THE ESTIMATION OF FLOOD AREA BASED ON A FEW SELECTED AND WEIGHTED PARAMETERS: CASE STUDY OF THE NANGKA RIVER BASIN, BALIKPAPAN (INDONESIA)

Totok Sulistyo[1]*, Sara Respati[1]

[1]Balikpapan State Polytechnic, Department of Civil Engineering, Balikpapan, Indonesia; e-mails: totok.sulistyo@poltekba.ac.id; sara.wibawaning@poltekba.ac.id

Abstract: In several previous studies on flood analysis and estimation, there was no clear rationale for why different researchers used a different combination of parameters in the determination of flood zones. Such research results raise the question of how to select a few dominant parameters without reducing the objectivity of the analysis. This research proposes the standardization of parameters selection by using Pareto Analysis in screening a few vital flood parameters from numerous parameters that prevail in certain areas. The selection of the right dominant parameters is the key to achieving the analysis goal and it will also simplify the analysis processes. This flood zone estimation study uses a combination of Pareto Analysis, Analytic Hierarchy Process (AHP) and Geographic Information System (GIS). The results of the study include a flood zonation map. The study area can be classified by its level of vulnerability as follow: very low vulnerability zones (0.003 km^2), low vulnerability zones (5.588 km^2), medium vulnerability zones (11.876 km^2), high vulnerability zones (8.629 km^2), and very high vulnerability zones (2.198 km^2). The validation shows that the estimation of the most vulnerable zone is consistent with field validation and the flood event history of several locations in the study area. As a result, the developed model can provide an accurate flood zonation map, enabling stakeholders to take appropriate mitigation measures for different areas.

Keywords: flood parameters; Pareto Analysis; AHP; GIS; river basin

1. Introduction

The interaction of rainwater and the complex natural and anthropogenic phenomena on the earth in certain areas can cause flood disasters. Therefore, the interaction between rainwater and all aspects of the earth's surface should be modeled so that floods can be estimated and mapped. Flood monitoring based on remote sensing data and the delineation of flood risk zone for large areas are increasingly getting interest in disaster mitigation (Notti et al., 2018). Such mapping and monitoring are based merely on serial satellites, and not all required satellite images are available in any area on the globe; so, more alternative flood mapping methods and algorithms should be developed using other terrestrial data (Yoshida et al., 2022), or combination of terrestrial and aerial data.

---

*Corresponding author, e-mail: totok.sulistyo@poltekba.ac.id

Many researchers have conducted flood estimations and mapping using various methods and parameters. However, the selection of parameters was mostly based on the researcher's personal judgment, where selected parameters varied among each researcher. They used between seven and twelve selected parameters in their research (Das & Gupta, 2021; Handini et al., 2021; Jahangir et al., 2019; Mujib et al., 2021; Nsangou et al., 2022; Sinha et al., 2008). The reasons why some parameters were taken and others were neglected, were mostly left unclear and with no further explanation. On the other hand, taking an abundance of parameters as input can cause model overfitting. This raises the question of how to create a procedure in selecting the right dominant parameters without reducing the quality and accuracy of the analysis.

To address this issue, Pareto Analysis can be used in the selection of the parameters. Then the weighting of parameters can be determined by combining the Analytic Hierarchy Process (AHP) method and the Geographical Information System (GIS; Boulomytis et al., 2019; Das & Gupta, 2021; Nsangou et al., 2022). The latest technology in remote sensing such as the combination of optical and radar satellites is available. This technology offers better accuracy in flood mapping (Li et al., 2022). Moreover, Copernicus Open Access Hub and Google Earth Engine also give chances to researchers to process the satellite's image quickly to generate flood zonation with high accuracy to support decision-making (Uddin & Matin, 2021). To make an accurate estimation of flood zonation, artificial intelligence, especially Artificial Neural Networks (ANN), has also been used to model the flood accurately (Jahangir et al., 2019).

This research aims to simplify the creation of flood-prone zonation. The flood is mapped without reducing its accuracy by selecting a few dominant parameters and reducing selection bias. The study is conducted in the Nangka River Basin. The mapping process uses Pareto Analysis to select the dominant parameters of the flood in the research area. AHP is then applied to weigh the parameters to represent their roles in flood disasters, and GIS is used to overlay the raster layers. The combination of Pareto Analysis, AHP, and GIS was used to predict flood zonation with only a few selected essential parameters. The novelty of this research is the parameter selection based on Pareto Analysis, which is less subjective compared to the previous research. Pareto Analysis has an assumption that handling 20% of the most significant parameters of flood in the research area will resolve 80% of others. This study proposes the standardization of flood parameter selection for future analyses. Other implications of this study will facilitate the mitigation of flood hazards in the study area and in other areas as well. Furthermore, it presents reasonable and simplified data and procedures for estimating flood hazards by using only one-fifth of the important factors without reducing the quality of the output.

2. Methods and data

This study employs Pareto Analysis and AHP method in GIS environment in aseries of stages to estimate flood vulnerability. Flood events are influenced by a multitude of interacting factors, and their relative importance can vary over places. Therefore, weighting the parameters is crucial. The following paragraphs describe the stages of the study, which are also illustrated in Figure 1.

In the first stage, the study selects a location that meets several criteria. The primary and secondary data areas must be accessible for the research so that further stages can be accomplished. The primary data include stone and soil variation and texture that are collected from the study area. The data are used to validate the geological formation.

The secondary data include flood events and their history such as the flood location, its severity, its frequency. Impacts on the surrounding area are also collected.

The second stage of the study involves collecting the experts' perspectives regarding corresponding aspects that cause floods on the study area. The aim of this stage is to obtain their common knowledge and judgment on the factors that cause floods in the area rather than just based on the researcher's perspective. Accurate prior knowledge of the physical characteristics of the susceptible flood area from various sources is crucial for flood mapping (Sami et al., 2020), and the involvement of stakeholders in hazard assessment and zonation can improve the effectiveness of the end products (Grigorecu et al., 2021). Therefore, by involving experts in this stage, the study aims to gather accurate and relevant information for subsequent analysis. In this study, the data are collected through

Figure 1. Methodological flowchart in estimation of flood.

questionnaires and personal discussions with hydrologists, civil engineers, geologists, hydraulic engineers, and policymakers who have knowledge of the study area. The perspective of 28 respondents was collected. Most of the respondents are over 35 years old. Eighteen respondents are lecturers, three respondents are bureaucrats and decision-makers, and seven respondents are company workers.

