Wang, X. J., Shourav, S. A., Minhans, A., & Bin Ismail, T. (2017). Impacts and adaptation to climate change in Malaysian real estate. *International Journal of Climate Change Strategies and Management*, *9*, 87–103.

Shahid, S., Wang, X. J., Harun, S. B., Shamsudin, S. B., Ismail, T., & Minhans, A. (2016). Climate variability and changes in the major cities of Bangladesh: Observations, possible impacts and adaptation. *Regional Environmental Change*, *16*, 459–471.

Singh, L. K., Jha, M. K., & Chowdary, V. (2018). Assessing the accuracy of GIS-based multi-criteria decision analysis approaches for mapping groundwater potential. *Ecological Indicators*, *91*, 24–37.

Tang, J.-Q., & Wang, C.-G. (2013). Entropy analysis on adaptability between supply structure and demand structure of passenger transportation. *Journal of the China Railway Society*, *35*, 1–7.

Wang, W., Liu, S., Zhang, S., & Chen, J. (2011). Assessment of a model of pollution disaster in near-shore coastal waters based on catastrophe theory. *Ecological Modelling*, *222*, 307–312.

Wang, X.-J., zhang, J.-Y., Shahid, S., Xia, X., He, R.-M., & Shang, M.-T. (2012). Catastrophe theory to assess water security and adaptation strategy in the context of environmental change. *Mitigation and Adaptation Strategies for Global Change*, *19*(4), 1–15.

Wong, C.-L., Yusop, Z., & Ismail, T. (2018). Trend of daily rainfall and temperature in peninsular Malaysia based on gridded data set. *International Journal of GEOMATE*, *14*, 65–72.

Yaseen, Z. M., Ebtehaj, I., Kim, S., Sanikhani, H., Asadi, H., Ghareb, M. I., ... Shahid, S. (2019). Novel hybrid data-intelligence model for forecasting monthly rainfall with uncertainty analysis. *Water*, *11*, 502.

Yli-Viikari, A., Hietala-Koivu, R., Huusela-Veistola, E., Hyvönen, T., Perälä, P., & Turtola, E. (2007). Evaluating Agri-environmental indicators (AEIs)—Use and limitations of international indicators at national level. *Ecological Indicators*, *7*, 150–163.

Zhou, W., Ma, T., Chen, L., Wu, L., & Luo, Y. (2018). Application of catastrophe theory in comprehensive ecological security assessment of plastic greenhouse soil contaminated by phthalate esters. *PLoS One*, *13*, e0205680.