

Article

Formulation of Parsimonious Urban Flash Flood Predictive Model with Inferential Statistics

Lloyd Ling ^{1,*} , Sai Hin Lai ^{2,*} , Zulkifli Yusop ³, Ren Jie Chin ^{1,*} and Joan Lucille Ling ⁴

¹ Centre of Disaster Risk Reduction (CDRR), Civil Engineering Department, Lee Kong Chian Faculty of Engineering & Science, Universiti Tunku Abdul Rahman, Jalan Sungai Long, Kajang 43000, Malaysia

² Department of Civil Engineering, Faculty of Engineering, University of Malaya, Kuala Lumpur 50603, Malaysia

³ Centre for Environmental Sustainability and Water Security, Universiti Teknologi Malaysia, Skudai 81310, Malaysia; zulyusop@utm.my

⁴ American Degree Programme, Department of Liberal Arts and Sciences, Taylor's University, No. 1, Jalan Taylors, Subang Jaya 47500, Malaysia; linglucille@gmail.com

* Correspondence: linglloyd@utar.edu.my (L.L.); laish@um.edu.my (S.H.L.); chinrj@utar.edu.my (R.J.C.)

Abstract: The curve number (CN) rainfall–runoff model is widely adopted. However, it had been reported to repeatedly fail in consistently predicting runoff results worldwide. Unlike the existing antecedent moisture condition concept, this study preserved its parsimonious model structure for calibration according to different ground saturation conditions under guidance from inferential statistics. The existing CN model was not statistically significant without calibration. The calibrated model did not rely on the return period data and included rainfall depths less than 25.4 mm to formulate statistically significant urban runoff predictive models, and it derived CN directly. Contrarily, the linear regression runoff model and the asymptotic fitting method failed to model hydrological conditions when runoff coefficient was greater than 50%. Although the land-use and land cover remained the same throughout this study, the calculated CN value of this urban watershed increased from 93.35 to 96.50 as the watershed became more saturated. On average, a 3.4% increase in CN value would affect runoff by 44% (178,000 m³). This proves that the CN value cannot be selected according to the land-use and land cover of the watershed only. Urban flash flood modelling should be formulated with rainfall–runoff data pairs with a runoff coefficient > 50%.

Keywords: curve number; flash flood model; inferential statistics



Citation: Ling, L.; Lai, S.H.;

Yusop, Z.; Chin, R.J.; Ling, J.L.

Formulation of Parsimonious Urban Flash Flood Predictive Model with Inferential Statistics. *Mathematics* **2022**, *10*, 175. <https://doi.org/10.3390/math10020175>

Academic Editors:

Araceli Queiruga-Dios, Maria

Jesus Santos, Fatih Yilmaz, Deolinda

M. L. Dias Rasteiro, Jesús

Martín Vaquero and Víctor

Gayoso Martínez

Received: 31 October 2021

Accepted: 30 December 2021

Published: 6 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Flood and its related disasters are caused by excessive volumes of water (runoff) which are not absorbed by the ground. Residents at low-elevated regions are often at risk of inundation, financial loss, and even the loss of lives. As the pace of urbanisation accelerates around the world, flash flood damage takes place more frequently. Between 1961 and 2020, nearly 10,000 cases were reported with 1.3 million deaths and a minimum of USD 3.3 trillion of financial losses at an equivalent loss rate of almost USD 1800 per second [1]. On average, the total reported deaths worldwide were 23,000/year for the past 6 decades at an equivalent rate of one death every 24 min [1]. Thus, it is important to study the relationship between rainfall and runoff in order to quantify the runoff amount from rainfall with equations or predictive models for water resources management, flood prediction, and risk mitigation to benefit mankind. Although there are many rainfall–runoff models for runoff prediction, this study assessed a popular rainfall–runoff model from the United States Department of Agriculture (USDA), Soil Conservation Services (SCS) for flash flood prediction and benchmarked its runoff prediction accuracy against two parsimonious models which also used two modelling parameters. The main objective is to formulate and identify a parsimonious runoff predictive model which requires the least modelling parameters for urban flash flood prediction.

Near the end of the 1930s and beginning of the 1940s, infiltrometer tests were carried out by the Natural Resources Conservation Service (NRCS) agency, also known as USDA, SCS in order to assess the impact of watershed treatments along with soil conservation measures on the rainfall–runoff process. The US Congress administered a Watershed Protection and Flood Prevention Act (Public Law PL-566) years later in August 1954. Therefore, the SCS had to establish a procedure for national implementation and thus, hydrologic methods that were once an agency procedure to tackle particular scenarios were taken and applied with immediate effect. Having said that, the method did not emerge in the archived literature, nor underwent professional review and critical procedures for a decade and a half [2,3]. The procedure established by the SCS was developed according to available data primarily from watersheds overseen with rain and streamflow gages in the USA. However, the SCS had to overcome most issues in ungaged watersheds. Hydrological models and procedures constructed by early pioneers were therefore adopted and grew into a runoff equation for the SCS curve number (CN) as:

$$Q = \frac{(P - I_a)^2}{P - I_a + S} \quad (1)$$

where

Q = Amount of runoff depth (mm).

P = Depth of rainfall (mm).

I_a = The initial abstraction amount (mm).

S = Maximum potential water retention of a watershed (mm).

Furthermore, the SCS developed a hypothesis whereby $I_a = \lambda S$, λ being the initial abstraction ratio coefficient. The equation was vaguely supported by daily rainfall and runoff data and the sole source of accredited evidence was the NRCS's National Engineering Handbook, Section 4 (NEH-4). In 1954, the SCS also created the CN methodology, although certain preliminary field data and core assumptions emerged undocumented and untraced [2–4].

In addition to that, the SCS proposed that $I_a = \lambda S = 0.20S$ where the initial abstraction ratio coefficient (λ) was 0.20, in which the criterion correlates I_a and S . Field data acquired from different regions in the USA also contributed to the proposed relationship of I_a to S . They determined that $\lambda = 0.20$ due to the simple correlation between I_a and S data points used by the SCS. Notwithstanding the sizeable scatter in the data, $\lambda = 0.20$ was adopted by the SCS with the conclusion that half of the data points fell in range of $0.095 < \lambda < 0.38$ [2,4]. The initial abstraction (I_a) is otherwise the depth of the rainfall event necessary for runoff commencement. Replacing $I_a = 0.20S$ simplifies Equation (1) to become the conventional SCS runoff forecast model that is commonly used in textbooks, certified hydrological design manuals, and is widely used in design software as well as programs following its establishment in 1954 [2–4]. The conventional (simplified) SCS runoff prediction model is:

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

Equation (2) holds a restriction where $P > 0.20S$, or else Q would be equal to zero. Howbeit, proliferating studies lean against the accuracy of Equation (2)'s predictions as well as the hypothesis where $I_a = 0.20S$. Literature reviews exhibit inconsistency in using Equation (2) to predict runoff results and a number of researchers encouraged the calibration of regional hydrological conditions to be carried out instead of simply following that of the SCS [5,6].

In the past six decades, the CN technique gained wide acceptance and appliance to hydrological problems that were not originally intended to be solved by the SCS. The technique became the most popular method to predict runoff and was widely taught in colleges and universities worldwide. The model was also integrated into many software and USDA SCS systems like the Chemicals, Runoff and Erosion System simulation from the Agricultural Management Systems (CREAMS) model. It is also used in other models, such as the Agricultural

Nonpoint Source model (AGNPS), Hydrologic Engineering Centre–Hydrologic Modelling System (HEC-HMS), USDA Technical Release 20 (TR-20), and USDA Technical Release 55 (TR-55). In fact, it presented a runoff component for successive water quality and erosion models comprising the Areal Nonpoint Source Watershed Environment Response System (ANSWERS), the simulator for Water Resources in Rural Basins (WRRB), the Erosion Productivity Impact Calculator (EPIC), the Pesticide Root Zone model (PRZM), the Water Erosion Prediction Project (WEPP), and the distributed Soil Water Assessment Tool (SWAT) [2–5]. Although not mentioned here, any software and technical handbooks which incorporated Equation (2) are very likely to produce inconsistent runoff prediction results.