The third stage is to analyze the results of the questionnaire about the factors that induce flood in the perspective of respondents and discussions to rank the parameters based on their scores. The top 20% of parameters are chosen using Pareto Analysis. The Pareto law constitutes an 80/20 ratio of the problem where 20% of top problem solutions will accomplish 80% of the problem (Erdil, 2019). Pareto Analysis has been broadly used in various subject areas because it enables risk estimation and estimation toward failure prevention (Can & Erbiyik, 2016).

The fourth stage is to collect secondary data from the City Planning Office, including geological and land use maps with a scale of 1:100,000 and 1:50,000 (Balikpapan City Government, 2021; Land and Spatial Arrangement Office Balikpapan City, 2018). The DEM Copernicus 30 dataset (GLO-30) of the study area is downloaded from EO Browser (European Space Agency, n.d.). In the fifth stage, QGIS software is used for photogrammetric processing of the dataset to extract river networks, slope maps, and elevation, and to delineate the river basin border based on surface flow direction. A zonation of distance from the river is generated using the buffering feature, and all layers are saved in raster format for further GIS overlay operation. The land use and geological maps were digitized, and various layers are classified based on the highest and lowest values. Risk values are assigned based on the
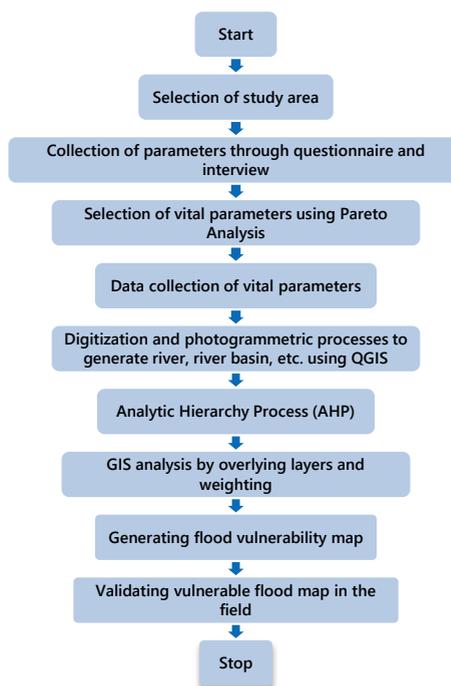
Table 1. Risk rank of layer classification

Rank	Risk	Pixel value	Representative color
1	Very low	10	Blue
2	Low	20	Green
3	Medium	30	Yellow
4	High	40	Orange
5	Very high	50	Red

contribution of each parameter, and the classification consists of five clusters of risk values (one to five), except for the rainfall layer, which is classified with one risk value because only one weather station is available. The geologic layer is classified into two classes because it is constructed by two formations. The rank, risk, pixel value, and representative color are detailed and described in Table 1.

The sixth stage determines the weight of layers using AHP by determining the comparison pairwise matrix, where the matrix size is the sum of vital aspects resulting from Pareto Analysis, and determines the Maximum Eigen Value, Consistency Index (*CI*), Eigen Vector (*w*), Random Consistency Index (*RI*), and Consistency Ratio (*CR*). This stage can be accomplished using Equations 1–6 (Brunelli, 2015; Saaty, 1987). This process can be done using online AHP software and a spreadsheet. AHP is employed in this research as its famous method has been proven to be successfully implemented to solve the probability of susceptibility to various natural hazards (Sinha et al., 2008).

The step of using AHP starts with the following pairwise comparison matrix A with element matrix a_{ij} and w_i/w_j is vector ratio in Equation 1 (Brunelli, 2015):

$$A = (a_{ij})_{n \times n} \text{ where } a_{ij} \cong \frac{w_i}{w_j} \tag{1}$$

the elements of the matrix are vectors ratio that is determined based on the fundamental scale for Pairwise Comparison: 1 is *equal importance*, 3 is described as *moderate importance*, 5 means *strong importance*, 7 is *very strong importance*, 9 is *extreme importance*, 2, 4, and 6 are intermediate values between the two adjacent judgments (Saaty, 1987).

After replacement element matrix a_{ij} with vector ratio w_i/w_j, follow Equation 2 matrix *A* multiplied by priority vector to get *Aw* (Brunelli, 2015).

$$A_w = A \times w = \begin{bmatrix} w_1/w_1 & \cdots & w_1/w_j & \cdots & w_1/w_n \\ \vdots & & \vdots & & \vdots \\ w_i/w_1 & \cdots & w_i/w_j & \cdots & w_i/w_1 \\ \vdots & & \vdots & & \vdots \\ w_n/w_1 & \cdots & w_n/w_j & \cdots & w_n/w_n \end{bmatrix} \begin{bmatrix} w_1 \\ \vdots \\ w_j \\ \vdots \\ w_n \end{bmatrix} = n \begin{bmatrix} w_1 \\ \vdots \\ w_j \\ \vdots \\ w_n \end{bmatrix} \tag{2}$$

Then, *n* and *w* are Eigen value and Eigen vector of *A* where the Maximum Eigen Value (λ_{max}) should satisfy Equation 3 (Brunelli, 2015; Saaty, 1987):

$$Aw = \lambda_{max} w \tag{3}$$

CI can be calculated using Equation 4 (Brunelli, 2015; Saaty, 1987):

$$CI = \frac{\lambda_{max} - n}{n-1} \tag{4}$$

where *n* is number of elements.

According to Saaty (1987), Random Consistency Index (*RI*) for different matrix size (*n*): *n* = 3 is 0.52, *n* = 4 is 0.89, *n* = 5 is 1.11, *n* = 6 is 1.25, *n* = 7 is 1.35, *n* = 8 is 1.40, *n* = 9 is 1.45.

Then *CR* can be counted using Formula 5 (Brunelli, 2015; Saaty, 1987):

$$CR = \frac{CI}{RI} \qquad (5)$$

The model will be accepted if *RI* meets this criterion (Brunelli, 2015; Saaty, 1987):

$$CR \leq 0.1 \qquad (6)$$

The seventh stage is the overlying process. Here, the layers are weighted using Eigen Vector that is obtained from the AHP model (*CR* ≤ 10%) resulting from the previous stage. This stage is accomplished using the raster calculators feature in QGIS to generate zonation of flood-prone areas from its calculation. Then the last stage of the research activity is the validation of the vulnerable flood zonation by checking the study area by its flood history and interviewing local inhabitants.

