Monitoring Spatiotemporal Changes in Urban Flood Vulnerability of Peninsular Malaysia from Satellite Nighttime Light Data

Ghaith Falah Ziarh[a], Eun Sung Chung[a,**], Ashraf Dewan[b], Md Asaduzzaman[c], Mohammed Magdy Hamed[d,e], Zafar Iqbal[f] and Shamsuddin Shahid[e]

[a]*Faculty of Civil Engineering, Seoul National University of Science and Technology, 232 Gongneung-ro, Nowon-gu, Seoul 01811, Korea*
[b]*School of Earth and Planetary Sciences (EPS), Faculty of Science and Engineering, Curtin University, Pert, Australia*
[c]*School of Digital, Technologies and Arts, Staffordshire University, UK*
[d]*Construction and Building Engineering Department, College of Engineering and Technology, Arab Academy for Science, Technology and Maritime Transport (AASTMT), B 2401 Smart Village, 12577, Giza, Egypt*
[e]*Department of Water & Environmental Engineering, Faculty of Civil Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia*
[f]*NUST Institute of Civil Engineering-SCEE, National University of Sciences and Technology (NUST), H-12, Islamabad, 44000, Pakistan*

Abstract

Urban flood vulnerability monitoring requires a large amount of socioeconomic and environmental data collected at regular time intervals. However, collecting such a large volume of data poses a significant constraint in assessing changes in flood vulnerability. This study proposed a novel method to monitor spatiotemporal changes in urban flood vulnerability from satellite nighttime light (NTL) data. Peninsular Malaysia was chosen as the research region as floods are the most devastating and recurrent phenomena in the region. The study developed a flood vulnerability index (FVI) based on socioeconomic and environmental data from a single year. This FVI was then linked to NTL data using an Adaptive neuro-fuzzy inference system (ANFIS) machine learning algorithm. The model was calibrated and validated with administrative unit scale data and subsequently used to predict FVI at a spatial resolution of 10 km for 2000–2018 using NTL data. Finally, changes in estimated FVI at different grid points were evaluated using the Mann-Kendall trend method to determine changes in flood vulnerability over time and space. Results showed a nonlinear relationship between NTL and flood vulnerability factors such as population density, Gini coefficient, and percentage of foreign nationals. The ANFIS technique performed well in estimating FVI from NTL data with a normalized root-mean-square error of 0.68 and Kling-Gupta Efficiency of 0.73. The FVI revealed a high vulnerability in the urbanized western coastal region (FVI ~ 0.5 to 0.54), which matches well with major contributing regions to flood losses in Peninsular Malaysia. Trend assessment showed a significant increase in flood vulnerability in the study area from 2000 to 2018. The spatial distribution of the trend indicated an increase in FVI in the urbanized coastal plains, particularly in rapidly developing western and southern urban regions. The results indicate the potential of the technique in urban flood vulnerability assessment using freely available satellite NTL data.

Keywords: flood vulnerability; remote sensing; nighttime light; machine learning; trend analysis

1. Introduction

Floods are the most costly and common natural disasters, affecting nearly 250 million people and causing more than $40 billion in damage worldwide yearly (OECD, 2016). About a third of natural catastrophe economic losses are

* Corresponding author: Areum Hall 305, Gongneungro 232, Nowongu, Seoul 01811, Republic of Korea
Email address: eschung@seoultech.ac.kr (E.-S. Chung)

http://dx.doi.org/10.1016/j.ijepes.2021.000000

attributable to flood damage (Bosello et al., 2018). A significant increase in flood risk, mainly urban flood risk, has been reported globally in recent years (Guo et al., 2021; Miller and Hutchins, 2017; Pour et al., 2020). Rapid population expansion in urban areas and the concentration of construction in flood plains contribute significantly to these regions' vulnerability to flooding (OECD, 2016; Cooper, 2019; Guerreiro et al., 2018; Liu and Niyogi, 2019). Changes in human settlements, infrastructure, public policy and administration, organizational capacity, social inequities, gender, economic patterns, etc., have all been identified as variables that affect flood susceptibility in study regions (Feloni and Baltas, 2020; Ibrahim et al., 2017; Romali and Yusop, 2021). Despite adopting various mitigation measures, population exposure to floods and flood damages is increasing rapidly worldwide (Glass, 2013; Opperman and Duvail, 2013; Keating et al., 2017). More than 50% of the global population is projected to be in flood hazard zones by 2030 (OECD, 2016). Reducing flood vulnerability is most appropriate to mitigate property damage and reduce flood risk. Assessing flood vulnerability and understanding its changes over time is vital for planning flood vulnerability reduction measures.

Assessing flood vulnerability requires a lot of hydrometeorological, physical, and socioeconomic data (Zhang et al., 2021; Soltani et al., 2021). Traditionally, the acquisition of socioeconomic and environmental data is time-consuming and costly. The flood vulnerability also changes with the changes in socioeconomic and environmental conditions. Therefore, monitoring the changes in flood vulnerability needs to collect a large amount of data at regular intervals. However, collecting such massive data is costly and not practicable for many developing countries. Improved remote sensing and computational capabilities have made it possible to accurately predict a wide range of physical and hydrological parameters. It made environmental change monitoring easier for the areas lacking sufficient observations (Lees et al., 2021; Mohanty and Simonovic, 2021).

Nighttime light (NTL) from earth-observing satellites is a valuable data source for understanding anthropogenic activities (Bennett and Smith, 2017; Coesfeld et al., 2020; Elvidge et al., 2020; Zhao et al., 2019). In recent years, NTL has been used for monitoring urban growth (Bharti and Tatem, 2018; Chen et al., 2019; Li et al., 2017), mapping dynamics of human population spread (Bennet and Smith, 2017; Chen et al., 2019; Enenkel et al., 2020; Tian et al., 2014; Kohiyama et al., 2004), studying the impacts of disasters on human settlements (Kohiyama et al., 2004; McDermott et al., 2015), explaining economic expansion, monitoring carbon emission from cities (Yang et al., 2020) and understanding human movements (Chen et al., 2019). Such data can be used as a proxy to assess socioeconomic change and understand how population vulnerability to a potential disaster changes (Guo et al., 2018; Ceola et al., 2015; de Miguel et al., 2020).

However, studies related to vulnerability assessment using NTL data are very limited. Only a few attempts have been made in recent years to assess vulnerability to different hazards using NTL data (Ceola et al., 2015; Fan et al., 2021; Mård et al., 2018). Fan et al. (2021) estimated the seismic vulnerability of buildings using NTL and MODIS vegetation index data. Y. Li et al. (2021) estimated the heating degree days (HDD) from NTL and satellite air pollution data to proxy for climate change vulnerability. Using NTL as a proxy for human settlement development along rivers linked to increased flood risk, Ceola et al. (2015) demonstrated how flood risk has fluctuated through time. Mård et al. (2018) used NTL to assess humans settling away from rivers after floods to understand human responses to flooding. However, those studies were limited to qualitative assessment of flood vulnerability. The only attempt taken so far to quantify flood susceptibility modeling using NTL data is by Fang et al. (2021). They used a multicriteria decision analysis (MCDA) tool to estimate the flood susceptibility index from elevation, drainage density, rainfall, and vegetation density data. They then developed a linear regression model to establish the relation between NTL and the susceptibility index. The major drawback of the study was the use of simple linear regression. Previous studies reported nonlinear relationships between socioeconomic changes and NLT data (Yang et al., 2020; Qiang et al., 2020).

Increased spatial heterogeneity in artificial nighttime lighting sources due to variations in development, land use, human activities, and cultural practices at fine spatial scales significantly affect NTL brightness. Consequently, NTL corresponds to socioeconomic factors nonlinearly (Ma, 2018; Wan et al., 2023). The nonlinear relationship is a major challenge in modeling socioeconomic changes from NTL data. For instance, a positive exponential trend exists between NTL and population density (Chen et al., 2019). However, the exponential curve is often very sharp for higher

values associated with NTL. Deriving such exponential equations reliably using conventional statistical methods is often very difficult. Thus, the demand for sophisticated modelling methods is increasing. The machine learning (ML) algorithm allows the development of highly nonlinear and complex relationships between variables. In order to better represent complex systems, several ML techniques have been created during the past four decades. Studies reported that integrating two or more classical ML can improve the capability of deriving complex nonlinear relationships compared to classical ML algorithms. An adaptive neuro-fuzzy inference system (ANFIS) is one such algorithm that combines two ML algorithms: artificial neural network (ANN) and fuzzy system (FS) (Jang, 1993; Nayak et al., 2004). The ANFIS mitigates the shortcomings of FS by leveraging ANN's capacity for learning and adding fuzzy if-then rules. This has made ANFIS highly applicable for mapping complex input-output relationships.

The purpose of this research is to calculate flood risk using NTL information. Since Peninsular Malaysia is one of Southeast Asia's most quickly developing countries, it was chosen as a case study to illustrate the effects of land use and economic development on flood vulnerability in the region. The increase in flood vulnerability has correspondingly resulted in increased economic damages. The existing literature indicates that this is possibly the first attempt to monitor spatiotemporal changes in flood vulnerability using NTL data.

