Article

Statistical and Type II Error Assessment of a Runoff Predictive Model in Peninsula Malaysia

Lloyd Ling [1,*], Zulkifli Yusop [2] and Joan Lucille Ling [3]

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 Centre for Environmental Sustainability and Water Security, Universiti Teknologi Malaysia, Skudai 81310, Malaysia; zulyusop@utm.my
3 Department of Liberal Arts and Languages, American Degree Programme, Taylor's University, Jalan Taylors, Subang Jaya 47500, Malaysia; linglucille@gmail.com
* Correspondence: linglloyd@utar.edu.my

Abstract: Flood related disasters continue to threaten mankind despite preventative efforts in technological advancement. Since 1954, the Soil Conservation Services (SCS) Curve Number ($CN_{0.2}$) rainfall-runoff model has been widely used but reportedly produced inconsistent results in field studies worldwide. As such, this article presents methodology to reassess the validity of the model and perform model calibration with inferential statistics. A closed form equation was solved to narrow previous research gap with a derived 3D runoff difference model for type II error assessment. Under this study, the SCS runoff model is statistically insignificant (alpha = 0.01) without calibration. Curve Number $CN_{0.2}$ = 72.58 for Peninsula Malaysia with a 99% confidence interval range of 67 to 76. Within these $CN_{0.2}$ areas, SCS model underpredicts runoff amounts when the rainfall depth of a storm is < 70 mm. Its overprediction tendency worsens in cases involving larger storm events. For areas of 1 km², it underpredicted runoff amount the most (2.4 million liters) at $CN_{0.2}$ = 67 and the rainfall depth of 55 mm while it nearly overpredicted runoff amount by 25 million liters when the storm depth reached 430 mm in Peninsula Malaysia. The SCS model must be validated with rainfall-runoff datasets prior to its adoption for runoff prediction in any part of the world. SCS practitioners are encouraged to adopt the general formulae from this article to derive assessment models and equations for their studies.

Keywords: rainfall-runoff model; curve number; inferential statistics; 3D runoff difference model; model calibration

1. Introduction

Nearly 8.5 million casualties attributed to flood related disasters were reported between 1990 and 2020 all over the world, which is equivalent to one death every seven minutes. In the recent six decades, about 10,000 cases were reported with 1.3 million deaths and at least $3.3 trillion of financial losses. This financial loss is estimated to be an equivalent rate of almost USD$1800/s [1]. Floods are not only a nuisance to people but also impede the financial well-being, economic development, and natural and cultural heritage preservation efforts of a country. The impact is more profound amidst the COVID-19 pandemic. Uncertainties regarding different scenarios surrounding climate change also require us to safeguard agricultural production and manage water resources wisely to ensure sustainable development for the future. As such, there is an imminent need for hydrologists and modelers to reassess the rainfall-runoff model and improve the modelling approach for better applications in flood prediction.

In order to comply with the federal flood control program in 1954, the United States Department of Agriculture (USDA), Soil Conservation Services (SCS) developed a Curve Number (CN) runoff estimation procedure to implement across the nation. The hydrologic

methods which were originally developed to address specific situations were adopted immediately without professional review and critics [2–5]. The work became the basic CN rainfall-runoff model:

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

Q = Amount of runoff depth (mm)
P = Depth of rainfall (mm)
S = Watershed maximum water retention potential (mm)
I_a = Rainfall initial abstraction amount (mm)

SCS also hypothesized that I_a = λS = 0.2S where λ is the initial abstraction ratio coefficient and fixed at λ = 0.2 as a constant. This equation was tenuously justified with daily rainfall and runoff data. The only official documentation source is the NRCS's National Engineering Handbook, Section 4 (NEH-4) [5]. Its substitution simplifies Equation (1) into the existing SCS CN model as:

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

if P < 0.2S, Q = 0.

The SCS CN methodology has been widely accepted since its inception in 1954. It has been incorporated in various types of software, adopted by many government agencies in design and even appears in every hydrology textbook. However, studies around the world from recent decades reported that Equation (2) inconsistently under and over-predicted runoff results. Curve Number (CN) selection from the SCS handbook for a watershed runoff prediction modelling were reported as subjective and often could not represent other watershed with similar land cover [2–4].