3. Study area

The study area for this research is the Sungai Nangka River Basin, located in Balikpapan City, East Borneo Indonesia (Figure 2). The river basin covers an area of approximately 28.294 km^2 and is situated in the middle of the city and central business district of Balikpapan.
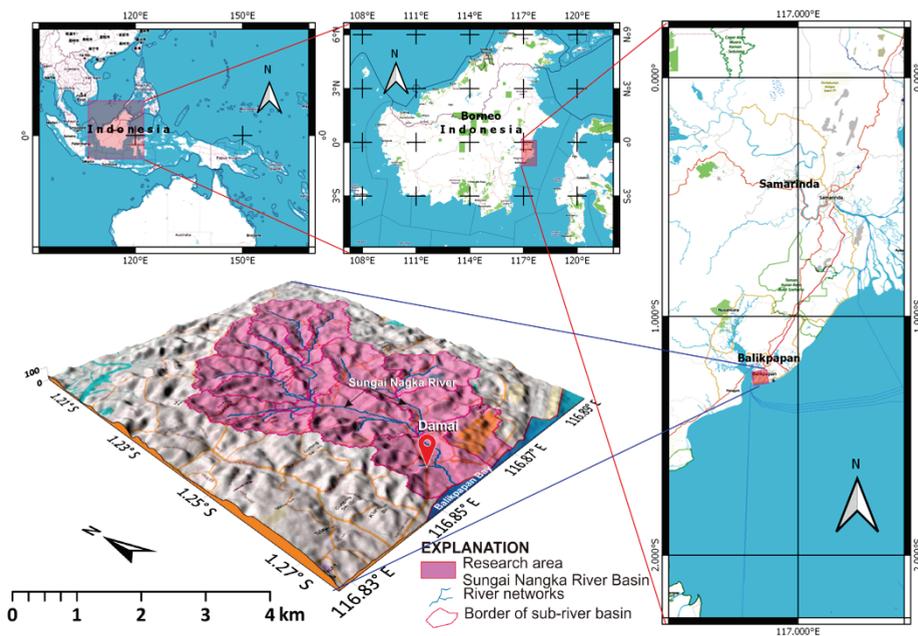

Figure 2. Location and morphological 3D view of research area.

During the rainy season, flooding is a frequent occurrence in some parts of the river basin, leading to submerged housing complexes and roads. This annual disaster causes significant economic losses and traffic disruptions. The local government has allocated funds

to mitigate flood disasters by improving the drainage system, increasing road elevation, and dredging riverbed sediments. However, the flood problem has worsened over time due to population and economic growth. This research focuses on the watershed system where rainwater is collected and flows downstream of the Sungai Nangka River before ultimately reaching the sea of Makassar Strait. To provide a better understanding of the terrain and relief of the research area, a 3D morphological view is presented in Figure 2, marked location is Damai, the central business district in the area of study.

4. Results and discussion

The results of a questionnaire and the professional judgment of experts who possess extensive knowledge about the physical geographical features of the research area can be represented in a Pareto Diagram. This diagram highlights the top 20% of vital parameters that should be considered in flood estimation. These critical parameters include the distance from the river, elevation, land use, rainfall, river basin, geology, and slope. Other parameters are not considered in the estimation. Figure 3 shows the results of the Pareto Analysis used to select these vital parameters. Referring to Figure 3, the role of each selected parameter concerning the study area are further explained in fallowing subsection.

4.1. Selected parameters

4.1.1. Distance from the river

Considering the density of the river network, 28.744 km/km^2 (very fine), flood plain width of the morphologic cross-section of the rivers, the author proposes the following criterion of the scale of risk for the distance from the center of the river as follows: within 0–100 m is a very high risk (rank 5), 100–200 m is a high risk (rank 4), 200–300 m is a medium risk (rank 3), 300–400 m is a low risk (rank 2), and more than 400 m is a very low risk (rank 1). This criterion is only appropriate for this river basin.
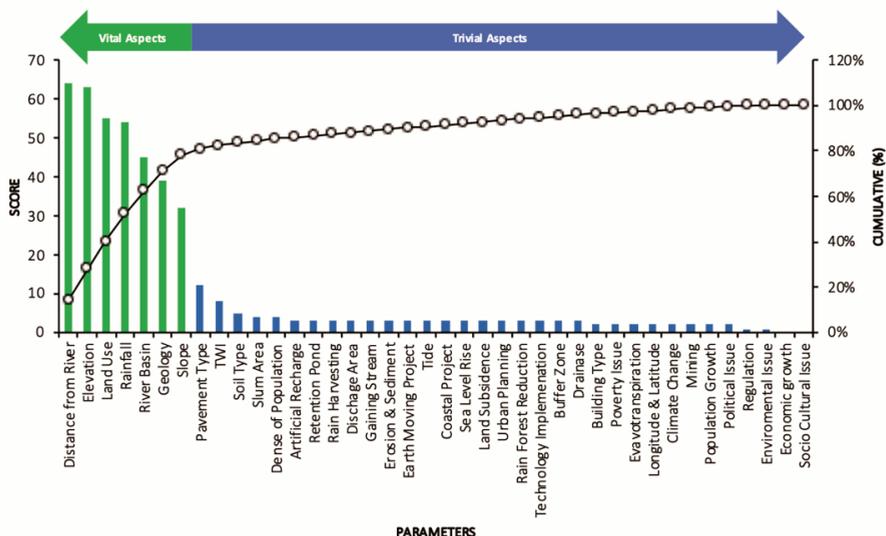

Figure 3. Pareto chart to determine 20% of the most vital aspects.

4.1.2. Elevation parameter

The elevation of the study area ranges between 1 and 51 m a.s.l. To assess the risk of flooding based on elevation, it can be classified into five risk ranks. Areas with the lowest elevation, between 1 and 11 m, are typically found in the valley and downstream of the river. These areas have the highest risk rank of 5 out of 5 (frequently inundated by level flood reaching 1–3 m). The other elevation levels between 11 and 41 m are classified into three risk ranks of 4, 3, and 2 (frequently reported inundated by flood 0.1 up to 1 m). Conversely, areas with the highest elevation level, more than 41 m, are located at the top of the hilly morphology and have the lowest risk rank of 1 out of 5 (never reported to be inundated by flood). The elevation range of such classification was based on the minimum and maximum elevation and the history of flood in several locations that are collected by the Balikpapan City Government (2021) and summarized in the map. This classification can be used to identify areas at higher risk of flooding and inform the development of flood management strategies. This classification is uniquely made for this river basin.