2. Study area and data

2.1. Research area

Peninsular Malaysia, situated in the tropics between latitudes 1.20°-6.40° N and longitudes 99.35°-104.20° E (refer to Fig. 1), covers an area of 130,598 km². Thailand is located to the north of Peninsular Malaysia, while Singapore is to the south, with the Melaka Strait to the west and the South China Sea to the east. The terrain of Peninsular Malaysia comprises irregular mountainous forests sloping down to the coastline. Its highly varied topography and irregular coastlines contribute to the complexity of rainfall patterns in the region (Nashwan et al., 2018). The region receives an annual average rainfall of 1460 to 2579 mm (Tangang, 2001). The monsoon winds, steep terrain, and complex land-sea interactions profoundly influence the rainfall distribution. Malaysia experiences two distinct monsoon systems: the northeast monsoon (NEM) and the southwest monsoon (SEM). The southwest monsoon occurs from May to August, bringing relatively drier conditions, while the northeast monsoon from November to February brings heavy rainfall, particularly to the eastern coastal areas. The annual mean daily maximum temperature in the region ranges from 31 to 33°C, with relative humidity varying between 80.3% and 85.7% (Muhammad et al., 2021).

[Fig. 1. about here.]

Around 78% of Malaysia's people reside in cities. Most residents live in 51 cities, covering only 4% of the total land. In some West Coast states, nearly 95% of the residents live in urban areas. The locations of major urban centres of Peninsular Malaysia are shown in Fig. 2. Urban floods are the study area's most frequent and damaging natural disasters (Hirschmann, 2020). Floods of large magnitude often affect electricity, food supplies, clean water, health care, sewerage, and other emergency services, resulting in extensive economic damage. The total economic damage resulting from floods and post-flood-related damages in Peninsular Malaysia during 1965-2020 shows increased damage in recent years (Fig. 3). This is mainly due to increased economic activities in the flood-risk area.

[Fig. 2. about here.]

[Fig. 3. about here.]

2.2. Geospatial data

Geospatial data in terms of NTL images come from Defence Meteorological Satellite Program (DMSP) - Operational Linescan System (OLS) sensors for the period 1992-2012 and Visible Infrared Imaging Radiometer Suite (VIIRS) device for the period 2012-2018. The DMSP NTL data range between 0 and 63, where 0 indicates no light, and 63 indicates complete saturation. VIIRS, on the other hand, is an improved dataset. Downloaded from http://creativecommons.org/publicdomain/zero/1.0/, the harmonized NTL data used in this study has a spatial resolution of 1 km and spans from 2000 to 2018.

2.3. Flood vulnerability data

This research used the FVI map of Peninsular Malaysia that was created by Ziarh et al. (2021). Their analysis of the available literature determined the social, physical, and environmental aspects contributing to people being at risk of being flooded. A public opinion poll was the basis for the seven variables used: population density (person/km^2), the proportion of the vulnerable population (children aged below 12 years and elderly aged above 65-year) (%), family income (Malaysian Ringgit), local economy (GDP in Ringgit), percentage of foreign nationals (%), altitude (m), and forest cover (% to total area of a district) (Ziarh et al., 2021). Subsequently, they employed a novel data-driven MCDA by integrating two state-of-the-art data-driven techniques, namely catastrophe and entropy theories, to estimate flood vulnerability on a scale of 0 to 1. Fig. 4. shows the FVI map of Peninsular Malaysia. The higher FVI value in the map indicates higher vulnerability. The procedure followed in developing the FVI map can be found in Ziarh et al. (2021).

3. Method

3.1. Adaptive Neuro-Fuzzy Inference System

This study employed a single-input ANFIS model (Jang, 1993), a powerful hybrid ML model, to estimate FVI from NTL data. Complex input-output relationships can be mapped well, allowing for efficient problem resolution. Using a sequence of if-then rules and membership functions (MF), the inference engine of an ANFIS is based on the Takagi-Sugeno-Kang (TSK) fuzzy inference system (Soroush et al., 2019). Fig. 5. shows the structure of a typical single-input ANFIS model. It takes the input variable (X) and passes it to the fuzzification layer. Every node in the fuzzification layer is an adaptive node with its own node function. The membership degree of the relevant fuzzy set is calculated after mapping X to fuzzy sets x1 and x2 (Jiang et al., 1996). The multiplicative effect of the fuzzified values is applied to the output of the product layer, which is composed of fixed nodes. The corresponding firing strength in this layer is determined using a set of rules analogous to "if-then" and "AND" statements. The product value (w) is transferred to the third layer (normalization), consisting of neurons (N) employed for estimating normalized firing strength using Eq. (1):

$$\overline{w_1} = \frac{w_1}{w_1 + w_2} \tag{1}$$

where, w_1 and w_2 represent the output of the product layer. The nodes in the normalization layer calculate part of the contribution of the rule (Adnan et al., 2022). The normalized firing strength (defuzzification) at the output of the nodes in the fourth layer can be expressed using Eq. (2).

$$\overline{w_i} f_i = \overline{w_i}(p_i x + q_i y + r_i) \tag{2}$$

where, y is the observed FVI, f_i is the defuzzification function, and p_i, q_i and r_i represent inference parameters. Finally, outputs of the fourth layer are aggregated to provide predicted FVI (Y) through the last layer using Eq. (3):

$$Y = \sum_i \overline{w_i} f_i \tag{3}$$

The skill of ANFIS in mapping the input-output relationship depends on the MF type and the number. In this study, those parameters were optimized using k-fold cross-validation (Kohavi, 1995).

[Fig. 5. about here.]

3.2. Estimation of change in flood vulnerability

The slope estimator (SSE) proposed by Sen (1968) is used to evaluate the change in FVI, and the modified Mann-Kendall (MMK) test proposed by Yue et al. (2002) is used to evaluate the statistical significance of the trend. SSE estimate change as the median of N-1 slopes between two consecutive points ($\boldsymbol{x_{t'}}$ and $\boldsymbol{x_t}$) of an N-point data series,

$$SSE = \frac{x_{t'} - x_t}{t' - t} \tag{4}$$

Mann Kendall statistic (S) for a time series is calculated as:

$$S = \sum_{k=1}^{n-1} . \sum_{i=k+1}^{n} \text{sign } (x_i - x_k) \tag{5}$$

$$where\ \ sign\ (x_i - x_k) = \begin{cases} +1\ if\ (x_i - x_k) > 0 \\ 0\ if\ (x_i - x_k) = 0 \\ -1\ if\ (x_i - x_k) < 0 \end{cases} \tag{6}$$

The normalized test statistic, Z is estimated from the variance of S (Var(S)) as follows:

$$Z = \begin{cases} \dfrac{S-1}{\sqrt{Var\ (S)}} & \text{if S} > 0 \\ 0 & \text{if S} = 0 \\ \dfrac{S-1}{\sqrt{Var\ (S)}} & \text{if S} < 0 \end{cases} \tag{7}$$

The null hypothesis of 'no trend' is rejected at the 95% significance level if the Z value is less than or higher than 1.96. To account for the predicted lack of serial correlation in FVI data, Yue et al. (2002) recommended adding a pre-whitening (PW) step to the standard MK test. It is possible to formulate the PW method as:

$$Y_i = x_i - (\beta \cdot i) \tag{8}$$

where β is the Theil-Sen's slope. The PW is calculated before applying the MK test using Eq. (9) and (10):

$$Y_i' = Y_i - r_1 \times Y_{i-1} \tag{9}$$

$$Y_i'' = Y_i' + (\beta \times i) \tag{10}$$

3.3. Assessing model performance

Three statistical indicators are employed to assess the model's efficacy. The ANFIS model's ability to predict FVI from NTL data is evaluated using the modified index of agreement (MD), the normalized root mean square error (NRMSE) and the Kling-Gupta efficiency (KGE).

$$MD = 1 - \frac{\sum_{i=1}^{n}(x_{obs} - x_{sim})^j}{\sum_{i=1}^{n}(|x_{sim} - \overline{x_{obs}}| + |x_{obs} - \overline{x_{0bs}}|)^j} \tag{11}$$

$$NRMSE = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(S_i - O_i)^2}}{\sigma} \quad (12)$$

$$KGE = 1 - \sqrt{(r-1)^2 + \left(1 - \frac{\mu_{sim}}{\mu_{obs}}\right)^2 + \left(\frac{\sigma_{sim}/\mu_{sim}}{\sigma_{obs}/\mu_{obs}}\right)^2} \quad (13)$$

where $\boldsymbol{x_{sim}}$ is the i-th NTL data and $\boldsymbol{x_{obs,i}}$ is the i-th observed data; n is the sample size; r is Pearson's correlation; and are the mean and standard deviation of NTL (sim) and observed (obs) data, respectively; r is Pearson's correlation. NRMSE values may be anywhere from 0 to ∞, MD values from 1.0 to -1.0, and KGE values from -∞ to 1.0. Both MD and KGE should be set to 1, while NRMSE should be set to 0.

