Geospatial Modelling of Watershed Peak Flood Discharge in Selangor, Malaysia

Ryan Cheah, Lawal Billa, Andy Chan, Fang Yenn Teo, Biswajeet Pradhan, and Abdullah M. Alamri

Abstract: Conservative peak flood discharge estimation methods such as the rational method do not take into account the soil infiltration of the precipitation, thus leading to inaccurate estimations of peak discharges during storm events. The accuracy of estimated peak flood discharge is crucial in designing a drainage system that has the capacity to channel runoffs during a storm event, especially cloudbursts and in the analysis of flood prevention and mitigation. The aim of this study was to model the peak flood discharges of each sub-watershed in Selangor using a geographic information system (GIS). The geospatial modelling integrated the watershed terrain model, the developed Soil Conservation Service Curve Cumber (SCS-CN) and precipitation to develop an equation for estimation of peak flood discharge. Hydrological Engineering Center-Hydrological Modeling System (HEC-HMS) was used again to simulate the rainfall-runoff based on the Clark-unit hydrograph to validate the modelled estimation of peak flood discharge. The estimated peak flood discharge showed a coefficient of determination, $r^2$ of 0.9445, when compared with the runoff simulation of the Clark-unit hydrograph. Both the results of the geospatial modelling and the developed equation suggest that the peak flood discharge of a sub-watershed during a storm event has a positive relationship with the watershed area, precipitation and Curve Number (CN), which takes into account the soil bulk density and land-use of the studied area, Selangor in Malaysia. The findings of the study present a comparable and holistic approach to the estimation of peak flood discharge in a watershed which can be in the absence of a hydrodynamic simulation model.

Keywords: GIS; watershed precipitation; curve number; peak flood discharge; Selangor; Malaysia

1. Introduction

Various methods, such as the widely used rational method published by Thomas Mulvaney in 1851, are used to estimate peak discharge. The rational method uses rainfall intensity, drainage area and a runoff coefficient to determine peak discharge [1–3]. This method is often conservative and very dependent on the runoff coefficient that is determined based on land use. Inaccuracies may arise in this
method if the wrong runoff coefficient is assigned or if the catchment area is incorrectly delineated [4,5]. Additionally, it does not take into account the infiltration rate of the precipitation as well as the soil type. The hydrological modelling methods are becoming the most useful and efficient techniques to address the human error element found in the rational method and they also provide the ability to include and assess the soil type when determining peak discharge. The primary attraction of the rational method has been its simplicity. However, with increasing capacity and capability of computer processing, computational designs and hydrological procedures, the use of appropriate hydrological software have become the preferred option for hydrological model generation [6–8]. Hydrological modelling is a powerful technique for hydrological system investigation, planning and the development of an integrated management approach for hydrological regimes and water resources [9,10]. Growing advances in computational power and the availability of increasingly fine resolution spatial data, has also made it possible to accurately describe watershed characteristics when determining runoff response to rainfall input [11].

Numerous hydrological models such as Simulator for Water Resources in Rural Basins (SWRRB), Environmental Policy Integrated Climate (EPIC), Groundwater Loading Effects of Agricultural Management Systems (GLEAMS) and Technical Release No.20 (TR20) have been developed, but a Geographic Information System (GIS) forms the baseline model for pre-processing and post-processing of data. A GIS as a computer-based tool that provides a suitable environment for efficient management of large and complex databases usually associated with hydrologic modelling and facilitates the processing, managing and interpretation of the hydrological data [12]. Applying geospatial modelling with the combination of digital elevation models (DEM) and the hydrological tools of ArcGIS, allows physical properties such as the catchment area, stream length and slopes of the study area to be accurately defined, and used in different hydrological models [13].