4.1.3. Land use parameter

The housing complex area is dominant in the land use around the Sungai Nangka River Basin and the rest of the area is green open space (Land and Spatial Arrangement Office Balikpapan City, 2018). The layer of land use is obtained by digitizing Land Use Map from Urban Planning Office.

In the land use aspect, dense housing areas and central business districts have the highest risk rank of 5 out of 5. Dense housing areas without green space in suburban parts have a risk rank of 4. Non-dense housing areas with green space such as bushes, city forest, mixed plantation, and other agricultural fields have a lower risk of 2 up to 3. Green open spaces with fewer or without buildings have the lowest risk rank of 1 out of 5. In the areas of central business, the land is covered by buildings, concrete, and asphalt causing an increment in impervious surface area (Zhu et al., 2019). As a result, the ground's ability to absorb water through infiltration is decreased. Meanwhile, the green open space also has less impact rather than the housing area when a flood disaster hits the area of study.

4.1.4. River basin and rainfall parameters

The Sungai Nangka River Basin is around 28.294 km2, which is a small part of Balikpapan City (512.25 km2). It consists of 12 sub-river basins. The river basin layer is generated from DEM after extracting river networks by using QGIS. This river basin acts as a catchment area that collects rainwater which then goes through the next water cycle. The last order of the sub-river basin of the Sungai Nangka River downstream has the highest risk level as it collects runoff from others. So the bigger order of sub-river basin has a greater risk rank than the smaller order of sub-river basin which just collects rainwater from its own area. Therefore, the biggest order of sub-river basins in the downstream of the river has the highest risk rank of 5 out of 5 and the upstream sub-river basin has the lowest risk rank of 1 out of 5.

There is only one weather station available for Balikpapan City. Therefore, this study uses the rainfall rate data from the weather station that recorded the average rainfall rate of 3098.45 mm/year in the period 2009–2021. The highest monthly average of rainfall was in June with 338.91 mm/month (Statistics of Balikpapan Municipality, 2010; 2015; 2022). The rainfall layer has only one risk level due to the one location of the source of data, so differentiation of rainfall in the study area cannot be done. The risk level of rainfall

parameter was set uniformly 5 out of 5 by taking the annual rainfall rate and, the highest daily rainfall which reached 105.2 mm in March 2014 (Statistics of Balikpapan Municipality, 2015), and the relatively small area of study (28.294 km^2) into consideration. The root cause of the flood in the study area is actually the rainfall parameter, Without the role of rainfall in the study area, the other parameters will not contribute to the flood event.

4.1.5. Geological parameter

The soil in the study area is dominated by fine cohesionless quartz sandstone as part of the Balikpapan Formation and claystone shale, and siltstone marl, and sandy claystone as part of the Kampung Baru Formation. The geological structures that were found minorly in the area are faults and joints. The area is situated in the limb of a gentle plunging anticline (Sulistyo, 2012; Sulistyo & Abrar, 2017). The lithology of Kampung Baru Formation has a higher risk rank 2 than Balikpapan Formation risk rank 1 because Kampung Baru Formation is dominated by sandy claystone, marl, and siltstone (Sulistyo, 2003), which provide no interconnected porosity which the infiltration will pass through. The textures of that formation make the formation relatively impermeable. Meanwhile, Balikpapan Formation is dominated by quartz sandstone which provides interconnected porosity as the conduit through which rainwater infiltrates downward to the ground. Geological structures in the area consist of gentle dips, with minor joints and faults (Hidayat & Umar, 1994) that seemingly have a minor contribution to the flood hazard. The previous research shows that the infiltration rate in a few locations of the Balikpapan Formation is relatively faster than in a few locations of the Kampung Baru Formation (Juwita & Santoso, 2019). However, the dominant geological effect of the flood hazard is the lithological textures such as grain size and porosity through which the surface water flows to the ground as infiltration or remains on the surface as runoff.

4.1.6. Slope parameter

The slope of the area is mostly hilly morphology where a small part of it is rugged surface (Jordan et al., 2021). The slope area has a minimum value of 0.16% and a maximum value of 21.22%. The highest slope value has the lowest risk value (1), whereas the minimum slope value has the highest risk value (5). The dominant slope in the area is between 12% and 21.22% with medium (3) to high importance value (4). This steep to moderate slope occupies 77% of the area of study.

4.2. Risk Analysis

The above description of the risk levels was applied to the parameters and gave the result as 7 parameter layers which are shown in Figure 4. Then the area of the level of risk in each parameter can be further known based on pixel count, as listed in Table 5.

Table 5. The level of risk area of each parameter

Level of risk	Distance from the river	Elevation	Land use	Rainfall	Sub-basin	Geology	Slope
	km^2	km^2	km^2	km^2	km^2	km^2	km^2
1	11.989	5.224	0.634	0.000	0.000	15.120	0.076
2	2.837	10.074	17.135	0.000	17.623	13.173	1.028
3	3.547	10.555	0.078	0.000	4.753	0.000	5.395
4	4.390	2.235	0.532	0.000	0.355	0.000	12.665
5	5.531	0.206	9.915	28.294	5.563	0.000	9.130
Total area	28.294	28.294	28.294	28.294	28.294	28.294	28.294

The next step is to determine the weight of each layer using AHP. The process starts with the arrangement of a pairwise comparison matrix. The matrix elements are arranged based on the comparison between each parameter referring to Equation 2. The pairwise comparison matrix is shown in Table 3.

Table 3. Pairwise comparison matrix

Parameter	Distance from river	Elevation	Land use	Rainfall	Catchment	Geology	Slope
Distance from river	1	1	3	3	5	7	9
Elevation	1	1	1	3	5	7	9
Land Use	1/3	1	1	1	3	5	7
Rainfall	1/3	1/3	1	1	3	5	7
Catchment	1/5	1/5	1/3	1/3	1	3	5
Geology	1/7	1/7	1/5	1/5	1/3	1	3
Slope	1/9	1/9	1/7	1/7	1/5	1/3	1

Table 3 shows the importance of one parameter compared to the others. For example, if we look at the first row and then compare the first column and last column, the Distance from the river parameter is nine times more important than Slope parameter.

The pairwise comparison matrix and matrix size as the inputs are processed using online AHP Calculation software by CGI (n.d.). Then, the calculation is conducted following the AHP steps as presented in Equation 1 to Equation 6. The outputs summary of the AHP Calculation is as follows:

- λ_{max} = 7.328,
- CI = 0.054,
- RI of matrix size 7 × 7 = 1.35, and
- CR = 0.044, which is less than 0.1, therefore, the AHP model is accepted. The Eigen Vector (w) or weight of each parameter is shown in Table 4.