4. Results

4.1. Correlation of NTL with flood vulnerability factors

The NTL and various flood vulnerability factors, identified by Ziarh et al. (2021), averaged for the district level, are correlated to explore the relation of NTL with flood vulnerability. The relationship of NTL with (a) the total population, (b) the percentage of the foreign population to the total population, and (c) the Gini coefficient for the year 2016 is presented in Figure 6. A positive exponential relationship of NTL with the total population and the percentage of the foreign population to the total population is noticed. In contrast, a negative exponential relationship of NTL with the Gini coefficient is observed. The higher population indicates higher light, while more income disparity in large urban areas than in non-urban areas justifies the relationship of NTL with population density and the Gini coefficient. Generally, economic activities are intense in large cities, and thus, concentrations of foreign populations are high. Since NTL increases exponentially with the proportion of non-native speakers, this justifies seeing NTL as positively correlated with the share of non-native speakers in the entire population.

The relationship of NTL with other flood vulnerability indicators is also expected. The center regions, where most of the population lives in rural areas, have the lowest median household income, while the districts with major cities have the highest. Most people live in the plain lands along the coast compared to the elevated central region. The population in the forested area is less compared to non-forested areas. These indicate a positive association between household income and NTL and a negative association of NTL with elevation and forest cover.

[Fig. 6. about here.]

4.2. Flood vulnerability prediction using NTL

ANFIS was used to develop a model for predicting FVI from NTL. The NTL data averaged for each FVI polygon was used for model development. Model calibration required 70% of the data, whereas model validation required 30%. A k-fold cross-validation approach was used to tune ANFIS hyperparameters to improve the model accuracy. The model's performance was evaluated using visual interpretation and three statistical metrics. Fig. 7. displays the scatter plot of observed and ANFIS-estimated FVI during model validation, revealing a strong correlation between the two sets of data.

The statistical performance of the model during calibration and validation based on different metrics is also presented in Fig. 7. During model validation, the ANFIS model demonstrated a low NRMSE of 0.007, indicating good agreement between observed and estimated FVI values. The md was high at 0.768, further supporting the model's

accuracy in capturing the spatial pattern of FVI. Additionally, the KGE value of 0.732 indicates that the model performed well in reproducing the mean, variability, and distribution of the observed FVI. The R2 for the validation period was 0.81, suggesting that 81% of the variability in FVI can be explained by the NTL data used in the model. This high R2 value indicates the model's strong predictive ability and skill in estimating FVI based on nighttime light data. Overall, the results indicate that the ANFIS model performed well in predicting FVI during the validation phase, with high accuracy and reliability in capturing flood vulnerability trends.

The scatter plot in Fig. 7 indicates that the model predicts a slightly higher value for the low FVI. However, the overestimation was significantly less. This suggests that caution should be exercised when interpreting the model's predictions, especially for regions where the FVI is expected to be low.

[Fig. 7. about here.]

4.3. Estimation of flood vulnerability for Peninsular Malaysia

The ANFIS model estimates FVI from NTL data at each NTL grid point for all the years (2000-2018). This produced a series of high-resolution FVI maps for Peninsular Malaysia. The derived FVI maps for four years are presented in Fig. 8. They show an increase in flood-vulnerable regions over time. The spatial distribution of average FVI is presented in Fig. 9. The urban areas of the western coastal region are more vulnerable to floods than other peninsula regions. The most vulnerable region is the central-western region (0.5 to 0.54), surrounding Malaysia's capital city. The vulnerability is also high in the northeast. Overall, the vulnerability is higher in major cities regions. This is mainly due to high population density. Besides, lower elevation and less forest cover could have made those areas more vulnerable.

[Fig. 8. about here.]

[Fig. 9. about here.]

4.4. Trends in flood vulnerability

The derived FVI time series for 2000-2018 at each grid location using NTL data was used to assess the trends in FVI. First, the FVI for the entire peninsula for each year was summed to prepare a yearly time series. Fig. 10. displays the FVI time series for the research region. The FVI sum of the peninsular was found to increase sharply over time. It increased from <140 in 2000 to 352 in 2018. Statistical analysis of the trend in the FVI total showed an upward trend of 11.1% per year at $p < 0.05$, indicating a fast rise in flood vulnerability in Peninsular Malaysia.

[Fig. 10. about here.]

The MK test also estimated trends in FVI at each NTL grid location. Fig. 11 displays the regional distribution of changes in the susceptibility of Peninsular Malaysia to flooding. Only the locations that showed significant change in FVI are shown on the map. An increase in flood vulnerability over a large area in Peninsular Malaysia was noticed. Overall, flood vulnerability was increased at more grid points in the west coastal region than in other parts of the country. A significant increase in vulnerability over a large area in the northeast region is also noticed. It should be noted that there is no increase in FVI in the high FVI regions over the west coast because FVI in those urban areas is already high. The NTL in those regions reached a saturation level. Therefore, further increases in flood vulnerability in those areas could not be detected using NTL.

[Fig. 11. about here.]

5. Discussion

This study assessed the association of NTL with various socioeconomic variables identified by Ziarh et al. (2021) for estimating FVI of Peninsular Malaysia. A model was subsequently developed to estimate FVI from NTL. The study revealed a clear positive exponential correlation between NTL and population density. Conversely, a negative exponential relationship was observed between NTL and the Gini coefficient. This can be explained by the fact that higher population densities lead to more light emissions, while areas with greater income inequality, indicated by higher Gini coefficients, exhibit lower levels of NTL. Considering the geographic distribution of the population, it is evident that more people reside in coastal plains than in elevated central regions or forested areas. This results in a

negative association of NTL with elevation and forest cover, as regions with higher elevations and denser forest cover tend to emit less NTL due to lower population densities and economic activity.

This study also showed that the relationship between NTL and socioeconomic variables is nonlinear. These findings align with previous studies (Ma et al., 2014; Ma, 2018; Dickinson et al., 2020; Singhal et al., 2020; Guo et al., 2021) which demonstrated nonlinear relationships between NTL and various socioeconomic factors. Levin and Zhang (2017) highlighted the nonlinear relationship between NTL and population density, while Singhal et al. (2020) observed a complex nonlinear relationship between NTL and the Gini coefficient in India. Yuan and Chen (2023) further demonstrated a nonlinear decreasing trend of NTL with increasing forest cover. The granularity of NTL data, where each satellite pixel represents a small area, contributes to the observed non-linearity. Increased spatial heterogeneity in artificial nighttime lighting sources and human activities at fine spatial scales significantly affect NTL brightness (Ma, 2018). Variations in development, land use, and cultural practices can also influence this non-linearity (Ma et al., 2014). Therefore, as regions become more developed, NTL may not linearly correspond to economic growth. Jia et al. (2020) showed that economic and demographic factors significantly positively affect light emissions, and living standards could further increase NTL. Additionally, they demonstrated that artificial green spaces can also cause variations in light levels at the city level.

Previous studies suggest that models capable of capturing nonlinear relationships are more suitable for estimating socioeconomic variables from NTL. Wan et al. (2023) demonstrated in China that random forests can better represent the nonlinear relationship between NTL and the Human Development Index. Similarly, Castro and Alvarez (2023) employed a transfer learning algorithm for predicting socioeconomic indicators in Brazil using NTL. This study employed a single-input ANFIS model to capture the nonlinear relationships between NTL and socioeconomic variables to estimate FVI from NTL. Guo et al. (2021) also reported the potential of NTL for flood vulnerability mapping, which agrees with the findings of this study. They evaluated the performance of linear and weighted linear regressions and simple ML methods, like ANN and support vector machine (SVM), for estimating flood vulnerability from NTL. Fan et al. (2021) also adopted different ML models from NTL data to simulate a building's seismic vulnerability. Their study showed a significantly high performance of ANN and SVM compared to the conventional statistical regression method. In this study, we employed ANFIS, an advanced hybrid ML algorithm, to overcome the learning difficulties of simple ML methods by integrating fuzzy rules (Yaseen et al., 2019). Therefore, a high performance (KGE ~ 0.73) in estimating FVI from NTL was possible.

The results indicate that the ANFIS model performed well in predicting FVI, with high accuracy and reliability in capturing flood vulnerability patterns. However, the model predicted a slightly higher level of flood vulnerability than observed in areas with low FVI values. This suggests that while the model provides valuable insights into flood vulnerability trends and patterns, particularly in areas where FVI is moderate to high, it may not be as precise in predicting very low levels of vulnerability. If areas are deemed more vulnerable to flooding than they are, resources may be allocated unnecessarily for flood prevention measures such as infrastructure upgrades. This can lead to inefficient use of resources and potentially increased costs for individuals, businesses, and governments. Therefore, caution should be exercised when interpreting the model's predictions, especially for regions where the FVI is expected to be low. This overestimation of low FVI values also highlights the need for further refinement of model parameters, input data, or algorithms to improve its accuracy and reliability, particularly in capturing variations in flood vulnerability across different settings and conditions.