Despite that, many recent studies started to develop and propose extended applications with Equation (2). Some researchers even proposed a global gridded CN concept for runoff modelling [6,7] while others incorporated land-use information in their studies and the GIS modelling technique [8–12]. Contrarily, some reported that the usage of CN in representing a watershed is often contradictory in describing related land cover areas [13]. Some researchers still reported difficulty to calibrate the existing model [14,15] while other studies started to incorporate soil moisture and saturation-excess concepts in their modelling approach [16–19]. US researchers [2,20] were first to conduct large scale studies on the SCS CN model by analyzing more than half a million rainfall events across 24 states in the USA and reported an optimum λ = 0.05 to achieve better runoff modelling results than Equation (2) in USA. To date, SCS practitioners do not have a systematic approach to assess the SCS CN model framework and analyze the impact on runoff prediction when the model is not calibrated.

2. Data and Methods

The SCS CN model (Equation (2)) has been adopted in Malaysia for runoff prediction studies and design. However, no attempt has been made to validate previous study findings by performing hydrological characteristics calibration on the SCS CN model and to derive the λ value with inferential statistics for the entirety of Peninsula Malaysia. The impact of not calibrating the SCS CN model and the blind adoption of Equation (2) for runoff predictions in Peninsula Malaysia are unknown. Therefore, this study extended study results from US researchers [2,20] to develop assessment methods of the SCS CN model for SCS practitioners.

Slightly larger than England (130,395 km^2), the land area of Peninsula Malaysia is 132,265 km^2. It shares a land border with Thailand to the north and Singapore across the strait of Johor to its south. The formation of the Malaysian Department of Irrigation and Drainage (DID) in 1932 assumed all works in connection with drainage and irrigation from the Public Works Department. Flood mitigation and hydrology was made an additional responsibility of DID from 1972 onwards after the declaration of a national disaster due to

severe floods in 1971. From 1986, coastal engineering has become an added function of the DID while river management became its official duty from 1990.

The Department has moved from the Ministry of Agriculture and Agro-based Industry (MOA) to Ministry of Natural Resource and Environment (NRE) on 27 March 2004. Over the years, DID took up new and expanded responsibilities. Today, the DID's duties encompass: River Basin Management and Coastal Zone, Water Resources Management and Hydrology, Flood Management and Eco-friendly Drainage projects in Malaysia.

The rainfall-runoff dataset from the DID, Hydrological Procedure no. 27 (DID HP 27) was used in this study. It is the latest official dataset published by this federal government agency that consists of 227 different storm events recorded between October 1970 to December 2000 from 41 different rural watersheds (Figure 1) across Peninsula Malaysia. The smallest storm event had a rainfall depth of 19 mm with a measurable runoff depth of 4.8 mm while the largest recorded storm event was 420 mm with 258 mm in runoff depth [21].
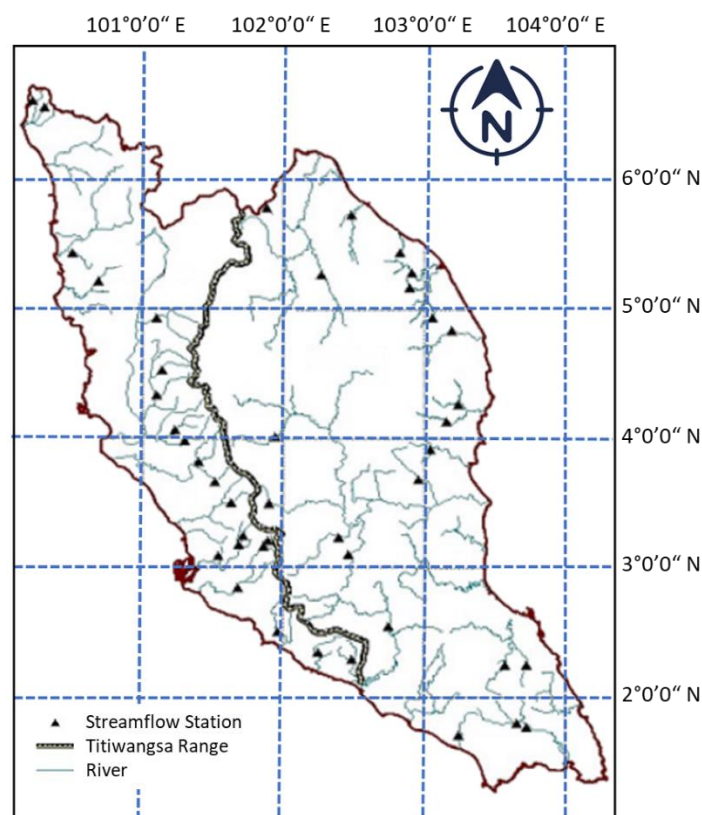
Figure 1. Locations of 41 streamflow stations with 227 rainfall-runoff (P-Q) data pairs used for λ derivation. Modified according to [21].