Recently, researchers proposed a global gridded CN concept for runoff modelling [7,8]. However, some reported that the usage of the CN in representing a watershed is arbitrary, vague, and often contradictory in describing related areas of land cover [2,9]. Other researchers emphasised the importance to apply a multi-modelling approach and statistical methods with land-use and cover variations in order to achieve better flood modelling results [10,11], while [12] even concluded that even very-fine-resolution topography and high-resolution land cover data may not be able to produce reliable urban flood modelling results using the HEC-RAS software. Some researchers also reported difficulty to determine an optimum λ value [13] in the SCS CN model. Therefore, it is crucial to ameliorate the modelling approach in order to achieve better model applications to manage water resources and river basins.

As Equation (2) was rooted into many fields, the model re-assessment prevents SCS practitioners from committing type II errors (a statistical term used within the context of hypothesis testing that describes the error that occurs when one accepts a null hypothesis that is actually false). The calibration methodology will derive a statistically significant rainfall–runoff model with better runoff prediction accuracy and produce a watershed-specific CN system for an area of interest. Instead of referring to the conventional CN table compiled by SCS which originated from USA, researchers will be able to derive watershed-specific CN anywhere with the presented methodology in this study. The CN derivation methodology will also replace the common unscientific CN adjustments or tweaking practices in order to achieve better runoff prediction results. The approach entails performing “trial and error” CN refining with the observed data to better the results of runoff prediction and assumes that $\lambda = 0.20$ for any watershed. However, such practice lacks statistical justification and often leads to inconsistent runoff prediction results. Study [14] cautioned that a $\pm 10\%$ CN variation could lead to $\pm 50\%$ change in runoff. It would create CN value(s) for a watershed; however, the “calibrated” CN value(s) may not even be able to predict runoff conditions of other watersheds with similar land-use and land cover conditions again. CN values are a better match for traditional agricultural watersheds but less accurate in the estimation of semiarid rangelands and are the worst for forested watersheds [2]. Therefore, CN values should be derived from the local P-Q dataset to reflect realistic situations [6,13,15].

It is not a common practice for practitioners to assess the statistical significance of a predictive model with their dataset. Many engineering students were not taught about the importance of such validation procedures prior to the use of any formulas and equations. The main aim of this study is to emphasise that any formula or predictive model should not be blindly adopted, in order to avoid committing type II error.

2. Methodology and Study Site

When the SCS runoff model was adopted to model rainfall and runoff conditions of a watershed, the assumption of the initial abstraction ratio coefficient ($\lambda = 0.20$) was also accepted. SCS practitioners do not question the validity of the runoff model and therefore, there is a potential of perpetrating a type II error. A hypothesis was used to assess the validity of the existing SCS model according to the rainfall–runoff dataset of this study. The model will be calibrated only when the hypothesis is rejected. In the event that the hypothesis cannot be rejected, the existing SCS model will be adopted for modelling in this study.

Null Hypothesis (H_0): $I_a = \lambda S$ where $\lambda = 0.20$ is valid and applicable for this study. The null hypothesis assesses the validity of Equation (2) as pertaining to the collected dataset. H_0 must be statistically significant at least at alpha (α) significance level of = 0.05 for it to be adopted to model the rainfall–runoff condition at the watershed of study. Otherwise, the model should be calibrated. According to the SCS, λ is a constant value of 0.20 in $I_a = \lambda S$.

In the event of H_0 rejection, a statistically significant initial abstraction ratio coefficient (λ) and total abstraction value (s) will be derived at alpha = 0.01 which will lead to the calculation of the best collective representative CN for the watershed as proposed by the authors' previous study that based model calibration on the non-parametric inferential statistics, the bootstrapping bias corrected, and an accelerated (BCa) procedure [6]. Unlike the existing antecedent moisture condition (AMC) concept which correlates CN I (CN in dry conditions), CN II (CN in normal conditions), and CN III (CN in wet conditions) to determine the runoff condition of a saturated soil, this study extends the model calibration method proposed previously [6] to model the runoff condition of an urban watershed under different ground saturation conditions without using return period data. Other than that, this study added a new control factor to calibrate the model by limiting the overall model prediction bias near to zero to prevent the newly calibrated model's bias toward a specific dataset under this study. As such, the supervised numerical optimisation algorithm will identify the optimum λ and S value under the model bias control within the bootstrap BCa confidence intervals while keeping the overall runoff predictive model's error and bias near to zero. The BCa technique was chosen as it is the only inferential statistic with a bias correction ability while able to provide a confidence interval (CI) at specific alpha levels for statistical assessments. It is also a data-distribution-free method which is compatible to the nature of any rainfall data distribution and available in the IBM statistical software SPSS used by this study.

The SCS rainfall–runoff model was calibrated previously by US researchers under the recommendation to only consider rainfall depths > 25.4 mm (1 inch) in order to avoid modelling bias towards small rainfall events resulting in higher CNs in rural watersheds [2]. However, the recommended guideline often excludes low-rainfall-depth field data, leading to an insufficient sample size to produce statistically significant results. Some researchers even lowered the limit to 15 mm (arbitrary) with a minimum of 10 events in their study plots [16]. Therefore, this study reviewed the recommended guidelines and demonstrated the possibility to extend the authors' previously proposed calibration methodology and to propose another new model calibration method based on watershed saturation conditions.

A 22.33 km² urban watershed (Sungai Kayu Ara) in the capital city of Malaysia, Kuala Lumpur, which consists of a large portion of impervious area, was chosen for this study. Over the years, this urban watershed is plagued by flash flood damages while the return period-based concept to assess the drainage capacity efficiency has neither rectified nor improved the condition. Sungai Kayu Ara watershed covers an area as shown in Figure 1.

The river originates from a forest reserve in the northern upstream and flows toward the relatively flattened developed suburban areas. The outlet of this watershed is marked by a dash circle in Figure 1 under the monitor of a water level station. Ninety-two storm events (1.4 to 90 mm) were collected in this study to produce the rainfall–runoff dataset through the separation of base flow from the hydrograph. A total of 61% (56 rainfall events) of this dataset has a rainfall depth of <25.4 mm (1 inch), but with a measurable runoff amount. Runoff coefficient (Q/P) of this dataset spans from 6% to 97% which is ideal to model runoff change according to the watershed saturation conditions, out of which, thirty-seven (37) rainfall–runoff data pairs have Q/P greater than 50% and twenty-four data pairs are greater than 60%.

This watershed was chosen to show that it is possible to include rainfall depth < 25.4 mm (1 inch) and incorporate watershed saturation conditions into the SCS runoff model calibration for urban runoff prediction. The dataset also contains multiple Q/P (%) data batches in order to reflect the runoff trend and characteristic change due to increasing watershed saturation conditions.

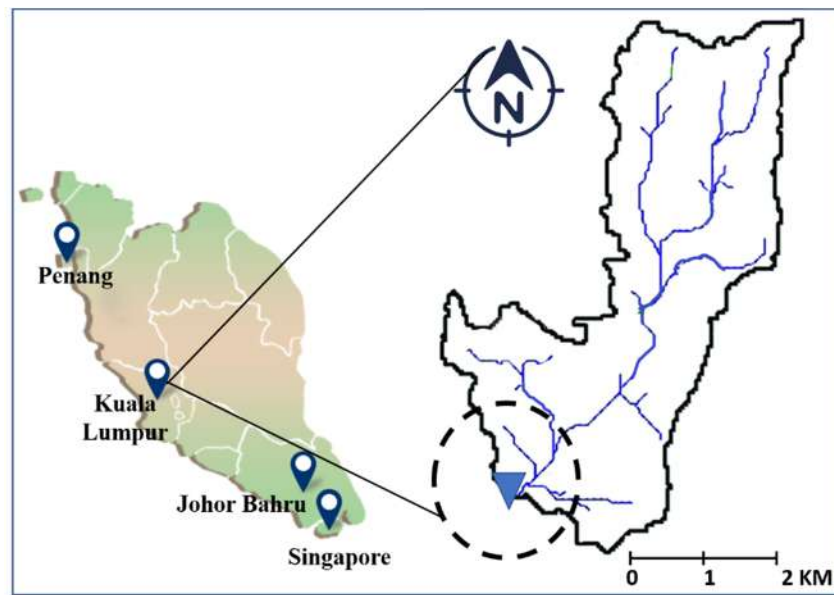


Figure 1. Location of Sungai Kayu Ara Watershed in Kuala Lumpur, Malaysia.

This study highlights the modelling effect on runoff predictions with different saturation-based datasets and advises practitioners against the blind adoption of the authors’ previously proposed model calibration methodology [6] for urban flash flood modelling with any rainfall–runoff dataset.