The present study revealed the west coastal region as the most vulnerable to floods in Peninsular Malaysia. The highest FVI values were noticed in the central-western region covering Kuala Lumpur, the capital of Malaysia, and its suburbs. Several studies identified this region as one of Malaysia's major contributors to total disaster losses (Bhuiyan and Pereira, 2018). The west coastal region of Peninsular Malaysia has experienced rapid deforestation and urbanization over the last four decades. The population and economic activities in the region increased rapidly with urbanization. The changes were more prominent in and around Kuala Lumpur, making the region most vulnerable to food (Yaakob and Masami, 2010; Shahid and Puan, 2014; Azari et al., 2022). The time series of FVI estimated in this study revealed a gradual increase in flood vulnerability in Peninsular Malaysia. The spatial distribution of the trend

showed a significant increase in vulnerability in the coastal plains of the peninsula. The study area experienced rapid economic development and population growth over the last few decades. This caused a significant expansion of settlements, mainly in the coastal plains, by clearing forests (Salim et al., 2018; Yong, 2012). Deforestation increases surface runoff and flood intensity. Therefore, deforestation, increased population density, and intense economic activities have made the area highly vulnerable to floods.

The results indicated the model's capability to estimate FVI in Peninsular Malaysia using NTL data. The model developed in this study can be used to monitor flood vulnerability in any region of Peninsular Malaysia at a resolution of NTL (0.1º×0.1º). The FVI map generated from NTL identifies areas with high flooding vulnerability based on factors such as topography, land use, population density, and other socioeconomic factors. This identification allows authorities to prioritize resources and efforts for mitigation measures in these high-risk zones. Therefore, the map can serve as a crucial decision-support tool for stakeholders, policymakers, and communities involved in flood risk mitigation efforts, enabling proactive measures to reduce vulnerability, enhance resilience, and minimize flooding impacts on human settlements and the environment. The recently available NTL products at a higher resolution (0.05º×0.05º) can be used to develop a similar model to provide a flood vulnerability assessment at a higher resolution (approximately 5 km). NTL data is available with a latency of 3 to 5 hours after acquisition; therefore, it can be used for near real-time monitoring of flood vulnerability. However, it should be noted that FVI, estimated by Ziarh et al. (2021), was used in this study. Hence, the results presented in this study reflect only socioeconomic flood vulnerability. However, the model is still important as it can estimate the changes in socioeconomic vulnerability to floods. The model can also be used to evaluate the real-time status of socioeconomic vulnerability.

The black-box nature of the model developed using machine learning algorithms limits its applicability compared to the empirical formula, which is generally developed using statistical methods. Practitioners and policymakers need understandable equations to compute flood vulnerability easily. In the future, any physical-law-based empirical model, like the symbolic regression model, can be developed for the reliable computation of flood vulnerability. The FVI estimated using socioeconomic data of a single year was used to develop the model in this study. The FVI can be estimated for other years for which statistical data are available to improve model performance. Other ML algorithms can also be attempted. In recent years, various hybrid models have been developed by integrating machine learning with optimization algorithms, which have shown excellent performance in solving highly nonlinear and complex relationships. Those can be used for a more reliable estimation of flood vulnerability from NTL data.

6. Conclusions

An ML model is developed in this study to monitor flood vulnerability using NTL data. The model performed well in estimating flood vulnerability using NTL data. This allowed for high-resolution flood vulnerability maps for different years for which NTL data is available. The study revealed significant changes in flood vulnerability in the urban areas of Malaysia. An increase in FVI was exceptionally high in the western and southern regions. Peninsular Malaysia is one of Southeast Asia's most rapidly changing regions in terms of urbanization and socio-economy. This caused a drastic increase in flood vulnerability and economic damages in the study area. The high-resolution maps of spatiotemporal changes in flood vulnerability generated in this study can mitigate flood damage and be used for policymaking. However, it should be noted that the FVI employed in this study is solely based on socioeconomic and environmental data without considering structural measures usually taken to reduce flood vulnerability. Therefore, the results presented in this study should be interpreted with caution. However, the model is still important as it can be used to evaluate the real-time status of socioeconomic vulnerability. Existing literature reviews indicate this is the first attempt to estimate spatiotemporal changes in flood vulnerability using high-resolution NTL data. In the future, the FVI map can be overlaid on flood inundation maps from different years to estimate the spatiotemporal changes in flood risk in Peninsular Malaysia. The method can be replicated in other regions for vulnerability assessment using freely available remote sensing data.

Acknowledgments

The authors thank Universiti Teknologi Malaysia (UTM) and Staffordshire University, United Kingdom, for supporting this research through the Global Challenge Research Fund (GCRF).

Funding

This study was supported by the Seoul National University of Science and Technology.

Conflict of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding authors.

References

Addison, D. M., and B. Stewart 2015. Nighttime lights revisited: the use of nighttime lights data as a proxy for economic variables. World Bank Policy Research Working Paper, (7496).

Adnan, R. M., Z. M. Yaseen, S. Heddam, S. Shahid, A. Sadeghi-Niaraki and O. Kisi 2022. *Predictability performance enhancement for suspended sediment in rivers: Inspection of newly developed hybrid adaptive neuro-fuzzy model*. International Journal of Sediment Research **37**(3): 383-398.

Azari, M., L. Billa and A. Chan 2022. *Multi-temporal analysis of past and future land cover change in the highly urbanized state of Selangor, Malaysia*. Ecological Processes **11**(1).

Bennett, M. M. and L. C. Smith 2017. *Advances in using multitemporal nighttime lights satellite imagery to detect, estimate, and monitor socioeconomic dynamics*. Remote Sensing of Environment **192**: 176-197.

Bharti, N. and A. J. Tatem 2018. *Fluctuations in anthropogenic nighttime lights from satellite imagery for five cities in Niger and Nigeria*. Scientific Data **5**.

Bhuiyan, T. R., M., Er, A. C.; Pereira, J 2018. *Direct impact of flash floods in Kuala Lumpur City: Secondary data-based analysis*. ASM Science Journal **11**: 145-157.

Bosello, F. I., A.; Termansen, M.; Jeuken, A.; Winsemius, H., De Cian, E.; . . . Garrote, L 2018. *Economy-wide impacts of climate mitigation and adaptation strategies across European regions. In Adapting to climate change in Europe*. Elsevier: 245-271.

Castro, D. A., and M.A. Álvarez 2023. Predicting socioeconomic indicators using transfer learning on imagery data: an application in Brazil. GeoJournal, 88(1), 1081-1102.

Ceola, S., F. Laio and A. Montanari 2015. *Human-impacted waters: New perspectives from global high-resolution monitoring*. Water Resources Research **51**(9): 7064-7079.

Chen, G. Y., Y. Liu, Y. J. Tian and H. Tian 2019. *Use of VIIRS DNB Satellite Images to Detect Nighttime Fishing Vessel Lights in Yellow Sea*. Proceedings of the Third International Conference on Computer Science and Application Engineering (Csae2019).

Coesfeld, J., T. Kuester, H. U. Kuechly and C. C. M. Kyba 2020. *Reducing Variability and Removing Natural Light from Nighttime Satellite Imagery: A Case Study Using the VIIRS DNB*. Sensors **20**(11).

Cooper, R. T. 2019. *Projection of future precipitation extremes across the Bangkok Metropolitan Region*. Heliyon **5**(5).

D'Ayala, D., K. Wang, Y. Yan, H. Smith, A. Massam, V. Filipova and J. J. Pereira 2020. *Flood vulnerability and risk assessment of urban traditional buildings in a heritage district of Kuala Lumpur, Malaysia*. Natural Hazards and Earth System Sciences **20**(8): 2221-2241.

de Miguel, A. S., C. C. M. Kyba, J. Zamorano, J. Gallego and K. J. Gaston 2020. *The nature of the diffuse light near cities detected in nighttime satellite imagery*. Scientific Reports **10**(1).

Dickinson, B., Ghoshal, G., Dotiwalla, X., Sadilek, A., and H. Kautz 2020. Inferring Nighttime Satellite Imagery from Human Mobility. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 34, No. 01, pp. 394-402).

Elvidge, C. D., F. C. Hsu, M. Zhizhin, T. Ghosh, J. Taneja and M. Bazilian 2020. *Indicators of Electric Power Instability from Satellite Observed Nighttime Lights*. Remote Sensing **12**(19).

Enenkel, M., R. M. Shrestha, E. Stokes, M. Román, Z. Wang, M. T. M. Espinosa, I. Hajzmanova, J. Ginnetti and P. Vinck 2020. *Emergencies do not stop at night: Advanced analysis of displacement based on satellite-derived nighttime light observations*. Ibm Journal of Research and Development **64**(1-2).

Fan, X. W., G. Z. Nie, C. X. Xia and J. X. Zhou 2021. *Estimation of pixel-level seismic vulnerability of the building environment based on mid-resolution optical remote sensing images*. International Journal of Applied Earth Observation and Geoinformation **101**.

Fang, J., C. Y. Zhang, J. Y. Fang, M. X. Liu and Y. B. Luan 2021. *Increasing exposure to floods in China revealed by nighttime light data and flood susceptibility mapping*. Environmental Research Letters **16**(10).

Feloni, E. M., I.; Baltas, E. 2020. *Flood vulnerability assessment using a GIS-based multicriteria approach*. Journal of Flood Risk Management **13**.