Objectives of this study are:

1. To assess the 1954 SCS assumption of: $I_a = 0.2S$ in $Q = \frac{(P-I_a)^2}{P-I_a+S}$ and determine its validity for runoff prediction use in Peninsula Malaysia according to the DID HP 27 dataset.
2. To solve the closed form mathematical equation of the "critical rainfall amount" and develop a statistically significant SCS CN model calibration methodology.
3. To assess the impact of not calibrating the existing SCS CN runoff predictive model (Equation (2)) for runoff prediction in Peninsula Malaysia with the official rainfall-runoff dataset from DID HP 27 [21].

2.1. The Reverse Derivation of λ and S Value

In hydrology, the difference between I_a and P is the effective rainfall depth (P_e) to initiate Q thus $P_e = P - I_a$. Substitute this relationship into SCS CN model (Equation (1)), it can be re-arranged to calculate the two key parameters of S and λ values according to the respective P-Q data pair [2,5,22]. Equation (1) can then be expressed as below after the substitution of $P_e = P - I_a$:

$$Q = \frac{(P_e)^2}{P_e + S} \tag{3}$$

rearrange Equation (3) to isolate S as:

$$S = \frac{(P_e)^2}{Q} - P_e \tag{4}$$

Equation (4) is subjected to the constraint where S must be a positive integer. SCS also proposed the correlation of $I_a = \lambda S$ thus λ can be calculated once I_a and S are known by rearranging the equation as:

$$\lambda = \frac{I_a}{S} \tag{5}$$

Equation (5) is subjected to the constraint defined by SCS that $S \geq I_a$ [5], and therefore the range of λ must be (0, 1). The upper limit for λ value is equal to 1 (where $I_a = S$) which is hardly realized in the real world as it implies the condition of a thick canopy interception. The infiltration during early parts of the storm and surface depression storage is equal to the maximum potential retention value (S) of a watershed [5].

Past studies reported different λ values in their work for model calibration. However, the statistical assurance of those new values was hardly mentioned [4]. Latest studies in this area started to report that the modelling approach with multiple CN and I_a values can reflect the heterogeneity of a watershed and the SCS CN model must be calibrated according to local rainfall-runoff data to improve the runoff prediction accuracy. Equation (2) may no longer be valid for runoff prediction modelling [23–25]. SCS defined $I_a = \lambda S$, the existence of multiple I_a values implied that multiple λ and S values can be found within a watershed. These latest study results [24,25] escalate the SCS CN model calibration difficulty to another level as SCS practitioners must identify a best collective representative I_a value to calibrate Equation (1). Therefore, this study proposed to use non-parametric inferential statistics as the guide to make a statistically significant selection of the two key parameters (S and λ values) to calibrate the fundamental SCS CN runoff framework (Equation (1)).

Under the SCS CN hydrological framework, the initial abstraction (I_a) amount must be less than the P value because I_a must first be fulfilled to initiate runoff. Therefore, a reasonable collective representative I_a value for runoff modelling must be less than the minimum P value from the entire P-Q dataset [5]. Given the P-Q dataset, an initial "I_a" value which was less than the minimum P value from the dataset was chosen as the first iterative value in order to calculate the corresponding S and λ values for each P-Q data pair according to Equations (4) and (5). In the event where either constraint in Equation (4) or (5) were to be violated, the "collective representative I_a" value must be reduced until every calculated λ and S values abide to their constraints for each P-Q data pair according to the SCS CN model framework [5].

The alpha value was set at a stringent level of 0.01 in this study to reduce the type I error in null assessment so that the SCS CN model will not be unnecessarily calibrated due to wrong null rejection under objective 1. It will also justify the urgent SCS CN model calibration need to the DID for runoff prediction work in Malaysia, review any past studies and projects that used Equation (2) when the null hypothesis is rejected. This study is only willing to accept 1% error chance because these DID processes are too costly to initiate by mistake.

According to the U.S. Geological Survey (USGS) Statistical methods in water resources guide, the minimum required sample size is 100 to be considered as a large dataset for

water resources related study at the 0.01 alpha level [26]. As such, the DID HP 27 dataset will be sufficient for this study. Given the 227 rainfall-runoff (P-Q) data pairs from DID HP 27, corresponding λ and S values can be calculated. These 227 λ and S values will be bootstrapped independently with the Bias Corrected and Accelerated (BCa) procedure by using the IBM Predictive Analytics software (PASW) version 18.0 (commonly known as SPSS) [27]. The method neither assumes data normality nor has limitation to certain data distribution and performs random sampling with replacement in SPSS [27,28]. In this study, the Mersenne Twister seed number for random sampling generation was set at 2 million (by default) and 10 million to conduct 2000, 5000, and 10,000 sampling for the calculated λ and S dataset.