Unlike the Hortonian runoff model, hydrologists have recognised that runoff was the result of rainfall from very wet parts of a watershed since 1960 [17,18]. When the effective soil water-storage capacity is exceeded, runoff will be produced. This runoff generation process is referred to as the saturation-excess runoff. Q/P (%) of a watershed will become larger when effective soil water-storage is reducing [19–22]. Even a medium-sized storm event over a saturated watershed with a high runoff coefficient may cause severe flooding; therefore, it is critical to assess the performance of different runoff models under different runoff coefficient conditions in order to select the best performing parsimonious urban rainfall–runoff model for urban runoff estimation.

This study formulated a feasible runoff predictive model for urban flash flood prediction without using the return period data. Inspired by the saturation-excess runoff concept used by past studies [23–25], the SCS rainfall–runoff model’s calibration methodology presented previously [6] was extended to model urban runoff under different ground saturation conditions according to the runoff coefficient which was calculated by dividing runoff amount (Q) to the corresponding rainfall depths (P). The P - Q dataset was then sorted in descending order of the calculated runoff coefficient of each P - Q data pair. A minimum of twenty P - Q events were used to ensure a sufficient sample size for analyses [2,26,27]. The significance of this extended application paved the way for the calibrated SCS runoff model to model urban runoff under different watershed ground saturation conditions (i.e., $Q/P > 50\%$, 60% , and 70% etc.) for urban flash flooding forecasting, the design of hydraulic infrastructures, and drainage capacity assessment.

This study also presented the possibility of developing an urban rainfall–runoff model with a P - Q dataset which includes any rainfall depth that is less than 25.4 mm but with a significant measurable runoff amount. Equation (1) is calibrated with urban hydrological constraints according to the rainfall–runoff dataset to illustrate that the calibrated SCS runoff model could be utilised for modelling urban runoff. In the authors’ previous works [6,27], Equation (1) was rearranged into an S general formula, whereby the S and λ values can be derived according to the corresponding P - Q data pairs, the formula is:

$$S_{\lambda} = \frac{\left[P - \frac{(\lambda-1)Q}{2\lambda} \right] - \sqrt{PQ - P^2 + \left[P - \frac{(\lambda-1)Q}{2\lambda} \right]^2}}{\lambda}. \text{ The corresponding } S \text{ values will be denoted by } S_{\lambda}$$

to differentiate from the conventional S value where λ value = 0.20. S_λ must be correlated to the $S_{0.2}$ values prior to the calculation of the CN value [2,6,27]. The non-parametric bootstrapping (BCa) technique was set in SPSS for 2000 random samplings with replacements from the 92 data pairs and produced 99% confidence interval (CI) for parameter optimisation. In another words, the parameter of interest underwent 2000 random validations against its dataset. By utilizing Equation (1), the λ and S optimisation was achieved using the approach of numerical analyses. Model calibration of this study was conducted by removing data pairs of Q/P > 60%, between 40 and 60%, and <40% for model calibration, respectively, with validation against the remaining dataset. Through several iterations of the minimisation of the overall predictive model fitting bias and residual sum of square error (RSS) between the predicted Q against its observed values, the final model was formulated.

Runoff models are compared for their model predictive accuracy in this study. The residual spread of a model is indicated by residual sum of squares (RSS) in which a lower RSS is equivalent to an improved conjecturing model. On the other hand, the model efficiency index (E also known as the Nash–Sutcliffe) lies on a spectrum of minus infinity to 1.0 whereby index value = 1.0 shows an ideal conjectured model. In the instance where $E < 0$, it is inferior to utilizing an average to predict the dataset. The average residual of a predictive model (BIAS) indicates an altogether model prediction error quantified by the average of its residual to stipulate a pattern in the overall model prediction. A zero value means that it is an ideal error-free overall model prediction, and a negative value demonstrates the under-predictive overall model tendency, and conversely. In addition, a better predictive model will also have smaller residual range and interval, smaller standard deviation in residual, and mean residual near to zero.

2.1. Linear Regression Model

Equation (1) will be calibrated (only if H_0 was rejected) and benchmarked against the linear regression model for runoff prediction accuracy in this study with models of parsimonious interest. For urban rainfall–runoff, the one-dimensional linear regression model was proposed by early pioneers and often utilised, whereby the slope of the regression equation represents a hydrological reduction variable [28] or a proportion in relation to a whole of an impermeable area of a watershed [29]. The intercept on the x-axis estimates a watershed’s local depression amount or depression loss [30]. The linear regression model’s base form is:

$$Q = mP + c \quad (3)$$

where

Q = Amount of runoff (mm).

P = Depth of rainfall (mm).

m = Gradient (slope).

c = Constant (intercept on the x-axis).

2.2. Asymptotic Curve Number Modelling

US researcher [31] first pointed out that the most common use of Equation (2) was to calculate runoff from the rainfall depth with the same return period and proposed to perform “frequency matching” by pairing P-Q data pairs with same return periods, while another US researcher [32] reported that CN values decreased against increasing rainfall amounts, and there was a notable pattern in that CN would eventually approach a constant value in most cases. He also proposed a two-parameter asymptotic fitting method (AFM) to sort the rainfall and runoff dataset separately in descending order and pair them up again as “ordered data”. The method accepted the SCS proposal where $\lambda = 0.20$. Three different patterns were observed and classified as standard behaviour, complacent behaviour, and violent behaviour.

AFM was proposed to determine the representative CN for the watershed of interest through its P-Q dataset (λ value remains as 0.20 under this method) under the theoretical

projection when P approaches infinity amount. CN cannot be determined for the complacent behaviour watershed, but standard behaviour watershed follows the following formula for CN determination:

$$CN(P) = CN_{\infty} + (100 - CN_{\infty})e^{(-\frac{P}{k})} \tag{4}$$

where

CN(P) = Fitted CN value of a specific rainfall depth.]

CN_∞ = CN of a watershed of interest.

P = Rainfall depth (mm).

K = Fitting parameter.

Violent behaviour watershed follows the following formula for CN determination:

$$CN(P) = CN_{\infty} [1 - e^{-k(P-P_{th})}] \tag{5}$$

where

P_{th} = Threshold Rainfall depth (mm).

AFM was chosen to benchmark against the linear regression model and the proposed calibrated model in this study because it is the only CN model that relies on two fitting parameters and has been widely adopted by SCS practitioners.

3. Results

3.1. Linear Regression Model

Using the IBM SPSS, the best fitted model of linear regression for the P-Q dataset of this study based on the basic linear intercept Equation (3) was identified as below while the descriptive and inferential statistics are tabulated in Table 1. Equation (6) has the highest R²_{adj} of 0.831 and the lowest fitting model standard error (SE) of 4.648.

$$Q = 0.62P - 3.182 \tag{6}$$

Table 1. Descriptive and Inferential Statistics of Equation (6) at α = 0.05.

Model	Coeff.	p Value	CI		p Value	BCa CI	
			Lower	Upper		Lower	Upper
Constant	−3.182	0.001	−4.956	−1.408	0.004	−5.484	−0.882
Gradient	0.62	0.000	0.561	0.678	0.000	0.531	0.703

From Table 1, both fitting parameters of the linear regression model are statistically significant (*p* < 0.01) and the BCa results reaffirmed the *p* value significances. Previous researchers [29] used the fitting gradient to estimate the percentage of impervious area while the constant value implies the local depression amount or the watershed’s depression loss. However, at alpha = 0.05, the confidence interval (CI) spans 252% and 83% in the respective constant and gradient fitting while the BCa CI results show 622% and 76% variation in those categories. At alpha = 0.05, stringent BCa CI results imply that the local depression, or the depression-loss, of the Sungai Kayu Ara watershed can be any value between 0.89 mm and 5.5 mm (BCa lower-to-upper confidence interval of the fitting constant parameter) while the interpretation of impervious area estimates within the watershed can be any value between 53.1% and 70.3% (BCa lower-to-upper confidence interval of the fitting gradient parameter). Hydrological implications from those two parameters are open to a wide range of interpretations at this point.

3.2. The Existing (Simplified) SCS Runoff Model

Using the entire dataset, the results of BCa established a CI span for derived λ values of the P-Q dataset at the Sungai Kayu Ara watershed (Table 2) can also be used to assess H_0 . To implement Equation (2), H_0 must be accepted. The λ confidence interval span and the standard deviation are utilised to assess H_0 .

Table 2. BCa results ($\alpha = 0.01$) of derived λ values at Sungai Kayu Ara watershed.