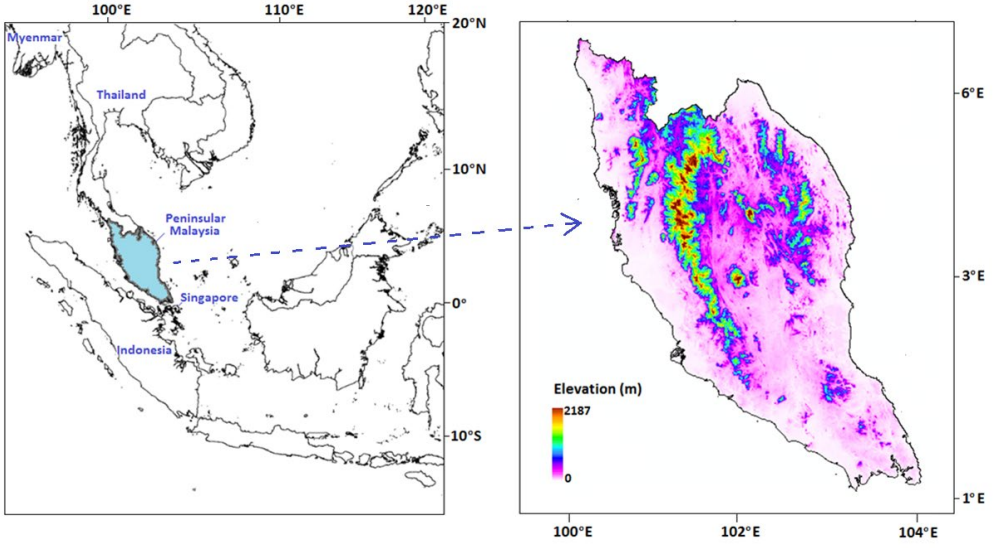
Glass, K. 2013. *Interpreting aerial photographs to identify natural hazards*. Elsevier.

Guerreiro, S. B., H. J. Fowler, R. Barbero, S. Westra, G. Lenderink, S. Blenkinsop, E. Lewis and X. F. Li 2018. *Detection of continental-scale*

intensification of hourly rainfall extremes. Nature Climate Change **8**(9): 803-+.
Guo, B. B., Y.; Zhang, D., Su, Y.; Wang, X.; Zhang, B.; . . . Luo, P. 2021. *Estimating socioeconomic parameters via machine learning methods using Luojia1-01 Nighttime Light remotely sensed images at multiple scales of China in 2018*. IEEE Access **9**: 34352-34365.
Hirschmann, R. 2020. *Travel and tourism in Malaysia-Statistics & Facts. In: Statista.*
Ibrahim, N. F., Zardari, N. H., Shirazi, S. M., Haniffah, M. R. B. M., Talib, S. M., Yusop, Z., & Yusoff, S. M. A. B. M. 2017. *Identification of vulnerable areas to floods in Kelantan River sub-basins by using flood vulnerability index*. GEOMATE Journal **12(29)**: 107-114.
Jang, J. S. R. 1993. *Anfis - Adaptive-Network-Based Fuzzy Inference System*. Ieee Transactions on Systems Man and Cybernetics **23**(3): 665-685.
Jia, T., K. Chen, and X. Li 2020. Exploring the factors controlling nighttime lights from prefecture cities in Mainland China with the hierarchical linear model. Remote Sensing, 12(13), 2119.
Jiang, Y. L., C. E. Metz and R. M. Nishikawa 1996. *A receiver operating: Characteristic partial area index for highly sensitive diagnostic tests*. Radiology **201**(3): 745-750.
Keating, A., K. Campbell, R. Mechler, P. Magnuszewski, J. Mochizuki, W. Liu, M. Szoenyi and C. McQuistan 2017. *Disaster resilience: what it is and how it can engender a meaningful change in development policy*. Development Policy Review **35**(1): 65-91.
Keola, S., M. Andersson and O. Hall 2015. *Monitoring Economic Development from Space: Using Nighttime Light and Land Cover Data to Measure Economic Growth*. World Development **66**: 322-334.
Kohavi, R. 1995. *A study of cross-validation and bootstrap for accuracy estimation and model selection*. Paper presented at the Ijcai.
Kohiyama, M., H. Hayashi, N. Maki and M. Higashida 2004. *Early damaged area estimation system using DMSP-OLS nighttime imagery*. International Journal of Remote Sensing **25**(11): 2015-2036.
Lees, K. J., R. R. E. Artz, D. Chandler, T. Aspinall, C. A. Boulton, J. Buxton, N. R. Cowie and T. M. Lenton 2021. *Using remote sensing to assess peatland resilience by estimating soil surface moisture and drought recovery*. Science of the Total Environment **761**.
Levin, N., and Q. Zhang, 2017. A global analysis of factors controlling VIIRS nighttime light levels from densely populated areas. Remote sensing of environment, 190, 366-382.
Li, X. M., Y. Y. Zhou, G. R. Asrar, M. Imhoff and X. C. Li 2017. *The surface urban heat island response to urban expansion: A panel analysis for the conterminous United States*. Science of the Total Environment **605**: 426-435.
Li, Y. Z., J. Y. Li, A. Xu, Z. Z. Feng, C. J. Hu and G. S. Zhao 2021. *Spatial-Temporal Changes and Associated Determinants of Global Heating Degree Days*. International Journal of Environmental Research and Public Health **18**(12).
Liu, J. and D. Niyogi 2019. *Meta-analysis of urbanization impact on rainfall modification*. Scientific Reports **9**.
Long, F. J., L. F. Zheng and Z. D. Song 2018. *High-speed rail and urban expansion: An empirical study using a time series of nighttime light satellite data in China*. Journal of Transport Geography **72**: 106-118.
Ma, T 2018. Multi-level relationships between satellite-derived nighttime lighting signals and social media–derived human population dynamics. Remote sensing, 10(7), 1128.
Ma, T., Zhou, Y., Wang, Y., Zhou, C., Haynie, S., & Xu, T. (2014). Diverse relationships between Suomi-NPP VIIRS night-time light and multi-scale socioeconomic activity. Remote Sensing Letters, 5(7), 652-661.
Mård, J., G. Di Baldassarre and M. Mazzoleni 2018. *Nighttime light data reveal how flood protection shapes human proximity to rivers*. Science Advances **4**(8).
McDermott, T. K., Michaels, G., & Rauch, F 2015. *CEP Discussion Paper No 1398 December 2015 Flooded Cities Adriana Kocornik-Mina.* .
Miller, J. D. and M. Hutchins 2017. *The impacts of urbanisation and climate change on urban flooding and urban water quality: A review of the evidence concerning the United Kingdom*. Journal of Hydrology-Regional Studies **12**: 345-362.
Mohanty, M. P. and S. P. Simonovic 2021. *Understanding dynamics of population flood exposure in Canada with multiple high-resolution population datasets*. Science of the Total Environment **759**.
Muhammad, M. K. I., Shahid, S., Ismail, T., Harun, S., Kisi, O., and Z. M. Yaseen 2021. The development of evolutionary computing model for simulating reference evapotranspiration over Peninsular Malaysia. Theoretical and applied climatology, 144, 1419-1434.
Nashwan, M. S., S. Shahid, E. S. Chung, K. Ahmed and Y. H. Song 2018. *Development of Climate-Based Index for Hydrologic Hazard Susceptibility*. Sustainability **10**(7).
Nayak, P. C., K. P. Sudheer, D. M. Rangan and K. S. Ramasastri 2004. *A neuro-fuzzy computing technique for modeling hydrological time series*. Journal of Hydrology **291**(1-2): 52-66.
OECD 2016. *Financial management of flood risk*. In: OECD Publishing Paris, France.
Opperman, J. G., G. E.; Duvail, S. 2013. *The multiple benefits of river-floodplain connectivity for people and biodiversity.* In: Academic Press.
Pour, S. H., A. K. Abd Wahab, S. Shahid and Z. Bin Ismail 2020. *Changes in reference evapotranspiration and its driving factors in peninsular Malaysia*. Atmospheric Research **246**.
Qiang, Y., Q. X. Huang and J. W. Xu 2020. *Observing community resilience from space: Using nighttime lights to model economic disturbance and recovery pattern in natural disaster*. Sustainable Cities and Society **57**.
Romali, N. S. and Z. Yusop 2021. *Flood damage and risk assessment for urban area in Malaysia*. Hydrology Research **52**(1): 142-159.
Salim, J. M. R., M. A.; Razali, S. M.; Cooke, F. M. 2018. *Coastal Landscapes of Peninsular Malaysia: The Changes and Implications for Their Resilience and Ecosystem Services*. In Landscape Reclamation-Rising From What's Left: IntechOpen.
Sen, P. K. 1968. *Estimates of Regression Coefficient Based on Kendalls Tau*. Journal of the American Statistical Association **63**(324): 1379-&.
Shahid, S. M., A.; Puan, O. C. 2014. *Assessment of greenhouse gas emission reduction measures in transportation sector of Malaysia*. Jurnal teknologi **70(4)**.
Singhal, A., Sahu, S., Chattopadhyay, S., Mukherjee, A., and S. N. Bhanja 2020. Using night time lights to find regional inequality in India and its relationship with economic development. PloS one, 15(11), e0241907.
Soltani, K., I. Ebtehaj, A. Amiri, A. Azari, B. Gharabaghi and H. Bonakdari 2021. *Mapping the spatial and temporal variability of flood susceptibility using remotely sensed normalized difference vegetation index and the forecasted changes in the future*. Science of the Total Environment **770**.
solubility in potassium and sodium based amino acid Salt solutions. Journal of Environmental Chemical Engineering **7**(1).
Soroush, E., M. Mesbah, N. Hajilary and M. Rezakazemi 2019. *ANFIS modeling for prediction of CO*
Tangang, F. T. 2001. *Low frequency and quasi-biennial oscillations in the Malaysian precipitation anomaly*. International Journal of Climatology