It was commented by [14] that, “With a larger number of parameters, there is a greater possibility that some parameters become to site-specific”. It was concluded by [2] that simple models involving fewer parameters forecasted discharge in the Nile River Basin better than models using more parameters and complex mathematical computations. Hence, the simple and straightforward Soil Conservation Service Curve Number (CN) method was adopted for their study. The soil conservation service Curve Number (SCS-CN) developed by the United States Department of Agriculture [15] is a measure of runoff potential that applies geospatial modelling without the use of hydrological models. The curve number is a dimensionless coefficient that ranges from 0 to 100, taking into account the soil and land-use types of the study area. While the curve number model can be used successfully as a calibration parameter in hydrological models, it was originally developed for use in determining streamflow for single storm events, and not for day-to-day analysis. Thus, it is a simple way to match observed rainfall data to a predicted stream flow. According to [9], the curve number was developed using data from a limited number of regions in the U.S. and may not be applicable for all regions, and this could be a limitation.

An infiltration approach should be used to determine the variation of runoff during a storm. The curve number runoff equation is not an infiltration equation. The runoff equation can, however, be used as a surrogate [15]. Data such as soil type and land use can be incorporated into the model hence further increasing the accuracy of the predicted peak discharge [16]. The importance of both soil type and land use in storm-runoff generation studies has also been proven in a research conducted by [17] while investigating the effects of climate change on the storm-runoff generation.

The main objectives of this study are to investigate the relationship between the peak flows of a watershed, watershed area, precipitation and land use by geospatial modelling and to quantify the aforementioned factors in a single empirical equation of peak flood discharge estimation.
2. Materials and Methods

2.1. Study Area

The state of Selangor in Malaysia has an area of 8200 km\(^2\) and is located between latitudes 2°35′–3°60′ N and longitudes 100°45′–102°00′ E (Figure 1). Selangor is affected by heavy and prolonged rainfall from June to September during the southwest monsoon and December to March during the northeast monsoon. In general, heavy localised rainfalls associated with thunderstorms of short durations occur throughout the year. The average annual rainfall ranges between 1900 mm to over 2600 mm.

The northern part of Selangor is predominantly underlain by igneous rock, mainly granite with some dacite, rhyolite, and micro granodiorite. The Northwest of Selangor is covered by granitic rocks, vein quartz of the mesozoic age and schist of Kajang formation. The central area of Selangor is underlain by limestone of the Kuala Lumpur formation, schist in the lands between Kelang and Gombak Rivers, while granite underlies most of the Ampang area [18]. Soils in Selangor vary from very thin sandy regolith covering quartzite ridges to the extremely deep weathering profile of gentle granite slopes. The soils are mostly derived from igneous, sedimentary and metamorphic rocks and also areas of marine and river alluvium [18].

Selangor consists of mainly lowland dipterocarp forests (LDF) which is the major forest type in Malaysia and are taller than any other broad-leaved tropical rainforest in the world. On the shorelines and riverbanks of Selangor, mangrove forests can be found. The state of Selangor has six main rivers of which the Selangor River is the largest, with a catchment area of about 1960 km\(^2\) covering about 25% of the state. The Selangor River runs from the upstream at Kuala Kubu Bharu in the east, near the Titiwangsa Mountains, to the Straits of Malacca at the downstream at Kuala Selangor in the west. The other major river is the Klang River that is 120 km long, with 11 major tributaries and a catchment area of 1288 km\(^2\). It flows through Kuala Lumpur and ends up in the Straits of Malacca. Thus Selangor comprises three major river basins, namely the Klang, Langat and, most importantly, the Selangor river basin.

Figure 1. Location of Selangor in peninsular Malaysia.
2.2. Data Collection and Geospatial Modelling

Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010)-type DEMs were used for the study [19]. The DEM with a spatial resolution of 75 m was processed in ArcGIS to determine and extract hydrological parameters such as slope, flow direction, flow accumulation, drainage features and to delineate the watersheds. Figure 2 shows the DEM and the catchments and drainage features of Selangor, respectively. Considering the size of Selangor, the DEM with a pixel size (75 x 75 m) was sufficient for the coverage of the multiple watersheds of the study area.