	λ	BCa 99% CI				
		Lower	Upper	S	Lower	Upper
Mean	0.021	0.007	0.051	36.3	28.2	46.7
Median	0.004	0.004	0.006	30.0	22.7	33.7
Skewness	9.020			3.2		
Kurtosis	84.332			16.8		
Std. Deviation	0.097			35.8		

The mean and median CI span (Table 2) did not comprise of a $\lambda = 0.20$. Furthermore, the derived λ dataset’s standard deviation is not zero (to indicate that it is a constant), hence H_0 can be rejected at $\alpha = 0.01$, ruling out the validity of Equation (2) to model runoff conditions at this watershed. Therefore, the SCS model must undergo model calibration to avoid the risk of type II errors.

3.3. Calibrated SCS Runoff Model for the Sungai Kayu Ara Watershed

Tabulated in Table 2 are the descriptive statistics regarding the nature of the data distribution of the values of λ and S. The supervised least-square-fitting algorithm has been configured to find the optimum value of λ and S in the confine of the median confidence intervals because of the nature of their skewed data distribution (Table 2). For the Sungai Kayu Ara watershed, the optimum λ value would be 0.004 and the best collective representation of S is 31.47 mm (denoted as $S_{0.004}$), with an overall model bias of zero ($\alpha = 0.01$) to model the entire dataset. As $I_a = \lambda S$, substituting λ and S values produces $I_a = 0.11$ mm.

The calibrated SCS urban rainfall–runoff prediction model, when I_a and S are substituted back to Equation (1), would be:

$$Q_{0.004} = \frac{(P - 0.11)^2}{P + 31.36} \tag{7}$$

where

$Q_{0.004}$ = Runoff amount (mm) of the new model formulated with $\lambda = 0.004$.

The urban runoff model calibrated by the SCS is displayed by Equation (7). It is bound to the condition of $P > 0.11$ mm, otherwise $Q_{0.004} = 0$ mm. Equation (7) has an overall model bias value of zero which implies that it does not have an over- or under-prediction tendency on runoff-amount prediction. The correlation between $S_{0.004}$ and $S_{0.2}$ can be determined with SPSS as: $S_{0.2} = 0.901 S_{0.004}^{0.87}$ (R^2_{adj} of 0.97, SE of 0.174, and $p < 0.001$). Subsequently, the equivalent $S_{0.2}$ can be calculated as 18.11 mm leading to the derivation of $CN_{0.2} = 93.35$ with the SCS CN formula: $CN_{0.2} = \frac{25,400}{S_{0.2} + 254}$. The 99% BCa CI of $S_{0.004}$ ranges from 22.7 to 33.7 mm (Table 2). Those values can also be used to calculate its equivalent upper and lower $CN_{0.2}$ limits in the same manner; therefore, the best collective $CN_{0.2} = 93.35$ ($\alpha = 0.01$, 99% CI ranges from 92.96 to 94.91) for the urban watershed under this study (The $CN_{0.2}$ value derivation and SCS CN model calibration steps were summarised and listed in a step-wise instruction format in Appendix A (or authors’ previous publication [6])).

3.4. Asymptotic Curve Number Modelling

Using the AFM, the derived $CN_{0.2}$ values versus rainfall depths graph (Figure 2) resembles the standard behaviour pattern, hence Equation (4) was adopted to derive CN_{∞} as the best representative $CN_{0.2}$ value for the Sungai Kayu Ara watershed and to verify the

modelling result in this section. Using the least-square-fitting method, the fitting parameter k was identified to be 19.23 and CN_{∞} is 90.27. Rounding to the nearest positive integer, $CN_{\infty} = CN_{0.2} = 90$. This CN value is in proximity to the equivalent $CN_{0.2}$ value of 93.35 which was derived by the calibrated SCS runoff model in Section 3.3.

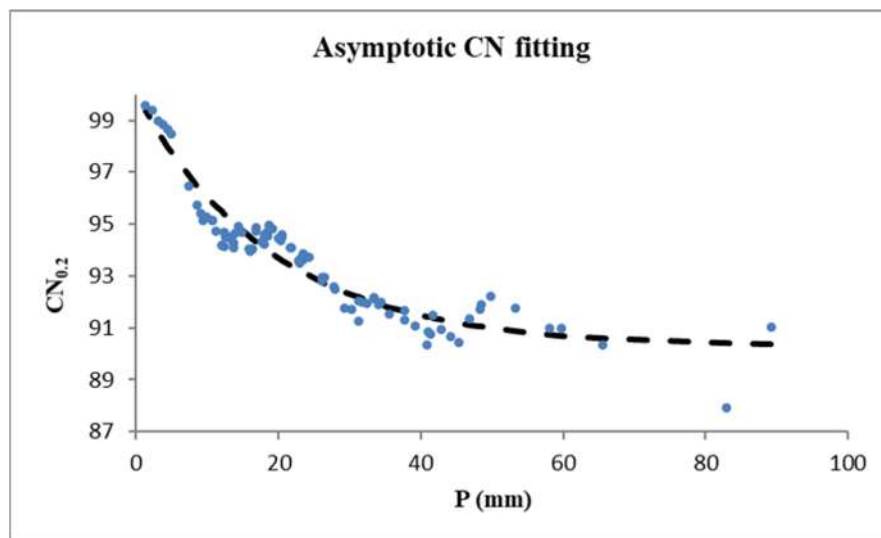


Figure 2. Asymptotic CN fitting of Sungai Kayu Ara watershed: The CN_{∞} resembles standard behaviour pattern with $CN_{\infty} = 90$ near to stable state at the higher rainfall depths.

From the SCS CN formula, the calculated $S_{0.2}$ value is 27.39 mm and $I_a = 0.20 \times 27.39$ mm = 5.48 mm. As such, the AFM runoff predictive model can be formulated from Equation (1) to benchmark the accuracy of the runoff prediction of the original (unsorted) rainfall–runoff dataset against Equations (6) and (7), with the descriptive and inferential statistics of the runoff predictive model’s residual listed in Table 3.

Table 3. Runoff Models’ BCa 99% residual analyses comparison.

	Equation (6)	Equation (7)	AFM Model
$CN_{0.2}$	N/A	93.35	90.27
E	0.832	0.823	0.784
RSS	1945.06	2052.37	2507.72
Residual Standard Deviation:	4.62	4.75	4.630
Standard Deviation CI:	[3.61, 5.70]	[3.64, 5.84]	[3.55, 5.70]
Skewness	−0.52	0.72	0.554
Mean Residual:	0.008	0.000	−2.460
Residual CI:	[−1.31, 1.23]	[−1.31, 1.29]	[−3.78, −1.19]
Residual: Range	26.84	27.98	28.10

Note: Smaller residual standard deviation and residual range imply a better model. Narrower residual interval shows a model with less error distribution while a residual interval that does not span across zero indicates a model with either an over- or under-prediction tendency.

3.5. Runoff Models Comparison

Inferential statistics results of the runoff models’ residual analyses were also generated and tabulated in Table 3. Equation (7), which is the calibrated runoff model (using $\lambda = 0.004$), was benchmarked with the linear intercept model of Equation (6) and AFM. The non-calibrated SCS runoff model of Equation (2) was ruled out as it is not statistically significant, hence it was excluded from the model comparison. The model’s prediction efficiency index (E) as well as its residual sum of squared error (RSS) and descriptive statistics were quantified to extract further comparisons.

From Table 3, the model residual of Equations (6) and (7) have skewness values near to zero, implying that each respective residual distribution is almost normally distributed; therefore, the mean residual can be a good indicator of the prediction accuracy of those

models. The mean residual’s BCa 99% confidence interval range of all three models spanned across zero to indicate the possibility of producing zero residual prediction is significant ($\alpha = 0.01$), hence all models (except AFM) can achieve an accurate runoff prediction.

The AFM runoff model has a higher RSS, with lower E index when compared to the other two models listed in Table 3. The AFM model shows the runoff amount’s under-prediction tendency, as its mean residual’s confidence interval range fluctuates within a negative range and it also has the widest residual range compared to the other models. According to the SCS, there will be no runoff until a rainfall depth is larger than I_a , but the calculated I_a value of the AFM runoff model is larger than six recorded events (6.5%) of the P-Q dataset, while Equation (7) does not have this issue. Although the AFM runoff model derived a proximate $CN_{0.2}$ as the value from Equation (7), a 3.4% increase in the $CN_{0.2}$ value from 90.27 (AFM model) to 93.35 (Equation (7)) was able to improve the runoff model’s E index efficiency by nearly 5% and reduce the RSS by 18% (Table 3).