Consequently, the BCa option in SPSS was used to generate a sampling distribution and 99% confidence interval (CI) to optimize the parameter of interest such as S and λ. Additionally, it provides standard error statistics and CI for the median value, which are unavailable under most parametric tests in SPSS [27]. BCa procedure was chosen by this study for its ability to correct for skewness and bias in the bootstrap distribution [29]. When the dataset has a high positive skewness, BCa can also correct the issue that the bootstrap CI range might be too small [26]. BCa 99% CI has wider range than the 95% CI. Therefore, this study used BCa option in SPSS to generate 99% CI (instead of 95% CI) for both λ and S dataset so that the assessment of the initial claim from SCS that λ = 0.2 can be inferred from the wider BCa CI.

2.2. Supervised Numerical Optimization Analyses

Past researchers faced the dilemma of choosing between the mean and median of a dataset [2,30]. To address this issue, this study utilized an algorithm of numerical analysis guided by inferential statistics for decision making.

λ and S were optimized using Equation (1) with a supervised numerical analyses approach. To prevent the optimization algorithm from focusing on residual sum of squares (RSS) minimization only, the overall model bias (BIAS) will be minimized near to the value of zero concurrently during the parameter optimization process. This acts as a check with the BCa technique to ensure that the optimized λ and S value are not biased towards the dataset during the SCS model calibration. In the event of skewed data nature, the supervised numerical optimization would be conducted to search for an optimum value within the BCa median's confidence interval limits of the derived λ and S dataset, respectively. The optimized S value and its confidence interval range will lead to the calculation of CN value to represent the entire DID HP 27 dataset in Peninsula Malaysia (see Section 3.2).

2.3. Null Hypotheses Assessments with Inferential Statistics

A Null hypothesis was set up to assess the 1954 SCS proposal with inferential statistics as below:

H_0: I_a = λS where λ must be 0.2 in Equation (1) (as proposed by SCS) to model runoff conditions according to the DID HP 27 dataset in Peninsula Malaysia.

H_0 assesses the validity of Equation (2) for this study as pertained to the DID HP 27 dataset. The assessment of H_0 will be inferred from the BCa confidence interval of λ [28]. The rejection of H_0 indicates that the SCS CN model (Equation (2)) is invalid to model the dataset of this study. It requires the acceptance of H_0 to adopt Equation (2) for rainfall-runoff modelling while the rejection of H_0 will pave a way to derive a new λ value for the DID HP 27 dataset. The optimized λ and S values will be used to formulate a new calibrated runoff prediction model for Peninsula Malaysia. SCS practitioners are encouraged to validate the existing SCS CN model (Equation (2)) prior to runoff modelling adoption.

2.4. The S General Formula

Equation (1) was re-arranged into a general form of $S_\lambda = f(P, Q, \lambda)$ in a previous study [4]. When $\lambda = 0.2$, the corresponding $S_{0.2}$ value leads to the derivation of conventional CN values in use by SCS practitioners. Any other λ values will result in S_λ leading to the derivation of CN_λ values which are different from the SCS tabulated CN values. The general S_λ formula (see [4] for derivation steps) used by this study is:

$$S_\lambda = \frac{\left[P - \frac{(\lambda-1)Q}{2\lambda}\right] - \sqrt{PQ - P^2 + \left[P - \frac{(\lambda-1)Q}{2\lambda}\right]^2}}{\lambda} \tag{6}$$

S_λ = Total abstraction amount of any λ value (mm).

2.5. Correlation Between S_λ and $S_{0.2}$

According to previous researchers, when the optimum λ value is different from the conventional value where $\lambda = 0.2$, a correlation between the newfound λ value and 0.2 must be used in order to calculate the curve number again [2,3,20]. US researchers termed the batch of curve numbers derived from any λ value other than $\lambda = 0.2$ as "conjugate curve numbers" denoted by CN_λ which are different from the SCS tabulated curve numbers [2–4,20]. Given the P-Q dataset, S_λ and $S_{0.2}$ can be calculated using Equation (6). A correlation between the S_λ and $S_{0.2}$ dataset must be established before the calculation of conventional CN value (see Section 3.2). SCS practitioners must use the correlation equation between the S_λ and $S_{0.2}$ to calculate the conventional CN value to avoid the mistake of using conjugate curve number in their study.