In preparation for the seven parameters selected by the Pareto Analysis, photogrammetric and digitizing processes were used to create raster values in QGIS. Each layer was then classified according to its risk rank using Table 1 as a reference for reclassifying pixel values and colors. All layers were classified based on their respective attributes and pixel values and assigned to the same class. For example, the elevation layer with a range of 1 to 11 m a.s.l., was assigned the highest risk rank of 5, and given a red color with a pixel value of 50. Similarly, the distance from the river center layer with a range of 0 to 100 m was also assigned the highest risk rank of 5, and given a red color with a pixel value of 50.

This study proposes Equation 7 for flood zonation pixels (fzp) or the sum of the parameter (P_i) times the weight (w_i) in Table 4 as follows:

$$fzp = \sum P_i \cdot w_i \qquad (7)$$

The prepared layers and their risk classifications are shown in Figure 4, which includes the Distance from the river layer, Elevation layer, Land use layer, Sub-river basin layer, and Slope layer. Each layer was classified into five clusters based on the range between their respective maximum and minimum values.

Table 4. Eigen Vector (w)

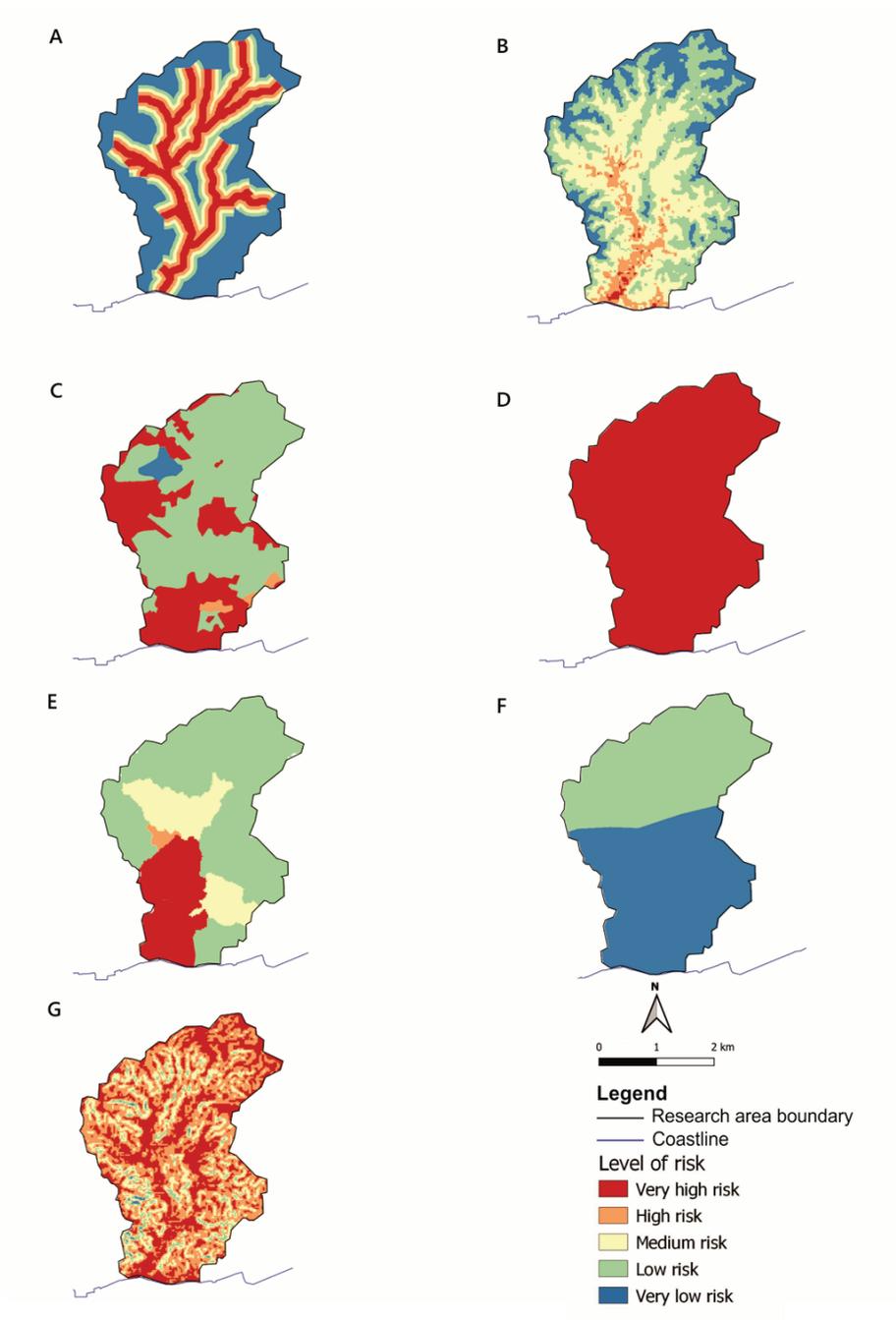| Parameters (P)	Weights (w)
Distance from river (P_1)	0.308
Elevation (P_2)	0.264
Land use (P_3)	0.163
Rainfall (P_4)	0.140
Sub-river basin (P_5)	0.068
Geology (P_6)	0.036
Slope (P_7)	0.021

Figure 4. GIS layers consist of Distance from river (A), Elevation (B), Land use (C), Rainfall (D), Sub-River basin (E), Geology (F), and Slope (G).

The Land use layer was defined based on the contribution of each element to the flood and their respective risk ranks. It was classified into five clusters, starting from the lowest to the highest risk: city forest, bush and agriculture, public open space and sports field, swamp, lake, housing, and building. The Rainfall layer had only one cluster as it was limited to data from one weather station. The Geological layer was divided into two clusters based on the two geological formations of the watershed system: Balikpapan and Kampung Baru.