2.2.1. Hydrologic Soil Group Classification

Soils play an important role in determining CN, with different soil types having different infiltration rates which are a function of surface intake rates and subsurface permeability. In developing the CN, the United States Department of Agriculture (USDA) utilized soil type data to derive the infiltration rate and categorized the soil based on the infiltration rate. In this study, soil data for Selangor was downloaded from SoilGrids [20] and further processed for soil bulk density. Thus, the bulk density of the soil was used in place of the infiltration rate. Soil bulk density is an important factor that affects soil infiltration capacity. As soil bulk density increases, soil porosity decreases, resulting in a decrease in soil infiltration capacity [21]. By having the bulk density data, the soil type is no longer significant. Figure 3 shows the processed soil bulk density of Selangor at 1m depth classified into 4 hydrologic soil groups (A, B, C and D). The details of the hydrological soil group classification are presented in Table 1.

Table 1. Hydrologic soil group classification

<table>
<thead>
<tr>
<th>Soil Group</th>
<th>Infiltration Rate (mm/hr)</th>
<th>Bulk Density (kg/m³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>8–12</td>
<td>555–960</td>
</tr>
<tr>
<td>B</td>
<td>4–8</td>
<td>960–1141</td>
</tr>
<tr>
<td>C</td>
<td>1–4</td>
<td>1141–1264</td>
</tr>
<tr>
<td>D</td>
<td>0–1</td>
<td>1264–1488</td>
</tr>
</tbody>
</table>

Figure 2. (a) Digital elevation model (DEM); and (b) catchment and drainage features of Selangor.

Figure 3. Soil bulk density map.
Table 1. Hydrologic soil group classification.

<table>
<thead>
<tr>
<th>Soil Group</th>
<th>Infiltration Rate (mm/hr)</th>
<th>Bulk Density (kg/m³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>8–12</td>
<td>555–960</td>
</tr>
<tr>
<td>B</td>
<td>4–8</td>
<td>960–1141</td>
</tr>
<tr>
<td>C</td>
<td>1–4</td>
<td>1141–1264</td>
</tr>
<tr>
<td>D</td>
<td>0–1</td>
<td>1264–1488</td>
</tr>
</tbody>
</table>

2.2.2. Combination of Hydrologic Soil Group and Land-Use Classes

The latest land use for Selangor was developed from Landsat Operational Land Imager images of the year 2016 downloaded from the United State Geological Services (USGS) website. The Landsat images were mosaicked to improve coverage over the study area. A pan-sharpening operation was conducted to improve the spatial resolution of the multispectral visible bands to 15 m using the panchromatic band. The Selangor land use was classified into six different classes based on a supervised classification using the ArcGIS image analysis tool. The land-use classification (Figure 4a) includes water bodies, residential, industrial and commercial areas, plantations, green open spaces and forests. The land-use classification was then combined with the hydrologic soil group classification (Figure 3) to process and determine the CN using a relationship between hydrologic soil group and land cover as described in Table 2. Figure 4b shows the output map of the generated soil conservation service curve number.

Figure 3. Soil bulk density map.

Figure 4. (a) Land-use classification of Selangor; (b) generated curve number map.

Table 2. Runoff Curve Number.

<table>
<thead>
<tr>
<th>Description of Land Use</th>
<th>Hydrological Soil Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
</tr>
<tr>
<td>Cultivated land</td>
<td>72</td>
</tr>
<tr>
<td>Natural vegetation</td>
<td>45</td>
</tr>
<tr>
<td>Urban commercial</td>
<td>77</td>
</tr>
<tr>
<td>Urban housing</td>
<td>61</td>
</tr>
<tr>
<td>Bare land</td>
<td>49</td>
</tr>
<tr>
<td>Streets and roads</td>
<td>98</td>
</tr>
<tr>
<td>Open water</td>
<td>0</td>
</tr>
</tbody>
</table>

2.2.3. Spatial Rainfall Data Processing

The daily rainfall data can be obtained from rainfall gauge stations throughout Selangor. For the study, rainfall data of 50 years from 1965 to 2015 were obtained from the Department of Irrigation and Drainage Malaysia (DID). Data from each rainfall station were checked and filtered for consistency. Stations with missing data were not taken into consideration. The mean maximum daily rainfall of
each station was then spatially interpolated based on the inverse distance weighting (IDW) method using ArcGIS. Figure 5 shows the results of the maximum daily precipitation of Selangor.