Up to this point, the linear regression runoff predictive model (Equation (6)) is statistically significant and outperformed against other models to model runoff conditions with the entire dataset. The AFM model is the worst model to predict runoff conditions. To test the robustness of the linear regression runoff predictive model, this study further assessed the runoff prediction ability of Equation (3) against the AFM technique and the newly proposed SCS runoff model calibration method of this study with the same dataset, but regrouped under two different Q/P conditions of >50% and 60%.

4. Further Assessment with Saturation-Excess Runoff Scenarios

Using the same P-Q dataset, the runoff coefficients were calculated to regroup the dataset. The original P-Q dataset has wide distribution range of Q/P from 6% to 97%. Of this, thirty-seven (37) events were found to have Q/P > 50% and twenty-four (24) events were greater than 60%. Thirty-seven (37) and twenty-four (24) P-Q data pairs with runoff coefficients of >50% and >60% were grouped separately to repeat the aforementioned calibration methodology for the rederivation of all rainfall–runoff models for runoff prediction re-assessment. This will further assess the reliability and robustness of all compared runoff predictive models when the watershed becomes increasingly saturated.

4.1. Linear Regression Model

According to the basic linear form of Equation (3), for the given rainfall–runoff dataset, IBM SPSS established the best fitting linear regression model as:

$$\text{For } Q/P > 50\%, Q = 0.654P - 0.216 \tag{8}$$

$$\text{For } Q/P > 60\%, Q = 0.747P - 0.913 \tag{9}$$

Q and P as defined previously. Equation (8) has R^2_{adj} of 0.936 and SE of 3.549 while Equation (9) has R^2_{adj} of 0.977 and SE of 2.242. The descriptive statistics are tabulated in Table 4.

Table 4. Inferential statistics of Equations (8) and (9) at $\alpha = 0.05$.

Model Q/P > 50%	Coeff.	p Value	Model Q/P > 60%	Coeff.	p Value
Constant	−0.216	0.834	Constant	−0.913	0.254
Gradient	0.654	0.000	Gradient	0.747	0.000

From Table 4, the constant coefficients are not significant ($p > 0.05$) under both runoff coefficient scenarios for both equations. When the constant becomes insignificant (x-intercept = 0), the fitting constant parameter will be discarded. The linear intercept runoff model relies on the interception on the x-axis (the constant term) to estimate the local depression, initial loss, or the depression loss of a watershed [30]. In the event when the fitting constant becomes statistically

insignificant, or local depression, initial loss, or the depression loss becomes zero, the only logically hydrological implication is that 100% rainfall becomes complete runoff ($Q = P$) thereafter.

In order to maintain the proposal of [30], the suitable structure of the alternative linear regression is a regression model through the origin (RTO), with the gradient as the only acceptable model-fitting parameter. The data were assessed with IBM SPSS again under both runoff coefficient conditions and identified the best-fitted RTO given by Equations (10) and (11), and their statistics shown in Tables 5 and 6.

$$\text{For } Q/P > 50\%, Q = 0.649P \tag{10}$$

$$\text{For } Q/P > 60\%, Q = 0.724P \tag{11}$$

Table 5. Descriptive statistics of Equations (10) and (11) at $\alpha = 0.05$.

Model	Coeff.	Confidence Interval		p Value
		Lower	Upper	
Q/P > 50% Gradient	0.649	0.616	0.681	0.001
Q/P > 60% Gradient	0.724	0.695	0.753	0.001

Table 6. Inferential statistics of Equations (10) and (11) at $\alpha = 0.05$.

Model	Coeff.	Confidence Interval		BCa p Value
		Lower	Upper	
Q/P > 50% Gradient	0.649	0.605	0.696	0.001
Q/P > 60% Gradient	0.724	0.680	0.756	0.001

According to the assessment statistics in Tables 5 and 6, although both RTO models are significant (p value < 0.01), neither confidence interval span shows a possible inclusion of 1.0 as a fitting gradient for both runoff coefficient scenarios. $Q = P$ is not statistically significant and does not fit the modelling dataset under either condition. As such, the 100% complete runoff scenario becomes impossible and posts a conflict with the proposal from [30]. The validity of Equation (3) for runoff prediction is now in question.

On the other hand, complacent behaviour patterns were detected for both runoff coefficient ($Q/P > 50\%$ and 60%) scenarios with the AFM. It failed to derive a representative $CN_{0.2}$ value as the value reduced according to the increasing rainfall depths and did not approach any stable value. Both the linear regression (two-parameters) model and the AFM model failed to model the runoff conditions and calculate the $CN_{0.2}$ value for the watershed in this study when $Q/P > 50\%$.

4.2. Calibrated SCS Runoff Model

Although Equation (7) is able to model the $Q/P > 50\%$ and $Q/P > 60\%$ datasets with zero bias and an E index of 0.854 and 0.851, respectively, the runoff under-prediction tendency is increasing with Q/P (%), which defeats the aim of developing an effective urban flood predictive model. As such, thirty-seven (37) and twenty-four (24) P-Q data pairs of the Sungai Kayu Ara watershed with $Q/P > 50\%$ and $>60\%$ were grouped separately for λ value rederivation in order to perform SCS model calibration and to formulate statistically significant runoff predictive models again. Optimisation of λ in Equation (1) was conducted within the median confidence interval because of skewed λ datasets of both runoff coefficient scenarios. Tables 7 and 8 illustrate the data distribution of λ values through descriptive statistics analyses conducted with IBM SPSS.

Table 7. Inferential statistics of derived λ at $\alpha = 0.01$ for Q/P > 50% scenario.

Q/P > 50% λ Dataset	Statistics	99% BCa CI of λ	
		Lower	Upper
Mean	0.043	0.011	0.117
Median	0.009	0.005	0.013
Skewness	5.815		
Kurtosis	34.691		
Std. Deviation	0.149		

Table 8. Inferential statistics of derived λ at $\alpha = 0.01$ for Q/P > 60% scenario.

Q/P > 60% λ Dataset	Statistics	99% BCa CI of λ	
		Lower	Upper
Mean	0.061	0.014	0.171
Median	0.013	0.007	0.026
Skewness	4.724		
Kurtosis	22.764		
Std. Deviation	0.184		

The Q/P > 50% dataset (54% dataset with P < 25.4 mm) gave an optimal λ value of 0.005 and an ideal collective S representation of 19.26 mm to yield $I_a = 0.104$ mm. The optimum λ value and best collective representation of S for Q/P > 60% dataset (63% dataset with P < 25.4 mm) are 0.007 and 12.42 mm, which yield $I_a = 0.092$ mm. By substituting I_a and S back into Equation (1), the formulation of the calibrated rainfall–runoff prediction models are:

$$\text{For } Q/P > 50\%, Q_{0.005} = \frac{(P - 0.104)^2}{P + 19.159} \tag{12}$$

$$\text{For } Q/P > 60\%, Q_{0.007} = \frac{(P - 0.092)^2}{P + 12.325} \tag{13}$$

$Q_{0.005}$ = Runoff (mm) where $\lambda = 0.005$.

$Q_{0.007}$ = Runoff (mm) where $\lambda = 0.007$.

Equation (12) bounds to a constraint where $P > 0.104$ mm, else $Q_{0.005} = 0$, whereas Equation (13) bounds to $P > 0.092$ mm, else $Q_{0.007} = 0$. These models were formulated with the extra model bias control factor during the supervised numerical optimisation process under the guiding control of BCa. Both equations have an overall model BIAS value of zero. Conceptualisation of the calibrated SCS runoff prediction Equations (12) and (13) with the optimum λ value hold an identical inherent significant level ($\alpha = 0.01$). Following the equivalent $CN_{0.2}$ derivation process as stated in Section 3.3, Equation (12) yields the equivalent $CN_{0.2}$ value of 94.89, while Equation (13) derives the value of 96.50 to represent the respective ground saturation conditions of the Sungai Kayu Ara watershed. In Tables 7 and 8, the BCa confidence interval ranges of both scenarios do not consist of the λ value of 0.20, hence Equation (2) is still invalid to model either runoff conditions. As such, it will be excluded from runoff model comparison in this section again.