All the required layers are arranged in the QGIS project layer manager, then their pixel values are weighted and summated based on Equation 7. The following pseudocode in the raster calculator is imposed to get *fzp*:

fzp : ("P_1" * 0.308) + ("P_2" * 0.264) + ("P_3" * 0.163) +
("P_4" * 0.140) + ("P_5" * 0.068) + ("P_6" * 0.036) + ("P_7" * 0.021)

The output-weighted summation of pixel value from seven layers is a new raster (*fzp*) of the vulnerable flood zonation as shown in Figure 5. The flood zonation is also overlaid on the 3D view to enhance the readability (Figure 5). Based on the pixel values, the flood zones can be classified as very low vulnerability zone (the area of 0.003 km^2), low vulnerability zone (the area of 5.588 km^2), medium vulnerability zone (the area of 11.876 km^2), high vulnerability zone (the area of 8.629 km^2), and very high vulnerability zone (the area of 2.198 km^2). Most of the very high vulnerability zone is located close to the downstream of the Sungai Nangka River. For further analysis such as the impact on the transportation system and to facilitate validation of the map in the field, the flood zonation map is overlaid with road networks.
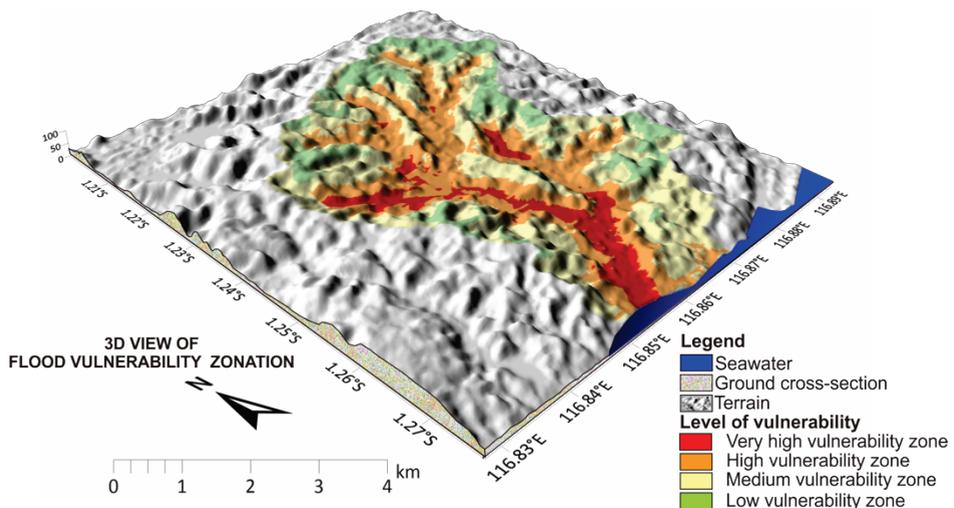

Figure 5. 3D view of flood vulnerability zonation.

To evaluate the flood vulnerability, random points in the map are visited and validated based on the flood history. This validation is conducted by checking the record in a local newspaper and interviewing the local inhabitants that live around the point of validation to get descriptions of flood history in that area. The location of field validation is at MT Haryono Street which is located at the longitude is 116°52'1"E, and a latitude is 1°15'1"S. Morphologically, the validation site is situated in the valley (Figure 6).
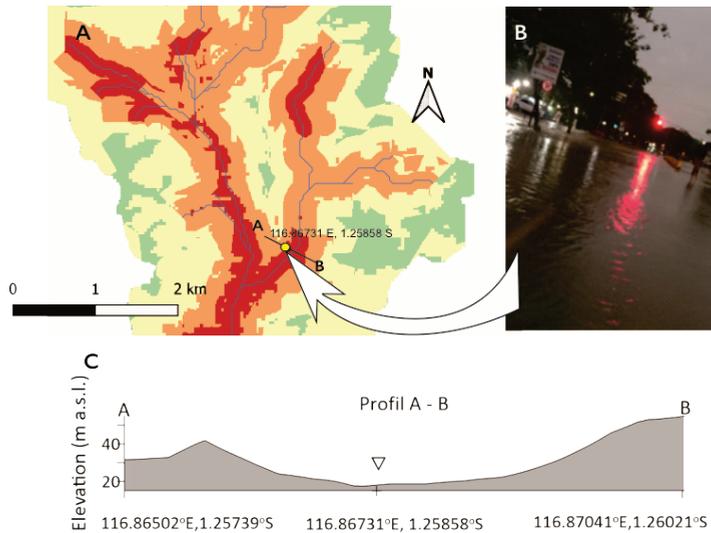
Figure 6. Validation was conducted in the field at MT.

Note. Panel A: Haryono Street South Balikpapan (shown by the yellow dot), the longitude is 116.86731° E, and the latitude is 1.25858° S (A). Panel B: Photo courtesy of local inhabitants (Anonymous), taken on Wednesday, March 16, 2022, 06.07 Local Time. Panel C: Topographical Profile of validation location.

According to the interviewees, the level of flood reached 0.5 m to 1 m. Based on other information from national daily news, the level of flood on March 16, 2022 reached 2 m and it was proclaimed to be the worst flood in the last decade (Sucipto, 2022). Another record of flood history from both national and local daily news was on August 25, 2022, and this flood inundated International Airport of Sultan Aji Muhammad Sulaiman Sepinggan and the parking lot of Balikpapan Super Block the biggest shopping mall in Balikpapan City located in part of the downstream of the Sungai Nangka River Basin.

5. Discussion

In this study, the perspectives from 28 respondents show that there are 40 possible aspects influencing the flood phenomenon in the study area. Moreover, 7 vital aspects are further selected and analyzed using AHP. The parameter of the Distance from the river has the highest weight (0.308), followed by Elevation (0.264), Land use (0.164), Rainfall (0.140), Sub-River Basin (0.068), Geology (0.036), and Slope (0.021) as listed in Table 4. Summation weighting pixel is the model interaction between each parameter in which it has a range of value between 10 and 50. After summation and weighting of pixels, they were reclassified using the following pseudocode:

If (fzp≤15) then { reclassify fzp to 10; }
Else If (fzp>15) or (fzp≤25) then { reclassify fzp to 20; }
Else If (fzp>25) or (fzp≤35) then { reclassify fzp to 30; }
Else If (fzp>35) or (fzp≤45) then { reclassify fzp to 40; }
Else then { reclassify fzp to 50; }

The result of the reclassified rasters can be classified as a flood-vulnerable zone based on Table 1. The result of pixel values summation in every single location of pixel (fzp) is the representation of the level vulnerability zone of flood in the study area.

The result of flood vulnerability zonation shows that a very highly vulnerable zone is 2.198 km^2. The major of the area of the very high vulnerable zone consists of 1.124 km^2 located downstream of the river basin, and the rest of 1.074 km^2 is located in the middle and upstream area. Such a result is proven visually by Figure 5 where the locations of the very highly vulnerable zone are valleys or flood plain. Otherwise, low vulnerability zones are the areas with high elevation where, geomorphologically, there are roughed hilly terrains or slopes that are laid extensively at the periphery of the river basin and upstream area. The 3D visualization of flood zonation in Figure 5 helps in assessing the accuracy and precision of the estimation of flood vulnerability zonation visually as it naturally has a correlation to the geomorphology.