![Figure 5. Maximum daily precipitation of Selangor.](image)

### 2.3. Hydrological Modelling

#### 2.3.1. Calculation of Curve Number Peak Discharge

The storm has to overcome a threshold determined by depression storage, interception as well as infiltration volume before runoff can occur. Initial abstraction is the amount of rainfall required to satisfy the aforementioned volumes. After runoff begins, additional losses due to infiltration will still occur. In general, accumulated infiltration increases with increasing rainfall up to a certain maximum retention, and runoff also increases as rainfall increases. The ratio of actual retention to maximum retention is assumed to be equal to the ratio of direct runoff to rainfall minus initial abstraction. This can be expressed mathematically as [15].

\[
\frac{F}{S} = \frac{Q}{P - I}
\]  

(1)

where,

- \(F\) = Actual retention after runoff begins, mm
- \(S\) = Watershed storage, mm
- \(Q\) = Actual direct runoff, mm
- \(P\) = Total rainfall, mm
- \(I\) = Initial abstraction, mm

The initial abstraction should take into consideration of depression storage, interception and infiltration occurring prior to runoff. The relationship between \(I\) and \(S\) was estimated by analysing rainfall-runoff data for many small watersheds to eliminate the necessity of estimating both parameters \(I\) and \(S\) in the above equation. The empirical relationship is given by [22].

\[
I = 0.2S
\]  

(2)

Substituting Equation (3) into Equation (2) gives:

\[
Q = \frac{(P - 0.2S)^2}{P + 0.8S}
\]  

(3)
The equation now has only two unknowns, one variable $P$ and one parameter, $S$. $S$ is related to curve number (CN) by:

$$S = \frac{25400}{\text{CN}} - 254$$

(4)

where,

$\text{CN} = $ Curve number (From Table 2)

Finally, Equation (5) was combined with the spatially distributed parameters of rainfall, soil and land–use to simulate the runoff in the watersheds. The total daily discharge was computed by the following equation:

$$Q_{CN} = \sum_{i=1}^{n} Q_i \times A_i$$

(5)

where,

$Q_{CN} =$ Total daily runoff discharge of the watershed, m$^3$

$Q_i =$ The depth of daily direct runoff of cell No. $i$, m

$A_i =$ The area of corresponding cell No. $i$, m$^2$

$n =$ Total cell numbers of the watershed

2.3.2. Simulation of Unit Hydrograph Peak Discharge

The Clark-unit hydrograph model in Hydrological Engineering Center-Hydrological Modeling System (HEC-HMS) was specifically chosen for peak flow based on the Malaysian Hydrological Procedure (HP) No. 27, published by the Department of Irrigation and Drainage [23]. The HP No. 27 proposes empirical equations for both the storage coefficient $R$ and the time of concentration $T_c$.

For simplicity and consistency, the catchment area, stream slope, and main stream length are used to estimate $T_c$ and $R$ for this procedure.

$$T_c = 2.32A^{-0.1188}L^{0.9573}S^{-0.5074}$$

(6)

$$R = 2.976A^{-0.1943}L^{0.9995}S^{-0.4588}$$

(7)

where,

$A =$ Catchment area [km$^2$]

$L =$ Main stream length [km]

$S =$ Weighted slope of main stream [m/km]

The weighted slope of main stream, $S$ is given by the equation:

$$S = \left( \frac{\sum l_i \sqrt{S_i}}{\sum l_i} \right)^2$$

(8)

where,

$l_i =$ Incremental stream length

$S_i =$ Incremental slope

The catchment area, main stream length and weighted slope of main stream is modelled in ArcGIS using the above stated methods. The design baseflow, $Q_B$ of each stream is then calculated with the equation from HP No. 27:

$$Q_B = 0.11A^{0.85889}$$

(9)

where,
A = Catchment area [km²]

The catchment area, A; precipitation, P; storage coefficient, R; time of concentration, \( T_c \); design baseflow, \( Q_b \) were then inputted into HEC-HMS to model the peak discharge, \( Q_{Clark} \) that will serve as a comparison against the ArcGIS-modelled peak discharge.