21(10): 1199-1210.
Tian, J. R., N. Z. Zhao, E. L. Samson and S. L. Wang 2014. *Brightness of Nighttime Lights as a Proxy for Freight Traffic: A Case Study of China.* Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing **7**(1): 206-212.
Wan, N., Du, Y., Liang, F., Yi, J., Qian, J., Tu, W., and S. Huang 2023. Nighttime light satellite images reveal uneven socioeconomic development along China's land border. Applied Geography, 152, 102899.
Yaakob, U. M., T.; Masami, F. 2010. *Ninety years of urbanization in Malaysia: a geographical investigation of its trends and characteristics.* . Journal of Ritsumeikan Social Sciences and Humanities **4(3)**: 79-101.
Yang, D., W. X. Luan, L. Qiao and M. Pratama 2020. *Modeling and spatio-temporal analysis of city-level carbon emissions based on nighttime light satellite imagery*. Applied Energy **268**.
Yaseen, Z. M., W. H. M. W. Mohtar, A. M. S. Ameen, I. Ebtehaj, S. F. M. Razali, H. Bonakdari, S. Q. Salih, N. Al-Ansari and S. Shahid 2019. *Implementation of Univariate Paradigm for Streamflow Simulation Using Hybrid Data-Driven Model: Case Study in Tropical Region*. Ieee Access **7**: 74471-74481.
Yong, D. 2012. *Massive deforestation in southern Peninsular Malaysia driving ecological change in Singapore*. Nature in Singapore **5**: 285-289.
Yuan, Y., and Z. Chen 2023. The impacts of land cover spatial combination on nighttime light intensity in 2010 and 2020: a case study of Fuzhou, China. Computational Urban Science, 3(1), 5.
Yue, S., P. Pilon, B. Phinney and G. Cavadias 2002. *The influence of autocorrelation on the ability to detect trend in hydrological series*. Hydrological Processes **16**(9): 1807-1829.
Zhang, J., J. L. Zhou, G. Q. Zhang, Y. Y. Ji, Y. Y. Zeng, W. Fan and A. Aikelamu 2021. *Climate- and human-driven variations in lake area and number in North Xinjiang, China*. International Journal of Remote Sensing **42**(2): 469-485.
Zhao, M., Y. Y. Zhou, X. C. Li, W. T. Cao, C. Y. He, B. L. Yu, X. Li, C. D. Elvidge, W. M. Cheng and C. H. Zhou 2019. *Applications of Satellite Remote Sensing of Nighttime Light Observations: Advances, Challenges, and Perspectives*. Remote Sensing **11**(17).
Ziarh, G. F., M. Asaduzzaman, A. Dewan, M. S. Nashwan and S. Shahid 2021. *Integration of catastrophe and entropy theories for flood risk mapping in peninsular Malaysia*. Journal of Flood Risk Management **14**(1).

Figure



(a) (b)

Fig. 1. Peninsular Malaysia's (a) position and (b) geography in the Southeast Asian region

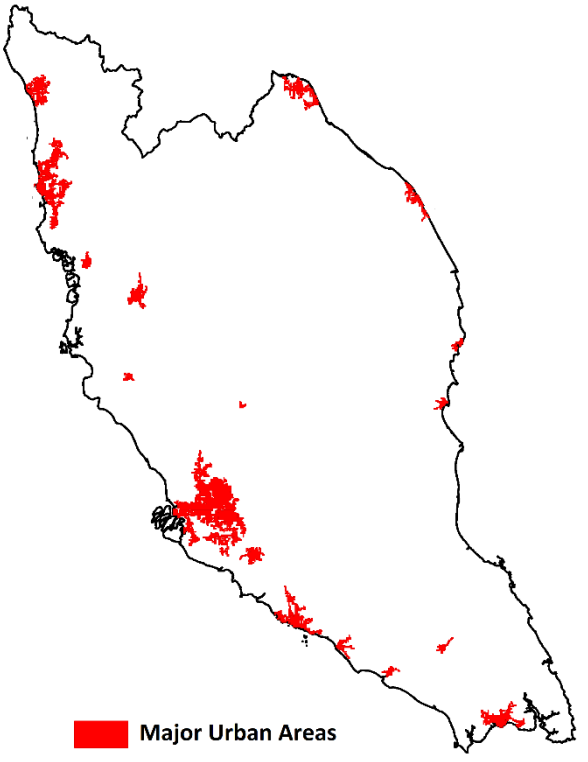


Fig. 2. Major urban centers of Peninsular Malaysia.

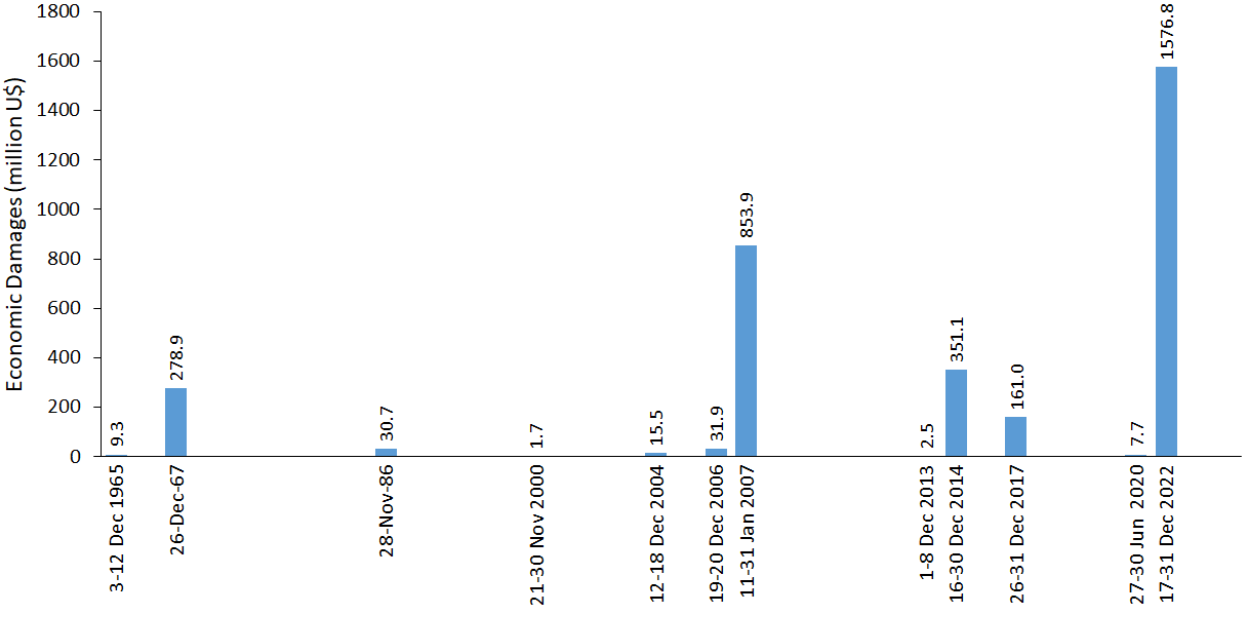


Fig. 3. Economic damage (U$) from floods and post-flood-related damages in Peninsular Malaysia during 1965-2022 (EM-DAT, 2020).

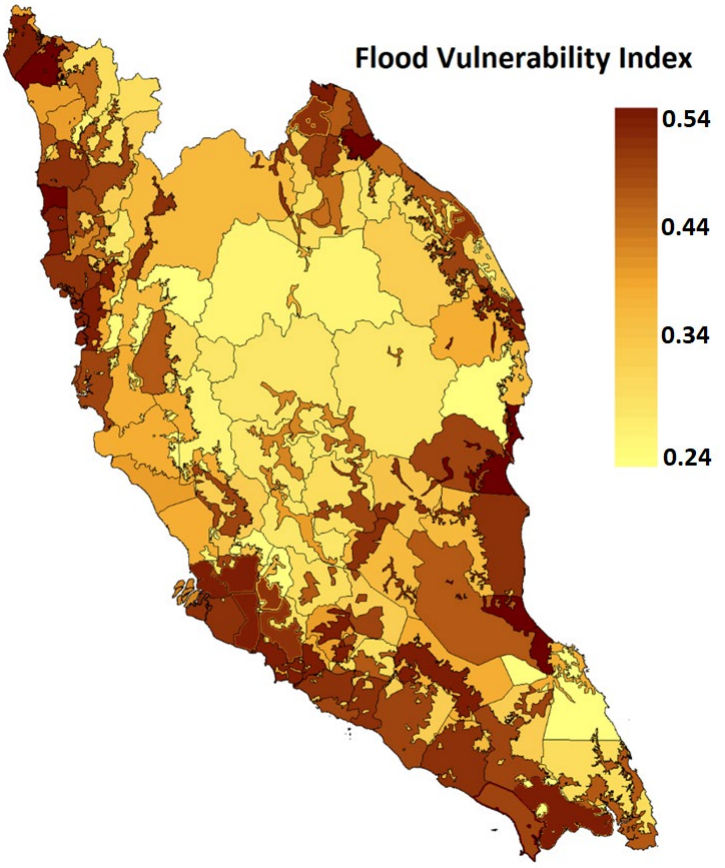


Fig. 4. Flood vulnerability index (FVI) map of Peninsular Malaysia (after Ziarh et al., 2021)