4.3. Model Comparison under Different Saturation-Excess Runoff Scenarios

The newly calibrated runoff models of Equations (12) and (13) were benchmarked against the RTO linear model $Q = P$ only (since the constant-fitting parameter becomes statistically insignificant, refer to Table 4 results) under two different runoff coefficient scenarios. The AFM was excluded as it failed to model the runoff conditions when Q/P > 50% and derive any $CN_{0.2}$ for the watershed. For further comparison, the model’s prediction efficiency index (E), RSS, as well as the predictive model BIAS have been formulated, as seen in Table 9.

Table 9. Runoff predictive model's comparison.

Q/P > 50%	Equation (12)	Q = P	Q/P > 60%	Equation (13)	Q = P
E	0.85	0.11	E	0.94	0.59
BIAS	0.000	10.419	BIAS	0.000	7.588
RSS	1051.23	6319.41	RSS	303.25	2056.83

RTO models failed to achieve runoff prediction accuracy with a runoff over-prediction tendency (positive model bias in Table 9). On the contrary, Equations (12) and (13) managed to predict runoff with high E index, and the E index value even improved from 0.85 to 0.94 as the ground saturation condition increased from 50% to 60%. In this study, the proposed calibrated model is the only runoff predictive model that is capable of modelling urban runoff conditions accurately, even when the watershed becomes increasingly saturated.

5. Discussion

5.1. Hydrological Implication of the New Runoff Predictive Models

The new runoff models (Equations (7), (12) and (13)) were derived to represent different hydrological conditions of the Sungai Kayu Ara watershed. Equation (7) modelled the overall runoff condition of the entire dataset. On the other hand, Equations (12) and (13) represent higher ground saturation conditions where runoff coefficients Q/P are >50% and >60%, respectively. The runoff and incremental trend of all the three models are shown in Figure 3. Equation (7) under-predicted runoff amount significantly when compared to Equations (12) and (13), even though it was under the bias control and guided by BCa.

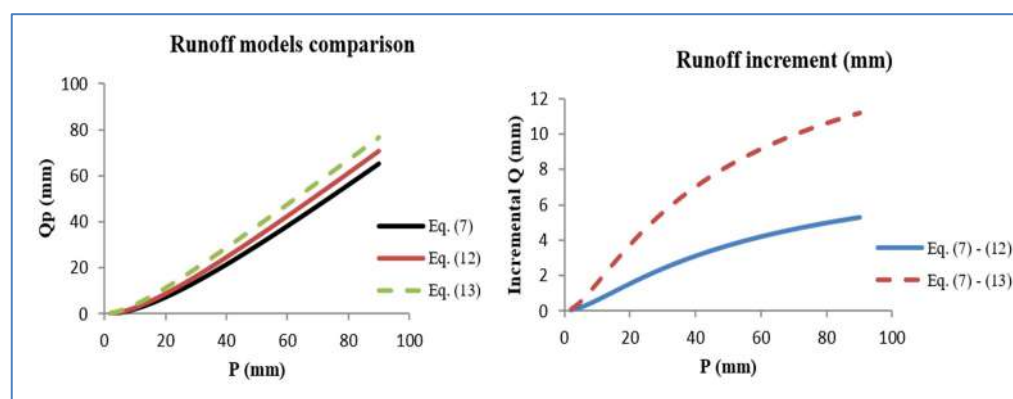


Figure 3. Runoff models using all P-Q data (Equation (7)), Q/P > 50% (Equation (12)), Q/P > 60% (Equation (13)) and runoff depth (mm) increment at Sungai Kayu Ara watershed. By comparing Equations (7) and (13), $CN_{0.2}$ increased 3.4% with an average of 44% (nearly 178,000 m^3) runoff amount increment while $CN_{0.2}$ increased 1.7% from Equations (12) and (13) with an average increment of 15.7% (about 75,000 m^3) in runoff amount. **Note: Runoff depth increment of 1 mm = 22,330 m^3 runoff volume increase at the study site.**

It is noteworthy to highlight that the land-use and land cover of the Sungai Kayu Ara watershed remained the same throughout this study. However, the $CN_{0.2}$ value of this watershed still increased from 93.35 to 94.89 and 96.50 as the watershed became more saturated. This proves that $CN_{0.2}$ value cannot be selected from any handbook according to the land-use and land cover of a watershed only, and it must be derived according to the rainfall–runoff dataset under different ground saturation conditions in order to reflect specific hydrological characteristics of a watershed.

The runoff coefficient distribution for Equation (7) diversifies across a wide range from 6% to 97% (47% on average). Equations (12) and (13) quantify scenarios where the Sungai Kayu Ara watershed becomes more saturated, and the runoff amount from Equation (13) is the highest. The actual incremental runoff from Equations (7)–(13) is also higher than the incremental runoff difference between Equations (7) and (12). Equations (12) and (13)

are as postulated by the “saturation-excess runoff” concept—that the runoff coefficient of a watershed becomes larger when effective soil water-storage reduces and induces higher runoff amounts. This proves that runoff predictive models must be formulated with appropriate datasets to reflect ground saturation conditions accurately. The proposed SCS model calibration methodology in this study cannot be adopted blindly for urban flash flood modelling if the Q/P of the dataset is less than 50%.

5.2. Comparison of Rainfall–Runoff Models

The prospects of the Sungai Kayu Ara watershed utilizing a linear regression model for urban runoff analyses is evaluated in this study. Table 3 tabulated comparisons for runoff models. During the saturation-excess assessment, linear intercept regression models had insignificantly fitted constant terms under both runoff coefficient scenarios (Table 4).

According to [30], the constant-fitting parameter is crucial to the x-intercept approach that represents an estimate of initial loss, local depression, or even watershed depression. The absence of an initial- or depression loss at the watershed according to its P-Q dataset is inferred by an insignificant-fitting constant. Further hydrological interpretation shows a completely saturated watershed with 100% of runoff by any volume of rainfall. The only reasonable linear regression runoff model takes the form of $Q = P$, however its RTO equation-fitting results state otherwise (Tables 5 and 6). Best-regressed RTO linear models from IBM SPSS show no possibility to model complete runoff conditions. For the $Q/P > 50\%$ dataset, at most $Q = 0.696 P$ where 30.4% of initial loss still exists. For the $Q/P > 60\%$ dataset, at most $Q = 0.756 P$ where 24.4% of initial loss still exists. $Q = P$ is statistically insignificant ($p > 0.05$) under both runoff coefficient scenarios. The hydrological implications of the linear regression model conflicts and detaches from the statistical justification. As such, hydrological-condition implications based on those two fitting parameters of the linear regression model become inconsistent and unreliable. The linear regression runoff model failed to describe the hydrological conditions of the Sungai Kayu Ara watershed when runoff coefficient is greater than 50%. On the other hand, the AFM did not outperform against compared models in this study (Table 3). It also failed to model the watershed and derive the $CN_{0.2}$ value when the runoff coefficient is greater than 50%.

The original SCS hypothesis where the value of $\lambda = 0.20$ was met with repeated rejections ($\alpha = 0.01$ level) as the BCa 99% confidence interval span did not show the possibility of having the value of 0.20 (Tables 2, 7 and 8), concluding that Equation (2) was not valid and therefore inapplicable for this study. H_0 also faced rejection ($\alpha = 0.01$ level) as its BCa findings indicated a standard deviation of λ that was not equivalent to zero (Table 2), showing the nature of λ 's value fluctuations, thus, λ does not meet the requirement of a constant as was suggested in 1954 by the SCS, and is, rather, a variable. As a matter of fact, H_0 in this study was rejected, and thus, it opens the opportunity for SCS model calibration. The notion of approaching this matter based on numerical analysis was also utilised in this study, along with guidance of non-parametric inferential statistics, identifying the ideal collective representation of λ and S values to formulate a calibrated runoff predictive model for the Sungai Kayu Ara watershed.

Many researchers in this field suggested different λ values to recalibrate the SCS runoff predictive model. However, the statistical significance of those new values was not reported [6,27]. This study is also in line with the latest findings in this area which reported the detection of multiple CN and I_a values within a watershed and suggested the practice of using multiple CN and I_a values to represent the heterogeneity of a watershed. Those studies also concluded that the SCS CN model must be calibrated according to local rainfall–runoff data to improve the runoff prediction accuracy [33–35]. As such, Equation (2) can no longer be blindly adopted for runoff prediction modelling according to the SCS with $I_a = 0.2S$. The latest findings of [34,35] increased the SCS CN model calibration difficulty level for SCS practitioners because they can only use one I_a value to calibrate Equation (1) [2,9–13]. The non-parametric inferential statistics model calibration guide proposed by this study offers a solution for SCS practitioners to select statistically significant key parameters of S and λ values

from its confidence interval range to calibrate the fundamental SCS CN runoff framework (Equation (1)) according to their rainfall–runoff dataset.