2.6. The 3D Runoff Difference Model

Using P-Q datasets from multiple watersheds or from multiple locations within a watershed, a 3D runoff difference model can be created as a collective visual representation of multiple rainfall depths to compare with different $CN_{0.2}$ scenarios. If Equation (2) fails the Null assessment, this 3D model can reflect the runoff difference between it and the new calibrated runoff model for further analyses. The model will be a guide to visualize the runoff under and over prediction zones between two models. In 1954, SCS correlated S and CN. The SI unit version of the formula is:

$$S = \frac{25,400}{CN} - 254 \tag{7}$$

Equation (7) was derived from the SCS assumption where $\lambda = 0.2$, and therefore it will be more appropriate to denote CN as $CN_{0.2}$ and S with $S_{0.2}$. Substituting Equation (7) into Equation (2), the SCS model can be simplified to become: $Q_{0.2} = f(P, CN_{0.2})$ and represented in SI form of:

$$Q_{0.2} = \frac{\left[P - 50.8\left(\frac{100}{CN_{0.2}} - 1\right)\right]^2}{\left[P + 203.2\left(\frac{100}{CN_{0.2}} - 1\right)\right]} \tag{8}$$

$Q_{0.2}$ = Runoff depth (mm) of $\lambda = 0.2$
where $P > 0.2\,S_{0.2}$ else $Q_{0.2} = 0$.

The general form of Equation (1) after the substitution of $I_a = \lambda S$ for any λ value becomes:

$$Q_\lambda = \frac{(P - \lambda S_\lambda)^2}{P - \lambda S_\lambda + S_\lambda} \tag{9}$$

where $P > \lambda S_\lambda$, else $Q_\lambda = 0$. As such, the runoff difference between SCS model (uncalibrated) and the new calibrated runoff model (with new λ) can be quantified as the difference between Equations (8) and (9) as:

$$Q_v = \frac{\left[P - 50.8\left(\frac{100}{CN_{0.2}} - 1\right)\right]^2}{\left[P + 203.2\left(\frac{100}{CN_{0.2}} - 1\right)\right]} - \frac{(P - \lambda S_\lambda)^2}{P - \lambda S_\lambda + S_\lambda} \qquad (10)$$

Q_v = Runoff depth prediction difference between 2 runoff models (mm)
$CN_{0.2}$ = the conventional curve number

As Equation (2) was widely adopted in many countries, it is important to assess the runoff prediction difference with Equation (10). It is a general equation that can be used by SCS practitioners to determine the impact of not calibrating Equation (2) for runoff predictions under their study.

In Equation (10), Q_v will be positive when the conventional SCS runoff model (Equation (2)) over-predicted runoff when compared to the calibrated new runoff equation and vice versa. If the newly derived $\lambda < 0.2$, Equation (10) is subject to the constraint where $P > \lambda S$. When the new derived $\lambda > 0.2$, Equation (10) will abide to the constraint of $P > 0.2S_{0.2}$, else $Q_v = 0$ because there is no runoff difference as I_a of the lower λ value model is yet to be fulfilled to initiate the runoff process [2,5] and produce a runoff difference between two runoff models. All in all, the smaller λ runoff model will initiate runoff ahead of the larger λ runoff model [5].

2.7. Outer Boundary Equation

Equation (2) is subject to a constraint where $P > I_a$ or $P > \lambda S_\lambda$, else $Q_\lambda = 0$. The 3D runoff difference model captures the runoff difference of two different runoff models. When the I_a constraint of the lower λ value model has been fulfilled, runoff will be initiated. Base on this concept, the I_a constraint of the lower λ value model becomes the outer boundary of the 3D runoff difference model which also represents the runoff indifference boundary with the following general equation:

$$P = \lambda S_\lambda \qquad (11)$$

2.8. Inner Boundary Equation

The second boundary is the "Inner Boundary" of the 3D runoff difference model. This boundary separates the runoff under-prediction zone from the over-prediction zone of the SCS runoff model. The runoff difference is equal to zero at the crossover boundary, which is also known as the runoff indifference boundary. Therefore, when $Q_v = 0$ (runoff indifference) in Equation (10), the form can be re-expressed as:

$$\frac{\left[P - 50.8\left(\frac{100}{CN_{0.2}} - 1\right)\right]^2}{\left[P + 203.2\left(\frac{100}{CN_{0.2}} - 1\right)\right]} = \frac{(P - \lambda S_\lambda)^2}{P - \lambda S_\lambda + S_\lambda} \qquad (12)$$

Equations (11) and (12) are also general equations that can be used by SCS practitioners to analyze the 3D runoff difference model (created with Equation (10)) in their study.