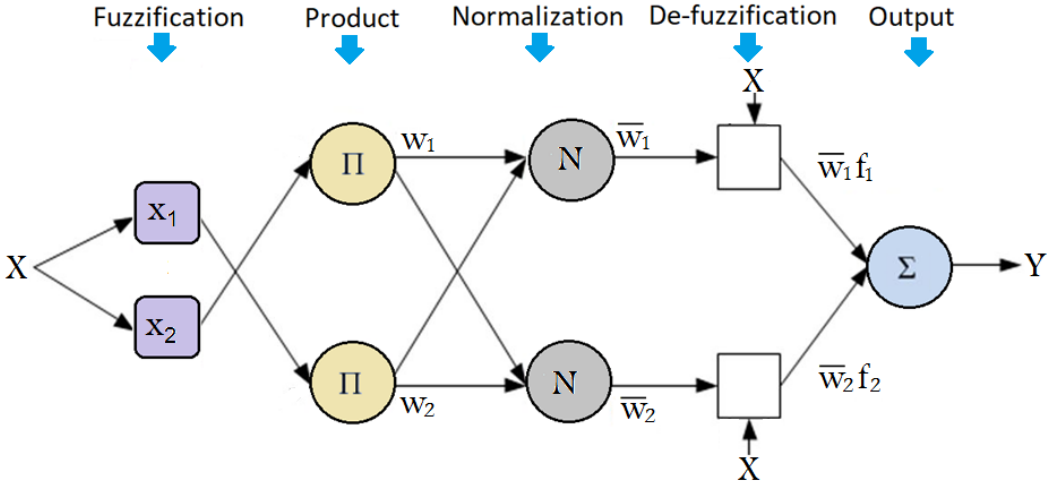


Fig. 5. A typical illustration of a single ANFIS model

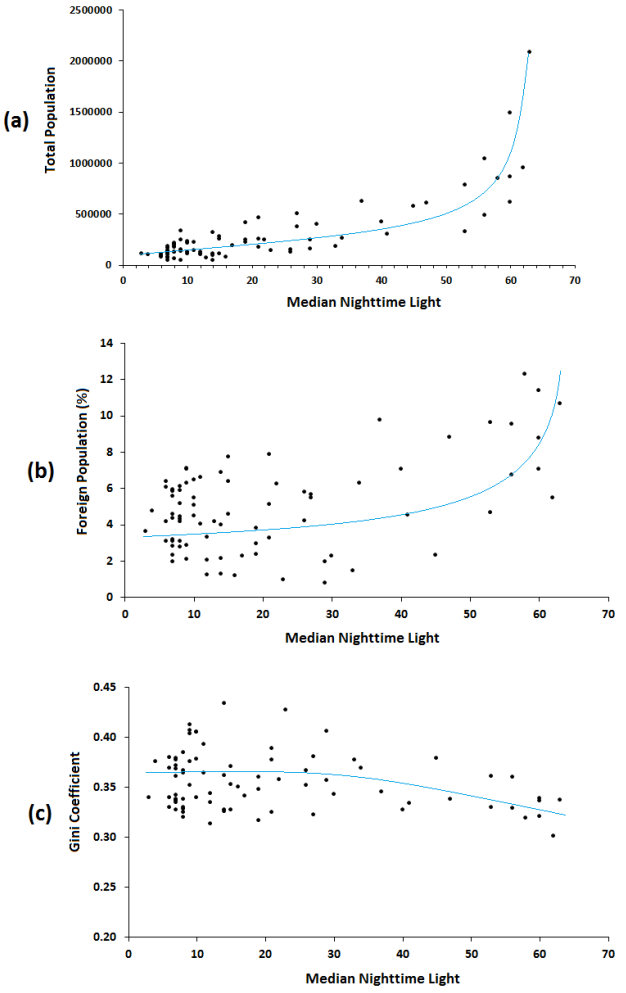
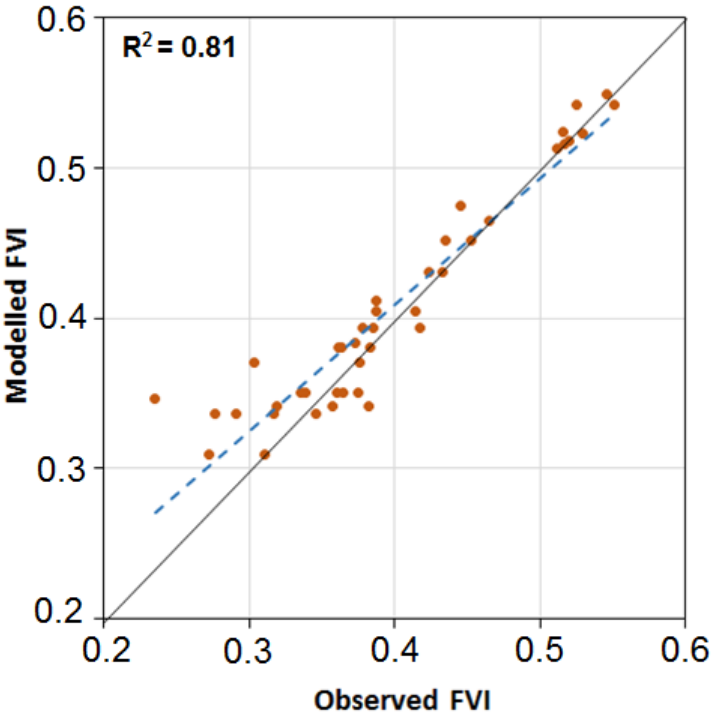


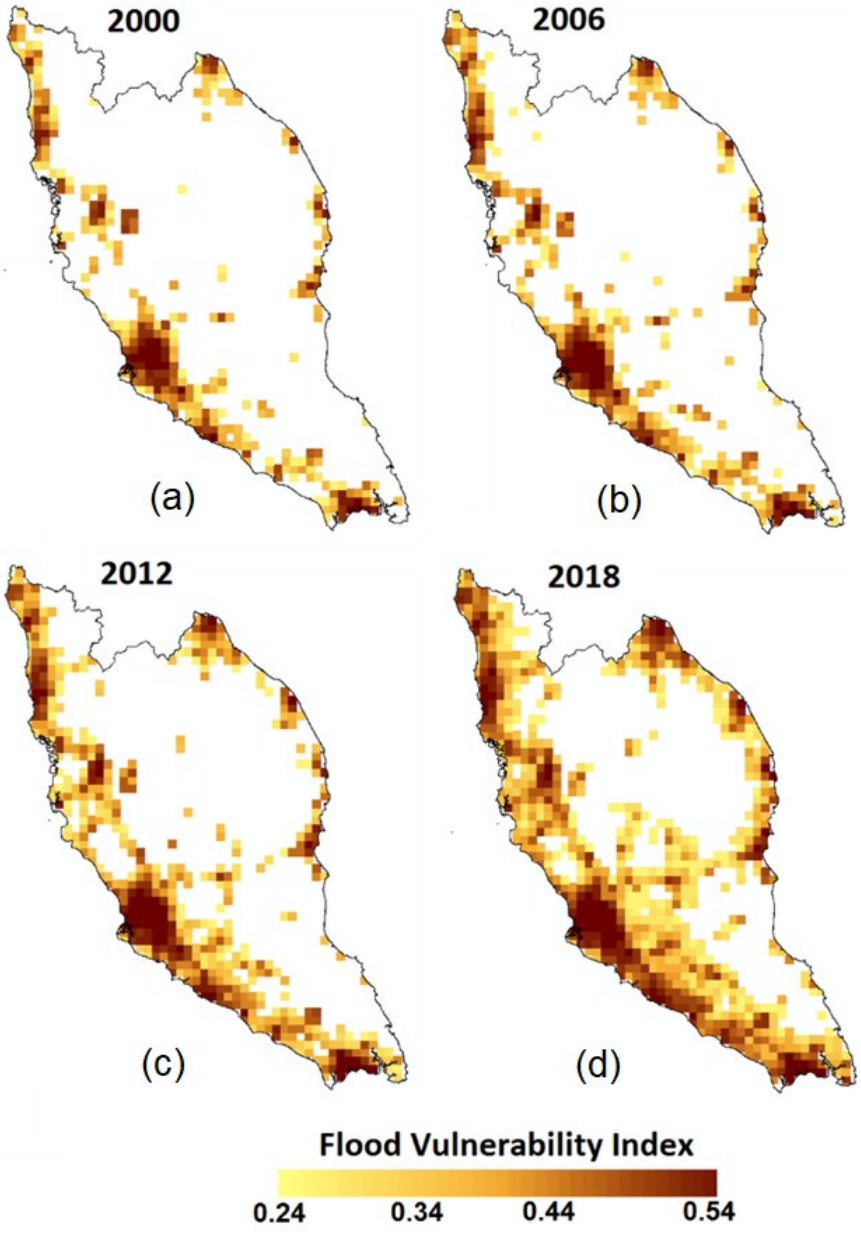
Fig. 6. The relationship of nighttime light with (a) the total population, (b) the percentage of the foreign population to the total population, and (c) the Gini coefficient for the year 2016 at the administrative unit scale.