6. Conclusions

1. The hydrological implications of the linear regression model conflicts and detaches from the statistical justification and, therefore, implications based on the two main-fitting parameters of the linear regression model become inconsistent and unreliable. The linear regression runoff model and the AFM technique failed to describe the hydrological conditions of this case study when the runoff coefficient (Q/P) is greater than 50%. On the other hand, $\lambda \neq 0.20$, and the simplified SCS runoff model (Equation (2)) is not statistically significant ($\alpha = 0.01$). These three models are unsuitable in terms of modelling the condition of urban runoff in this study;
2. The decimal $CN_{0.2}$ value can be calculated in order to reflect runoff conditions with higher accuracy. Unlike the antecedent moisture-condition CN concept, $CN_{0.2}$ can be calculated directly and independently to represent the runoff condition of the watershed when it reaches 50% and 60% ground saturation level. The optimum is $\lambda = 0.004$ and $CN_{0.2} = 93.35$ (at $\alpha = 0.01$ level) to model all P-Q data pairs runoff for the Sungai Kayu Ara watershed. Although the land-use and land cover remained the same throughout this study, the $CN_{0.2}$ value of this case study still increased from 94.89 ($\lambda = 0.005$) to 96.50 ($\lambda = 0.007$) as the watershed became more saturated. In line with [14], the $CN_{0.2}$ value was found to be a sensitive parameter. Comparing Equations (7) and (13), the $CN_{0.2}$ value increases by 3.4% with an average of 44% increase in runoff amount, which is almost 178,000 m³ of the runoff increment. Runoff prediction-difference is more profound toward higher-rainfall-depth storm events. As such, Equation (13) should be used to assess the urban drainage capacity and for flash flood prediction, while $CN_{0.2}$ value cannot be decided according to the land-use and land cover conditions only. It can be derived according to the rainfall–runoff dataset and ground saturation conditions in order to reflect the specific hydrological characteristics of a watershed with the proposed method in this study for urban runoff predictions;
3. This study also demonstrated that it is possible to calibrate Equation (1) and to include rainfall depths less than 1 inch (25.4 mm) in order to formulate a statistically significant urban runoff predictive model. Otherwise, nearly 61% (56 events with measurable runoff) of the P-Q data pairs of this study will be discarded for modelling as the corresponding rainfall values are less than 25.4 mm, out of which, 34% (19 events) with $Q/P > 50\%$ and 27% (15 events) with $Q/P > 50\%$ would not be available to formulate Equations (12) and (13) with a sufficient sample size. This study also preserved the parsimonious form of the SCS runoff model for calibration and it emerged as the simplest two-parameter rainfall–runoff predictive model in this study. Flood prediction through rainfall–runoff modelling should be formulated with rainfall–runoff data pairs with runoff coefficients $> 50\%$ instead of using datasets with low runoff coefficients;
4. In general, the proposed methodology in this study is applicable to any urban watershed with a measurable runoff amount (even with rainfall depth < 25.4 mm) and enough datasets (at least 20 events) with $Q/P > 50\%$ or above. It offers a quick and economical runoff assessment in developing areas with rapid land-use and cover change without relying on return period information. Future works may analyse the cost effectiveness of return period-based infrastructure design at urban scale with a longer period rainfall–runoff dataset. The SCS CN model has two parameters (λ and S) only. The proposed model calibration methodology offers a quick quantification of runoff depth from a storm event, whereby SCS practitioners can estimate flood volume at the watershed of interest. Instead of selecting the CN value subjectively by looking at the land-use and land cover, SCS CN model practitioners can derive a range of statistically significant CN values to estimate probable flood volume, assess drainage capacity, and identify probable flood prone area(s). In conjunction, GIS software can be used to assess flood risk, financial losses, and propose needed mitigation strategies;

- The authors caution that there are several limitations to the proposed methodology. The minimum sample size should be at least 20 to achieve meaningful inferential results. The choice of the statistical software must have the option to conduct bootstrapping BCa procedures and provide confidence intervals for median values. The SCS CN lump model must be used with caution when re-creating the specific features of an actual storm as it does not contain time parameters. It is not a precipitation runoff model to model runoff from snowmelt or rain on frozen ground conditions.

Author Contributions: Conceptualisation, L.L. and Z.Y.; methodology, L.L.; software, L.L. and J.L.L.; validation, L.L., S.H.L., R.J.C. and Z.Y.; formal analysis, L.L.; investigation, L.L., S.H.L., R.J.C. and Z.Y.; resources, L.L. and Z.Y.; data curation, L.L.; writing—original draft preparation, L.L. and J.L.L.; writing—review and editing, L.L., J.L.L., R.J.C. and Z.Y.; visualisation, L.L. and J.L.L.; supervision, Z.Y.; project administration, L.L. and Z.Y.; funding acquisition, L.L. and Z.Y. All authors have read and agreed to the published version of the manuscript.

Funding: The research was supported by Ministry of Higher Education (MoHE) through the Fundamental Research Grant Scheme (FRGS/1/2021/WAB07/UTAR/02/1) and was partly supported by the Brunfield Engineering Sdn. Bhd., Malaysia (Brunfield 8013/0002 & 8126/0001).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Please email to request from Ling Lloyd at: linglloyd@utar.edu.my.

Acknowledgments: The authors appreciate the guidance from R. H. Hawkins at The University of Arizona, Tucson, AZ, USA.

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

Appendix A

The CN_{0.2} value derivation and SCS CN model calibration steps can be summarized as below:

- Given that: Effective rainfall (P_e) = $P - I_a$ and $I_a = \lambda S$; Equation (1) can be rearranged as:

$$Q = \frac{P_e^2}{P_e + S}$$
 where $S = \frac{P_e^2}{Q} - P_e$ and $\lambda = \frac{I_a}{S}$
- For each P-Q data pair (P_i, Q_i), calculate corresponding λ_i and S_i value;
- Perform bootstrap, BCa procedure, and normality test in SPSS (version 18.0 or an equivalent statistics software) for (λ_i, S_i);
- Check the normality test results of S_i to see whether it is normally distributed or not:
 - If yes, refer to the mean BCa confidence interval for S_i optimisation.
 - Otherwise, refer to the median BCa confidence interval for S_i optimisation;
- Check the normality test results of λ_i to see whether it is normally distributed or not:
 - If yes, refer to the mean BCa confidence interval for λ_i optimisation.
 - Otherwise, refer to the median BCa confidence interval for λ_i optimisation;
- Substitute the λ_{optimum} and S_{optimum} value into Equation (1) to formulate the calibrated SCS runoff predictive model;
- Given (P_i, Q_i) and λ_{optimum} , compute S_{λ_i} values with
$$S_{\lambda} = \frac{\left[P - \frac{(\lambda-1)Q}{2\lambda} \right] - \sqrt{PQ - P^2 + \left[P - \frac{(\lambda-1)Q}{2\lambda} \right]^2}}{\lambda}$$
- Given (P_i, Q_i) and $\lambda = 0.2$, compute $S_{0.2i}$ values with
$$S_{\lambda} = \frac{\left[P - \frac{(\lambda-1)Q}{2\lambda} \right] - \sqrt{PQ - P^2 + \left[P - \frac{(\lambda-1)Q}{2\lambda} \right]^2}}{\lambda}$$
 again;
- Correlate $S_{0.2i}$ and S_{λ_i} to form a S correlation equation in SPSS (or an equivalent statistics software);
- Substitute the S correlation equation into the SCS curve number formula: $CN_{0.2} = \frac{25,400}{S_{0.2} + 254}$ to derive CN_{0.2} value.