2.9. Models Comparison

Runoff models are compared and benchmarked for their model predictive accuracy in this paper. Model's residual sum of squares (*RSS*), predictive model *BIAS* prediction and model efficiency index (*E*), also known as Nash–Sutcliffe index, were calculated with the following formulae to draw further comparison between them.

$$RSS = \sum_{i=1}^{n}\left(Q_{predicted} - Q_{observed}\right)^2 \qquad (13)$$

$$E = 1 - \frac{RSS}{\sum\limits_{i=1}^{n}\left(Q_{predicted} - Q_{mean}\right)^{2}} \tag{14}$$

$$BIAS = \frac{\sum\limits_{i=1}^{n}\left(Q_{predicted} - Q_{observed}\right)}{n} \tag{15}$$

n = Total number of data pairs.

Lower *RSS* implies a better model. Index *E* lies on a spectrum of minus 1.0 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. *BIAS* is the overall model prediction error indicator. Zero *BIAS* value indicates an error free model prediction while negative value indicates the overall predictive model's under-prediction tendency and vice versa.

2.10. Asymptotic Curve Number Fitting

Other than numerical optimization technique, many researchers [31–35] used asymptotic CN fitting method (AFM) to determine the best representative CN for the watershed of interest with P-Q dataset (λ value remains as 0.20 under this method). Therefore, AFM will be used to benchmark against the proposed method in this article. Under AFM, CN cannot be determined for the Complacent behavior watershed, but Standard behavior watershed follows the following formula [33]:

$$CN(P) = CN_{\infty} + (100 - CN_{\infty})e^{\left(-\frac{P}{k}\right)} \tag{16}$$

CN(P) = Fitted CN value of a specific rainfall depth
CN_{∞} = CN of a watershed of interest
K = Fitting parameter

Violent behavior watershed follows the following formula [33]:

$$CN(P) = CN_{\infty}\left[1 - e^{-k(P-P_{th})}\right] \tag{17}$$

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

2.11. Critical Rainfall Amount (P_{crit})

The concept of P_{crit} was initially suggested by US researchers [2,20,22] which can only be obtained through numerical analysis solving technique or by trial and error procedure. In their work, optimum λ was reported as 0.05 and the P_{crit} points were identified through the intersection of conjugate $CN_{0.05}$ and $CN_{0.2}$ curve on the graph in their study.

The concept of P_{crit} was built upon the runoff indifference between 2 runoff models. When $Q_v = 0$ (runoff indifference between two runoff models), Equation (10) becomes Equation (12). As such, this study introduces runoff difference curves which was created with numerical analysis technique as the visual presentation of Equation (12). Runoff difference curves can be plotted for specific $CN_{0.2}$ classes across multiple rainfall depth scenarios. Unlike previous research work, it combined two curves into a single curve and identify P_{crit} at where the curve crosses the x-axis.

2.12. The Closed Form Equation of Critical Rainfall Amount (P_{crit})

Through algebraic manipulation, this study successfully rearranged Equation (10) and solved the general closed-form equation of P_{crit} in terms of $CN_{0.2}$ when $Q_v = 0$. The breakthrough has also proven to be able to solve for P_{crit} value precisely of any pairing runoff models and replace the trial and error procedure used by previous researchers [2,20,22]. SCS practitioners can derive the P_{crit} equation for their study with proposed method in this article (see Section 3.10).

2.13. Critical Curve Number (CN$_{crit}$)

With a similar concept (based upon Equation (12)) as the critical rainfall amount (P$_{crit}$), this study also introduces "critical curve number(s)" (CN$_{crit}$) to supplement the use of P$_{crit}$. Under a specific rainfall scenario, critical curve number value(s) can also be identified from the points where Q$_v$ = 0 between 2 runoff models. Unlike the success of the P$_{crit}$ closed-form equation derivation, the effort to realize the closed-form equation of CN$_{crit}$ in term of P is still unfruitful to date. Therefore, the numerical analysis technique was applied to estimate CN$_{crit}$ value(s) with visual aid from the runoff difference curves graph. Runoff difference curves methodology as Section 3.9 covered can be adopted to show that Equation (2) or Equation (8) will under-predict runoff amount in any curve number areas below the critical curve number value and vice versa.