The Clark-unit hydrograph method was used as the control variable of the study. The curve number derived peak discharge was compared against the Clark-unit hydrograph peak discharge. The Clark-unit hydrograph rainfall-runoff relationships are derived based on average conditions and should be only used for design purposes. The same applies to the \( T_c \) and \( R \) values derived. The areal variability of catchment rainfall during a storm causes the time of concentration of a catchment to vary from storm to storm. This makes the assumption of uniform areal distribution of design storm invalid. Some unaccounted for storage depression (e.g., wetland, extremely flat catchment slopes) could lead to the overestimation of the peak discharge and the underestimation of the time to peak when using the equations [23].

3. Results

3.1. Peak Flood Discharge Estimation

It was observed in Figure 6a that the two peak discharges agree with each other fairly well, showing a linear regression of 0.76. The peak discharge derived from the CN method was an underestimation when compared to the Clark-Unit hydrograph peak discharge. The Clark-unit hydrograph is a lumped model that includes a lot of assumptions about hydrological processes. Because of these assumptions, lumped models have a tendency to over-or underestimate runoff values [24]. Hence, using the equation of the best fit line that intercepts the origin, a correction factor was applied to the peak discharge modelled after the CN method to calibrate it.

\[
Q_{peak} = 1.8118Q_{CN} \tag{10}
\]

Figure 6. (a) Graph of urbanized area per watershed against total curve number (CN); (b) graph of urbanized area per watershed against total CN.
The graph of urbanized area of a watershed was plotted against the total curve number of a watershed. This was done to establish a positive relationship between urban area per watershed, \( A_u (\text{km}^2) \) and the sum of the curve number per watershed, \( \text{CN} \) from Figure 6b, it was observed that the total curve number of the watershed increases as the urbanized area of a watershed increases. This strong correlation was further proven with the high linear regression of 0.7107. Hence, it can be concluded that an urbanized area increases the runoff potential of a watershed.

Having proven the hypothesis of a positive correlation between urbanized area and curve number, the relationship between the calibrated peak discharge \( (Q_{\text{peak}}) \) and Curve Number (CN), precipitation \( (P) \) and watershed area \( (A_w) \) was investigated to formulate Equation (12) based on the three factors. This was successfully done by obtaining the regression coefficients between peak discharge and CN, P, Aw.

By plotting the corrected peak discharge against the curve number, watershed area and precipitation, it can be seen that the curve number and watershed area are relatively more correlated to the peak discharge than the precipitation. This can be explained as the peak discharge being directly related to the surface runoff of the watershed. Thus, the larger the watershed, the more surface area that is able to intercept rainfall and contribute to the peak discharge. In the case of CN number, it was concluded that the urban area increases the runoff potential of the watershed. Precipitation, however, showed a weaker correlation because the individual watersheds have different characteristics, stream lengths, and base flows. Base flow forms the initial input baseline parameter for runoff simulations, as there will usually be stream discharge regardless of the existence of rainfall. Runoff occurs when there is rainfall, thus, precipitation was considered in the development of the equation.

By taking the linear regression coefficient as the exponent of the respective factors, the multiplication product of the three powered factors was plotted against the calibrated CN modelled peak discharge. From Figure 7 below, it can be observed that the plotted best-fit power line has a strong correlation, with a regression coefficient of 0.9445. Hence, this power best-fit line is taken as the product equation.

\[
Q_{\text{peak}} = 0.3257 (A_w^{0.8019} \cdot P^{0.2871} \cdot \text{CN}^{0.6486})^{0.7083}
\]  

(11)

Figure 7. Graph of calibrated CN peak discharge against the multiplication product of adjusted watershed area, precipitation and curve number.