References

1. EM-DAT, CRED/UCLouvain, Brussels, Belgium. International Disasters Database, 1900–2020 Hydrological & Meteorological Categories (Flood, Landslide & Storms). Available online: www.emdat.be (accessed on 16 December 2020).
2. Hawkins, R.H.; Ward, T.J.; Woodward, D.E.; Van Mullem, J.A. *Curve Number Hydrology: State of Practice*; ASCE: Reston, VA, USA, 2009.
3. Hawkins, R.H. Curve number method: Time to think anew? *J. Hydrol. Eng.* **2014**, *19*, 1059. [[CrossRef](#)]
4. Natural Resources Conservation Service (NRCS). *National Engineering Handbook, Part 630 Hydrology*; USDA: Washington, DC, USA, 2004.
5. D’Asaro, F.; Grillone, G. Empirical investigation of curve number method parameters in the Mediterranean area. *J. Hydrol. Eng.* **2012**, *17*, 1141–1152. [[CrossRef](#)]
6. Ling, L.; Yusop, Z.; Yap, W.S.; Tan, W.L.; Chow, M.F.; Ling, J.L. A Calibrated, Watershed-Specific SCS-CN Method: Application to Wangjiaqiao Watershed in the Three Gorges Area, China. *Water* **2020**, *12*, 60. [[CrossRef](#)]
7. Ross, C.W.; Prihodko, L.; Anchang, J.; Kumar, S.; Ji, W.; Hanan, N.P. HYSOGs250m, global gridded hydrologic soil groups for curve-number-based runoff modeling. *Sci. Data* **2018**, *5*, 150091. [[CrossRef](#)]
8. Jaafar, H.H.; Ahmad, F.A.; Beyrouthy, N.E. GCN250, new global gridded curve numbers for hydrologic modeling and design. *Sci. Data* **2019**, *6*, 145. [[CrossRef](#)] [[PubMed](#)]
9. Zelelew, D.G. Spatial mapping and testing the applicability of the curve number method for ungauged catchments in Northern Ethiopia. *Int. Soil Water Conserv. Res.* **2017**, *5*, 293–301. [[CrossRef](#)]
10. Hounkpe, J.; Diekrüger, B.; Afouda, A.A.; Sintondji, L.O.C. Land use change increases flood hazard: A multi-modelling approach to assess change in flood characteristics driven by socio-economic land use change scenarios. *Nat. Hazards* **2019**, *98*, 1021–1050. [[CrossRef](#)]
11. Feng, B.; Wang, J.F.; Zhang, Y.; Hall, B.; Zeng, C.Q. Urban flood hazard mapping using a hydraulic–GIS combined model. *Nat. Hazards* **2019**, *100*, 1089–1104. [[CrossRef](#)]
12. Yalcin, E. Assessing the impact of topography and land cover data resolutions on two-dimensional HEC-RAS hydrodynamic model simulations for urban flood hazard analysis. *Nat. Hazards* **2020**, *101*, 995–1017. [[CrossRef](#)]
13. Durán-Barroso, P.; González, J.; Valdés, J.B. Sources of uncertainty in the NRCS CN Model: Recognition and Solutions. *Hydrol. Process* **2017**, *31*, 3898–3906. [[CrossRef](#)]
14. Boughton, W.C. A Review of the USDA SCS Curve Number Method. *Aust. J. Soil Res.* **1989**, *27*, 511–523. [[CrossRef](#)]
15. Soulis, K.X.; Valiantzas, J.D. Identification of the SCS-CN parameter spatial distribution using rainfall-runoff data in heterogeneous watersheds. *Water Resour. Manag.* **2013**, *27*, 1737–1749. [[CrossRef](#)]
16. Lal, M.; Mishra, S.K.; Pandey, A.; Pandey, R.P.; Meena, P.K.; Chaudhary, A.; Jha, R.K.; Shreevastava, A.J.; Kumar, Y. Evaluation of the Soil Conservation Service curve number methodology using data from agricultural plots. *Hydrogeol. J.* **2016**, *25*, 151–167. [[CrossRef](#)]
17. Betson, R.P. What Is Watershed Runoff? *J. Geophys. Res.* **1964**, *68*, 1541–1552. [[CrossRef](#)]
18. Hewlett, J.D.; Hibbert, A.R. Factors Affecting the Response of Small Watersheds to Precipitation in Humid Areas. In Proceedings of the International Symposium on Forest Hydrology, State College, PA, USA, 31 August–2 September 1967.
19. Miller, J.D.; Kim, H.; Kjeldsen, T.R.; Packman, J.; Grebby, S.; Dearden, R. Assessing the impact of urbanization on storm runoff in a peri-urban catchment using historical change in impervious cover. *J. Hydrol.* **2014**, *515*, 59–70. [[CrossRef](#)]
20. Davidsen, S.; Löwe, R.; Ravn, N.H.; Jensen, L.N.; Arnbjerg-Nielsen, K. Initial conditions of urban permeable surfaces in rainfall-runoff models using Horton’s infiltration. *Water Sci. Technol.* **2018**, *77*, 662–669. [[CrossRef](#)] [[PubMed](#)]
21. Fidal, J.; Kjeldsen, T. Accounting for soil moisture in rainfall-runoff modelling of urban areas. *J. Hydrol.* **2020**, *589*, 125122. [[CrossRef](#)]
22. Sumargo, E.; McMillan, H.; Weihs, R.; Ellis, C.J.; Wilson, A.M.; Ralph, F.M. A Soil Moisture Monitoring Network to Assess Controls on Runoff Generation During Atmospheric River Events. *Hydrol. Process* **2020**, *35*, e13998. [[CrossRef](#)]
23. Yao, L.; Chen, L.; Wei, W. Assessing the effectiveness of imperviousness on stormwater runoff in micro urban catchments by model simulation. *Hydrol. Processes* **2015**, *30*, 1836–1848. [[CrossRef](#)]
24. Ebrahimian, A.; Gulliver, J.S.; Wilson, B.N. Effective impervious area for runoff in urban watersheds. *Hydrol. Processes* **2016**, *30*, 3717–3729. [[CrossRef](#)]
25. Hoang, L.; Schneiderman, E.M.; Moore, K.E.B.; Mukundan, R.; Owens, E.M.; Steenhuis, T.S. Predicting saturation-excess runoff distribution with a lumped hillslope model: SWAT-HS. *Hydrol. Processes* **2017**, *31*, 2226–2243. [[CrossRef](#)]
26. Schneider, L.; McCuen, R.H. Statistical Guidelines for Curve Number Generation. *J. Irrig. Drain. Eng.* **2005**, *131*, 282–290. [[CrossRef](#)]
27. Ling, L.; Yusop, Z.; Ling, J.L. Statistical and Type II Error Assessment of a Runoff Predictive Model in Peninsula Malaysia. *Mathematics* **2021**, *9*, 812. [[CrossRef](#)]
28. Harremoës, P.; Arnbjerg-Nielsen, K. Prediction of Hydrological Reduction Factor and Initial Loss in Urban Surface Runoff from Small Ungauged Catchments. *Atmos. Res.* **1996**, *8*, 137–147.
29. Abustan, I.; Ball, J.E. Modelling the Export of Phosphorous From Urban Catchments. In Proceedings of the HydraStorm 98, Adelaide, Australia, 27–30 September 1998.
30. Huber, W.C.; Dickinson, R.E. *Stormwater Management Model, Version 4, Users Manual*; Environmental Research Laboratory, US Environmental Protection Agency: Athens, GA, USA, 1988.

31. Hjelmfelt, A.T. Curve-number procedure as infiltration method. *J. Hydraul. Div.* **1980**, *106*, 1107–1111. [[CrossRef](#)]
32. Hawkins, R.H. Asymptotic determination of runoff curve numbers from data. *J. Irrig. Drain. Eng.* **1993**, *119*, 334–345. [[CrossRef](#)]
33. ASCE-ASABE (American Society of Agricultural and Biological Engineers)-NRCS (Natural Resources Conservation Service) Task Group on Curve Number Hydrology. *Report of Task Group on Curve Number Hydrology, Chapters 8 (Land Use and Land Treatment Classes), 9 (Hydrologic Soil Cover Complexes), 10 (Estimation of Direct Runoff from Storm Rainfall), 12 (Hydrologic Effects of Land Use and Treatment)*; Hawkins, R.H., Ward, T.J., Woodward, D.E., Eds.; ASCE: Reston, VA, USA, 2017.
34. Santikari, V.P.; Murdoch, L.C. Including effects of watershed heterogeneity in the curve number method using variable initial abstraction. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 4725–4743. [[CrossRef](#)]
35. Hawkins, R.H.; Theurer, F.D.; Rezaeianzadeh, M. Understanding the Basis of the Curve Number Method for Watershed Models and TMDLs. *J. Hydrol. Eng.* **2019**, *24*, 06019003. [[CrossRef](#)]