2.14. Soft Computing and Data Mining of the 3D Model

In general, Equation (10) represents the runoff prediction errors of Equation (2) under multiple P and CN$_{0.2}$ scenarios but it is difficult to visualize the quantified effect by looking at Equation (10) and solve for the global maxima and minima in order to represent the worst under and over runoff prediction amounts between two runoff models.

Based on the rainfall depth range of the dataset [21], a numerical table can be compiled with Equation (10) through the substitution of different P, CN$_{0.2}$ scenarios and the λ value to quantify runoff depth prediction difference between two runoff models in a table. A 3D model can also be constructed with the collective information from the table (Section 3.7). With the visual aid of a 3D runoff difference model, it is possible to extract all minimum and maximum runoff prediction difference amount and represent them with statistically significant equations. The minimum under-prediction difference amount equation represents the worst under-design case incurred by Equation (2) and vice versa.

3. Results and Discussion

3.1. The Reverse Derivation of Optimum λ and S for Peninsula Malaysia

In all, 227 λ and S values were calculated according to corresponding rainfall-runoff (P-Q) data pairs. The calculated λ dataset was checked for normality in SPSS with Kolmogorov–Smirnov and Shapiro–Wilk test statistics, both tests concluded the λ dataset to be non-normal ($p < 0.001$). Nearly 95% (214 out of the 227) storm events calculated λ value below 0.2 while none was equal to 0.2 as proposed by SCS.

According to Section 2.1, as defined by the SCS [5], the "collective representative I$_a$" was reduced to 5.9 mm to fulfil both constraints of Equations (4) and (5) for the entire dataset of DID HP 27 [21]. 227 calculated λ and S values were independently used for 2000, 5000, and 10,000 random samplings prior to CI generations and cross checking (This study found that the CI upper and lower limits only differ at the fourth decimal places with 2000, 5000, and 10,000 random samplings while there were no difference between the use of 2 million (by default) and 10 million Mersenne Twister seed numbers for random sampling generation) in SPSS. The inferential statistics of the derived λ and S values are tabulated in Tables 1 and 2.

Table 1. Inferential Statistics of the derived λ dataset from Malaysian Department of Irrigation and Drainage (DID) Hydrological Procedure (HP) 27.

λ	Statistics	Bootstrap, BCa 99%			
		Bias	Std. Error	Confidence Interval	
				Lower	Upper
Skewness	5.125				
Kurtosis	36.456				
Mean	0.071	−0.00006	0.006	0.056	0.089
Median	0.042	0.00023	0.003	0.034	0.051

From Table 1, neither the mean nor the median BCa λ's 99% CI include the λ value of 0.2 (In comparison, the BCa 95% mean and the median CI for λ span across smaller range (0.036, 0.084)). Therefore, H_0 can be rejected at alpha = 0.01 level. As such, Equation (2) is statistically insignificant (not even significant at alpha = 0.05) and cannot be used to predict runoff conditions in this study. λ dataset is skewed (skewness of 5.125 in Table 1) thus the search of the optimum collective representative λ value via numerical optimization technique focusses on median λ's confidence interval [0.034, 0.051].

On the other hand, data distribution of the S dataset is somewhat skewed with a skewness of 1.624 (Table 2). The definition of skewness is non-uniform, some guidelines suggested skewness value less than 3.0 to be considered as normal while some set a more stringent limit at 1.0. To avoid the ambiguity of skewness determination, the search of the optimum S value was widened to include the lowest and the highest confidence interval limit of both mean and median values (118.125, 196.332) on S [2,30].

Table 2. Inferential Statistics of derived S dataset from DID HP 27.

S	Statistics	Bootstrap, BCa 99%			
		Bias	Std. Error	Confidence Interval	
				Lower	Upper
Skewness	1.624				
Kurtosis	4.392				
Mean	172.297	0.002	8.649	150.952	196.332
Median	141.54	−0.053	10.005	118.125	170.170

The optimum λ value was recognized as 0.051 (rounded) while 150.46 mm was the optimum S value in formulating the best runoff predictive model (based on Equation (1)) according to the entire dataset of DID HP 27 with an overall predictive model's *BIAS* near to zero. The collective representation of the I_a for the entire dataset was found from the product of the optimum λ and S and therefore, the best collective representative value of I_a to model the entire dataset in Peninsula Malaysia is 8.3 mm from this study.