The flood vulnerability zonation shows that the area along the river basin poses the highest risk to be affected by flooding. The rainfall data show that the daily rainfall rate has reached 105.2 mm/day in a couple of hours, which causes inundation in the very highly flood-prone zones that cover housing areas, the main road (MT. Haryono Street), and the Central Business District. The overlay of the road layer on the flood vulnerability map implies information on the density of housing and buildings. The more severe flood can reoccur in the middle and downstream if the aggravated buffer area is not managed wisely. The high density of road networks, at the same time, also has a high density of houses or buildings. The government could mitigate the disaster by relocating the buildings and inhabitants in the red zone to a safer location to lessen the impact of the flood. The buffer area and flood plain of the river should be free of buildings and protected as green open space.

The other mitigation measure that could be taken is developing retention ponds in the upstream sub-river basin The steep slopes and river basin areas upstream must be protected by strict laws and regulations from deforestation and illegal housing. This protection is important to keep the natural retention of rainwater so the sudden peak of velocity and discharges of runoff can be avoided, and it will flow downstream stably. Vegetation loss, the increasingly paved terrain, and the growth of dense building areas have reduced the natural infiltration and the natural retention capacity. Such worsening conditions cause a sudden peak of water discharges that threaten the downstream of the river basin with severe flooding. The improvement of the flood estimation zonation model needs local government involvement in providing more weather stations and hydrological stations at some representative points, especially in very highly flood-prone zones. Those additions of more weather stations to record daily rainfall should also be a priority of government planning, taking into consideration that Balikpapan City has 512.25 km^2 (Statistics of Kalimantan Timur Municipality, 2015) with its flood disaster history and, ironically the city only has one weather station.

6. Conclusion

This research proposes a combination of Pareto Analysis and AHP in GIS environment to estimate flood zones by an objective selection of vital parameters among numerous and complex interacting parameters in the real world. The proposed method is tested to estimate flood vulnerability in the Nangka River Basin. The result shows that the method can accurately estimate the flood zone as the very highly vulnerable zone is confirmed by the records of flood history in the study area and from field validation. The validation shows that the result of the estimation using seven vital parameters has the highest accuracy in the study area.

The limitation of this research is the lack of rainfall data, so there is no differentiation of rainfall rate in each pixel of the rainfall layer. On the other hand, rainfall is the main parameter of flood phenomena in the study area. This limitation will possibly reduce its accuracy, especially in the area of the upstream river basin that is situated far away from the location of rainfall measurement. For future research, combining the selection input method of Pareto analysis with other algorithms such as Artificial Neural Networks, Fuzzy Logic, Deep Learning, and Machine Learning Regression Models, etc. is possible.

References

AHP Calculation software by CGI [Computer software] (n.d.). Retrieved May 26, 2022, from http://www.isc.senshu-u.ac.jp/~thc0456/EAHP/AHPweb.html

Balikpapan City Government. (2021). *Peraturan Daerah Kota Balikpapan Nomor 6 Tahun 2021 Tentang Rencana Pembangunan Jangka Menengah Daerah Tahun 2021-2026* [Balikpapan City Regional Regulation Number 6 of 2021 concerning Regional Medium-Term Development Plans for 2021-2026]. http://bappedalitbang.balikpapan.go.id/assets/globalimg/RPJMD_Kota_Balikpapan_Tahun_2021-_2026.pdf

Boulomytis, V. T. G., Zuffo, A. C., & Imteaz, M. A. (2019). Detection of flood influence criteria in ungauged basins on a combined Delphi-AHP approach. *Operations Research Perspectives, 6*, Article 100116. https://doi.org/10.1016/j.orp.2019.100116

Brunelli, M. (2015). *Introduction to the Analytic Hierarchy Process*. Springer International Publishing. https://doi.org/10.1007/978-3-319-12502-2

Can, E., & Erbiyik, H. (2016). Determination of the Risks That Are Emerged from the Use of Aerial Photographs in the Mapping Activities for Landslide Movements by FMEA and Pareto Analysis Methods and Suggested Solutions. *Procedia Engineering, 161*, 850–858. https://doi.org/10.1016/J.PROENG.2016.08.729

Das, S., & Gupta, A. (2021). Multi-criteria decision based geospatial mapping of flood susceptibility and temporal hydro-geomorphic changes in the Subarnarekha basin, India. *Geoscience Frontiers, 12*(5), Article 101206. https://doi.org/10.1016/j.gsf.2021.101206

European Space Agency. (n.d.). *DEM Copernicus 30* (GLO-30) [Data set]. https://doi.org/10.5270/ESA-c5d3d65

Erdil, A. (2019). An Evaluation on Lifecycle of Products in Textile Industry of Turkey through Quality Function Deployment and Pareto Analysis. *Procedia Computer Science, 158*, 735–744. https://doi.org/10.1016/j.procs.2019.09.109

Grigorecu, I., Sima, M., Dragotă, C.-S., & Dumitrică, C. (2021). Climate hazard maps-a tool to support communication with stakeholders. A showcase of Baia Mare depression, Romania. *Romanian Journal of Geography, 65*(2), 105–123. http://www.rjgeo.ro/issues/revue%20roumaine%2065_2/grigorescu%20et%20al..pdf

Handini, D. R., Hidayah, E., & Halik, G. (2021). Flash Flood Susceptibility Mapping at Andungbiru Watershed, East Java Using AHP-Information Weighted Method. *Geosfera Indonesia, 6*(2), 127–142. https://doi.org/10.19184/geosi.v6i2.24173

Hidayat, S., & Umar, I. (1994). *Peta Geologi Lembar Balikpapan, Kalimantan* [Geological Map of the Balikpapan Quadrangle, Kalimantan]. https://geologi.esdm.go.id/geomap/pages/preview/peta-geologi-lembar-balikpapan-kalimantan

Jahangir, M. H., Mousavi Reineh, S. M., & Abolghasemi, M. (2019). Spatial predication of flood zonation mapping in Kan River Basin, Iran, using artificial neural network algorithm. *Weather and Climate Extremes, 25*, Article 100215. https://doi.org/10.1016/j.wace.2019.100215