3.2. Classification of Developed Model

A runoff model is defined by [25] as a set of equations that aid in the estimation of the amount of rainfall that turns into a runoff as a function of various parameters used to describe the watershed. According to [26], the runoff model, Equation (11), would be classified as an empirical model or more specifically, a regression equation. Regression equations find the functional relationship between inputs and outputs [25]. It is suggested by both [2] and [27] that empirical models are black-box
models, meaning very little is known about the internal processes that control how runoff results are determined.

In terms of spatial structuring, according to [26] this developed model would be classified as a lumped model, as it treats the catchment area as a single homogeneous unit. In lumped models, the spatial variability of the catchment is disregarded [28,29]. Both the Clark-unit hydrograph and SNS-CN methods exhibit this property too. By assuming homogeneity over the catchment, the model loses the spectral resolution of the data. For example, rainfall and runoff patterns vary over space and time, but in lumped models, they are considered stationary [26].

3.3. Model Validation

Equation (12) was verified by plotting the estimated peak discharge against the peak discharge modelled with the curve number method. From Figure 8, it can be seen that the estimated peak discharge agrees fairly well with the calibrated CN modelled peak discharge, with a strong linear regression coefficient of 0.7854. Hence, the equation developed to estimate the peak discharge during a storm event that relates the watershed area, precipitation, and curve number can be considered as a valid equation.

![Figure 8. Graph of CN modelled peak discharge against estimated peak discharge.](image)

The equation to estimate flood peak discharge taking into account the watershed area, precipitation, soil bulk density, and land use was successfully implemented. A higher resolution DEM and soil map should further improve the reliability of the curve number grid. The use of rainfall-runoff models was questioned by [30] as a better option for surface runoff modelling. Similar studies by [31] and [32] suggest that physically based models like MIKE-SHE (Système Hydrologique Européen) and TOPMODEL (TOPography based hydrological MODEL) have certain disadvantages as they require a large number of input parameters for calibration making their use highly data-dependent, time consuming and expensive. The approach presented in this study provides a straightforward method with limited parameter input to estimate peak discharges based on a GIS.

4. Conclusions

The hypothesis suggesting urbanization leads to increased runoff and short lag time of peak discharges holds true. In essence, the modelled peak discharge based on curve number was generally in agreement with the modelled peak discharge based on the hydrologic procedure No. 27. The curve number method, however, underestimated the peak discharge when compared to the Clark-unit hydrograph. Hence, a correction factor was introduced to calibrate the curve number modelled peak discharges. All three factors (curve number, precipitation and watershed area) were found to be correlated to peak discharge. This result demonstrated that the geospatial modelled peak flood discharge was reliable and provided an alternative method of monitoring potential flood in multiple
watersheds. Ideally speaking, modeling of spatial variation of runoff generation by the infiltration excess mechanism and the downslope infiltration of runoff onto unsaturated areas require high spatial data input and or field sampling. For data-limited and inaccessible regions of Malaysia and most parts of the world, this has been a limitation in estimating the runoff in various watersheds. This proposed method applied geospatial techniques in the modeling of watershed area, land use, soil bulk density, spatial interpolated rainfall and runoff curve number and the derivation of the equation for estimation of peak discharge. The model serves as an alternative for estimation of peak discharge in watersheds and areas with poor data availability.

Author Contributions: R.C. and L.B. conceived, designed and performed the experiments; they analyzed the data and wrote the paper with contribution from A.C. and F.Y.T., and B.P. and A.M.A. professionally revised and edited the manuscript. The manuscript was discussed and reviewed by all of authors as they enhanced the context with sufficient references.

Funding: The research is supported by the Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS), University of Technology Sydney under grant numbers: 323930, 321740.2232335, 321740.2232424 and 321740.2232357. This research was also supported by Researchers Supporting Project number RSP-2019/14, King Saud University, Riyadh, Saudi Arabia.

Acknowledgments: This research was supported by the University of Nottingham Malaysia Campus common project funds and learning resources. The metrological data were made available by the Department of Irrigation and Drainage (DID) Malaysia.

Conflicts of Interest: The authors declare no conflict of interest.

References

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).