Metics	Calibration	Validation
NRMSE	0.004	0.007
md	0.884	0.768
KGE	0.858	0.732

513

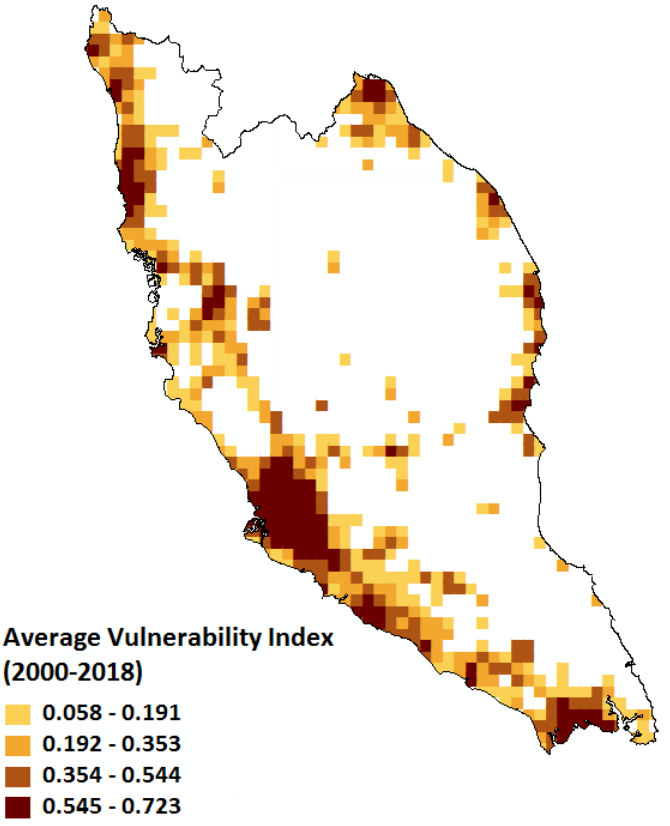
514 Fig. 7. Scatter plot of flood vulnerability index estimated using survey data (observed) and vulnerability index estimated from NTL data via
515 ANFIS approach.



516

517 Fig. 8. Geographic distribution of flood vulnerability metric during (a) 2002, (b) 2006, (c) 2012, and 2018.

518



519

520 Fig. 9. Spatial distribution of average flood vulnerability index in Peninsular Malaysia for the period 2000-2018

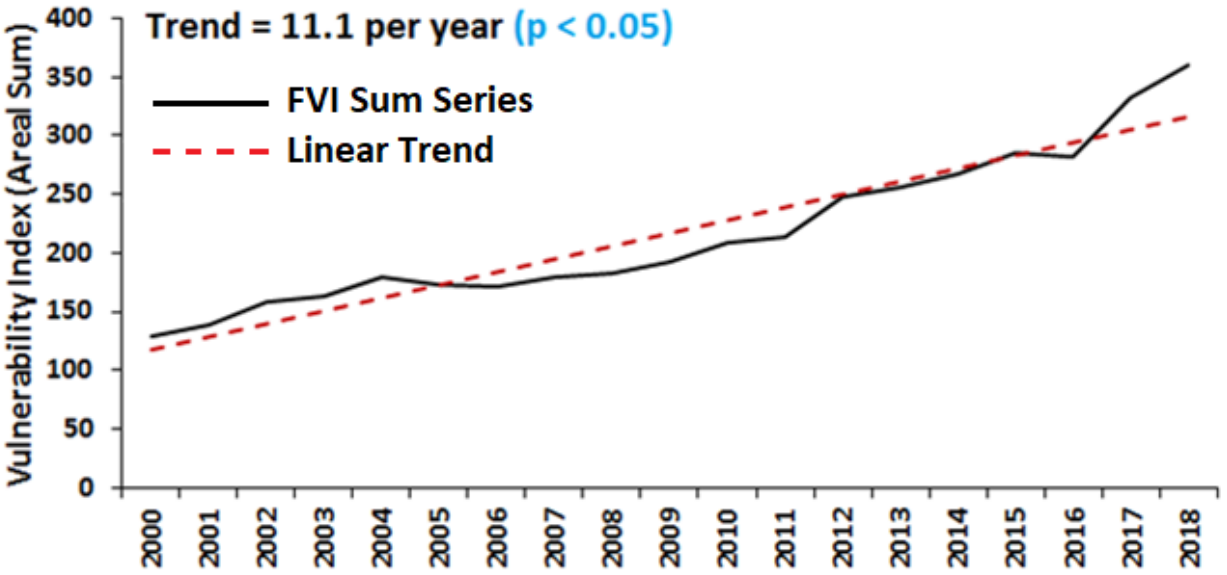


Fig. 10. Trends in flood vulnerability index for Peninsular Malaysia

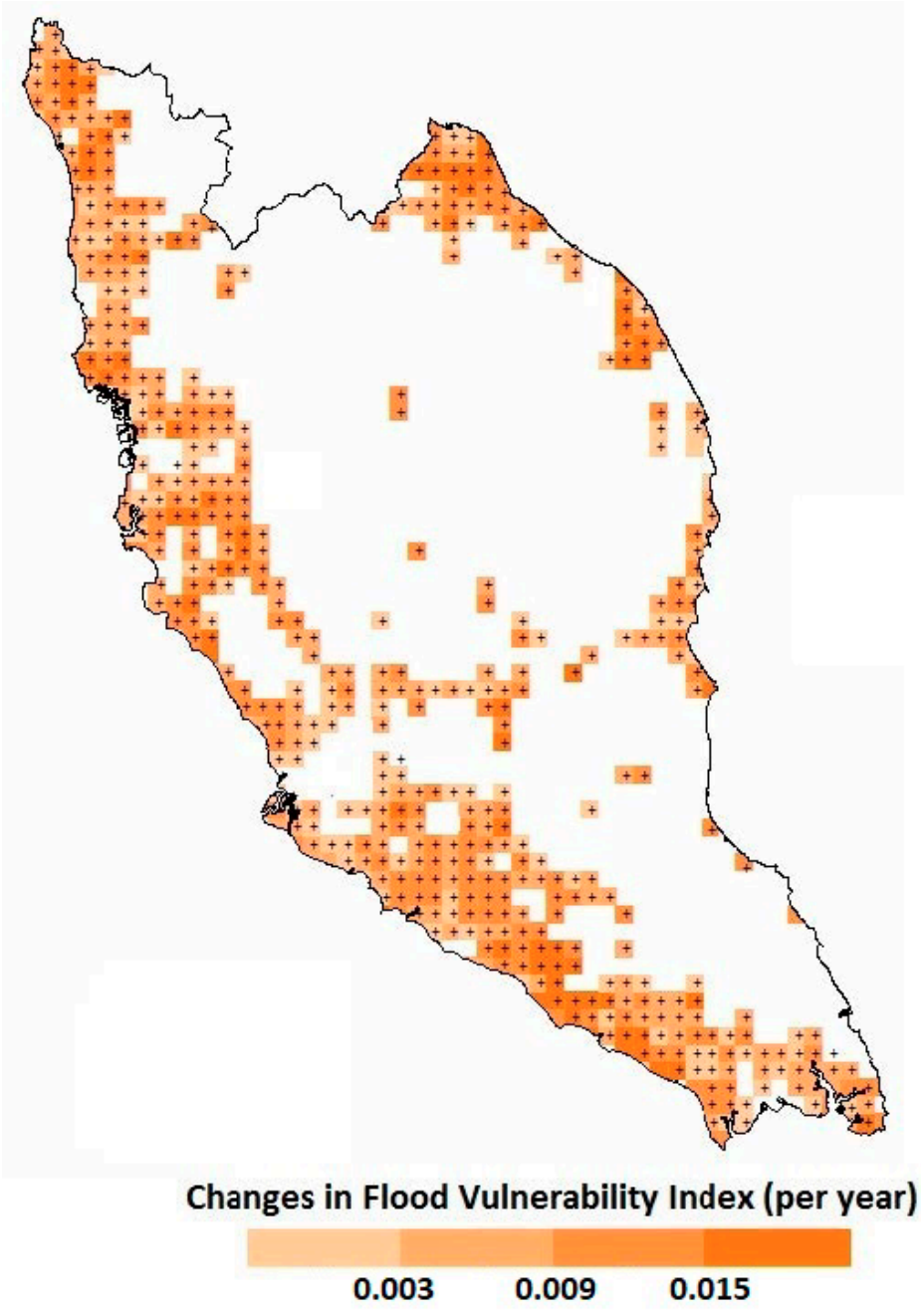


Fig. 11. Spatial distribution of changes (per year) in flood vulnerability index in Peninsular Malaysia during 2000-2018. The color ramp shows the magnitude of change, while the symbol "+" indicates a significant increase in FVI