Jordan, N. A., Sherlia, & Syafitri, E. D. (2021). Building age profile and figure-ground image: Defining the urban development pattern of Balikpapan City. *IOP Conference Series: Earth and Environmental Science, 778*(1), Article 012038. https://doi.org/10.1088/1755-1315/778/1/012038

Juwita, R., & Santoso, I. B. (2019). Assessment of Soil Infiltration Capability in Balikpapan City. *IPTEK Journal of Proceedings Series, 5*, 291–297. https://iptek.its.ac.id/index.php/jps/article/view/6341

Land and Spatial Arrangement Office Balikpapan City. (2018). *Peta Tutupan Lahan* [Land Use Map]. https://data.balikpapan.go.id/dataset/tutupan-lahan

Li, C., Dash, J., Asamoah, M., Sheffield, J., Dzodzomenyo, M., Gebrechorkos, S. H., Anghileri, D., & Wright, J. (2022). Increased flooded area and exposure in the White Volta river basin in Western Africa, identified from multi-source remote sensing data. *Scientific Reports, 12*(1), Article 3701. https://doi.org/10.1038/S41598-022-07720-4

Mujib, M. A., Apriyanto, B., Kurnianto, F. A., Ikhsan, F. A., Nurdin, E. A., Pangastuti, E. I., & Astutik, S. (2021). Assessment of Flood Hazard Mapping Based on Analytical Hierarchy Process (AHP) and GIS: Application in Kencong District, Jember Regency, Indonesia. *Geosfera Indonesia, 6*(3), 353–376. https://doi.org/10.19184/geosi.v6i3.21668

Notti, D., Giordan, D., Caló, F., Pepe, A., Zucca, F., & Galve, J. P. (2018). Potential and Limitations of Open Satellite Data for Flood Mapping. *Remote Sensing, 10*(11), Article 1673. https://doi.org/10.3390/rs10111673

Nsangou, D., Kpoumié, A., Mfonka, Z., Ngouh, A. N., Fossi, D. H., Jourdan, C., Mbele, H. Z., Mouncherou, O. F., Vandervaere, J.-P., & Ndam Ngoupayou, J. R. (2022). Urban flood susceptibility modelling using AHP and GIS approach: case of the Mfoundi watershed at Yaoundé in the South-Cameroon plateau. *Scientific African, 15*, Article e01043. https://doi.org/10.1016/j.sciaf.2021.e01043

Saaty, R. W. (1987). The analytic hierarchy process—what it is and how it is used. *Mathematical Modelling, 9*(3–5), 161–176. https://doi.org/10.1016/0270-0255(87)90473-8

Sami, G., Hadda, D., Mahdi, K., & Abdelwahhab, F. (2020). A Multi-Criteria Analytical Hierarchy Process (AHP) to Flood Vulnerability Assessment in Batna Watershed (Algeria). *Analele Universităţii Din Oradea, Seria Geografie, 30*(1), 41–47. https://doi.org/10.30892/auog.301105-810

Sinha, R., Bapalu, G., Singh, L., & Rath, B. (2008). Flood risk analysis in the Kosi river basin, north Bihar using multi-parametric approach of Analytical Hierarchy Process (AHP). *Journal of the Indian Society of Remote Sensing, 36*, 335–349. https://doi.org/10.1007/s12524-008-0034-y

Statistics of Balikpapan Municipalities. (2015). *Balikpapan In Figures 2015.* https://balikpapankota.bps.go.id/publication/2016/10/25/d04948397272b54a8d9cc93c/kota-balikpapan-dalam-angka-2015.html

Statistics of Balikpapan Municipality. (2010). *Balikpapan In Figures 2010.* https://balikpapankota.bps.go.id/publication/2016/01/06/d29f7695a456e7339c77de0e/balikpapan-dalam-angka-2010.html

Statistics of Balikpapan Municipality. (2022). *Kota Balikpapan Dalam Angka 2022* [Balikpapan Municipality In Figures 2022]. https://balikpapankota.bps.go.id/publication/2022/02/25/e78628bde521d84e52633df5/kota-balikpapan-dalam-angka-2022.html

Statistics of Kalimantan Timur Municipality. (2015). *Luas Wilayah Menurut Kabupaten/Kota* [Total Areas of Regencies and Cities 2015] https://kaltim.bps.go.id/indicator/153/74/1/luas-wilayah-menurut-kabupaten-kota.html

Sucipto, O. (2022, March 16). Banjir Rendam Balikpapan, Terparah dalam 10 Tahun Terakhir [The Worst Flood Inundation in Balikpapan in Last Decades]. *Kompas Daily News.* https://www.kompas.id/baca/nusantara/2022/03/16/banjir-di-balikpapan-genangan-tersebar-di-seluruh-kecamatan

Sulistyo, T. (2003). *Study on Slope Stability in Vulnerable landslide Area for Evaluation of General City Spatial Arrangement Plan in South Balikpapan.* Universitas of National Development "Veteran".

Sulistyo, T. (2012). Study on Slope Stability in Vulnerable Landslide Area for Evaluation of General City Spatial Arrangement Plan in South Balikpapan. *Jurnal Ilmiah Politeknik, 4*(1), 1–7. https://jurnal.poltekba.ac.id/index.php/JIP/article/view/71/841

Sulistyo, T., & Abrar, A. (2017). Characterization of Thin Alluvial Bed Aquifers in Manggar River Balikpapan East Kalimantan Indonesia. *Jurnal Teknologi Terpadu, 5*(1), 54–62. https://doi.org/10.32487/jtt.v5i1.212

Uddin, K., & Matin, M. A. (2021). Potential flood hazard zonation and flood shelter suitability mapping for disaster risk mitigation in Bangladesh using geospatial technology. *Progress in Disaster Science, 11*, Article 100185. https://doi.org/10.1016/j.pdisas.2021.100185

Yoshida, K., Pan, S., Taniguchi, J., Nishiyama, S., Kojima, T., & Islam, M. T. (2022). Airborne LiDAR-assisted deep learning methodology for riparian land cover classification using aerial photographs and its application for flood modelling. *Journal of Hydroinformatics, 24*(1), 179–201. https://doi.org/10.2166/HYDRO.2022.134

Zhu, H., Yu, M., Zhu, J., Lu, H., & Cao, R. (2019). Simulation study on effect of permeable pavement on reducing flood risk of urban runoff. *International Journal of Transportation Science and Technology, 8*(4), 373–382. https://doi.org/10.1016/j.ijtst.2018.12.001