As mentioned in Section 2.1 and 2.2, BCa technique produced confidence intervals (Tables 1 and 2) for the optimization of λ and S value to calibrate the SCS CN model. It also generated a range of λ and S value to enable the calculation of multiple I_a and CN values which is in line with the latest research development in this area [23–25]. Other than the best collective representative I_a value, SCS practitioners who use the proposed method in this article have an option to compare other possible I_a values with other research results in future.

3.2. The Correlation between S_λ and $S_{0.2}$ for Peninsula Malaysia

The derivation of S_λ formula (Equation (6)) proved mathematically that even with the same P-Q dataset, as λ varies, the corresponding total abstraction amount (S) varies as well and therefore, the corresponding CN value will change also. As such, it is more appropriate to re-represent Equation (7) in general form as:

$$CN_\lambda = \frac{25,400}{S_\lambda + 254} \tag{18}$$

CN_λ = Curve number of any λ value (dimensionless)
S_λ = Total abstraction amount of any λ value (mm)

Given the P-Q dataset and λ value, the corresponding CN_λ can be derived from Equation (18). When λ = 0.2, its corresponding $S_{0.2}$ value gives rise to deriving the conventional curve number compiled by SCS. To differentiate the conventional SCS CN, the notation of "$CN_{0.2}$" is used in the remaining of this paper. When $\lambda \neq 0.2$, its corresponding S_λ value

derives "Conjugate Curve Number" (CN_λ) [2,20,22]. As the optimum λ value = 0.051, the correlation between $S_{0.051}$ and $S_{0.2}$ was identified with SPSS for this study as:

$$S_{0.051} = 1.176 S_{0.2}{}^{1.063} \qquad (19)$$

$S_{0.051}$ = Total abstraction amount (mm) of λ = 0.051
$S_{0.2}$ = Total abstraction amount (mm) of λ = 0.2

Equation (19) has a R^2-adj of 0.946, standard error of 0.15 and $p < 0.001$. Equation (19) is also the key to convert $S_{0.051}$ back to its equivalent $S_{0.2}$ value for the calculation of $CN_{0.2}$ for SCS practitioners. The optimum $S_{0.051}$ is 150.46 mm (alpha = 0.01) from the range of 118.125 to 196.332 (Table 2) in Section 3.1. The equivalent $S_{0.2}$ value of $S_{0.051}$ = 150.46 mm is 95.97 mm (calculated from Equation (19)). By substituting $S_{0.2}$ = 95.97 mm into Equation (18), $CN_{0.2}$ = 72.58; thus, new λ of 0.051 derives an equivalent $CN_{0.2}$ value of 72.58 to model the entire DID HP 27 dataset. The 99% confidence interval of $S_{0.051}$ ranges from 118.125 to 196.332, those values can also be used to calculate its equivalent upper and lower $CN_{0.2}$ limits in the same manner through Equation (18) and therefore, for the DID HP 27 dataset [21], the best collective $CN_{0.2}$ = 72.58 (99% CI ranges from 67 to 76) for runoff predictions in Peninsula Malaysia.

3.3. Conjugate Curve Numbers (CN_λ) for Peninsula Malaysia

Given the P-Q data pairs from DID HP 27, conjugate curve number values (CN_λ) of each storm event can be calculated with aforementioned equations in the following steps:
Since the optimum λ value obtained was 0.051, Equation (18) becomes:

$$CN_{0.051} = \frac{25,400}{S_{0.051} + 254}$$

Substitute Equation (19) into Equation (18) will yield:

$$CN_{0.051} = \frac{25,400}{(1.176 S_{0.2}{}^{1.063}) + 254} \qquad (20)$$

where $S_{0.2}$ values can be calculated using Equation (6) (the S general formula) when P-Q data pairs are given. $CN_{0.051}$ is the conjugate curve number of $CN_{0.2}$. Equation (20) proves that conjugate curve number (CN_λ) is not the same as the conventional curve number $CN_{0.2}$ which was derived using Equation (7). Thus, it is inappropriate to use any conjugate curve number (CN_λ) with Equation (2) in any rainfall-runoff modelling work.

3.4. The 3D Runoff Difference Model for Peninsula Malaysia

According to the discussions from Sections 2.4 and 2.5, the S amount is specific to its corresponding λ value. The optimum λ value = 0.051 to model runoff conditions for the DID HP 27 dataset thus by substituting λ with 0.051 into Equation (9) yields a calibrated rainfall-runoff predictive model on Equation (1) in the form of:

$$Q_{0.051} = \frac{(P - 0.051 S_{0.051})^2}{P - 0.051 S_{0.051} + S_{0.051}}$$

The substitution of Equations (19) and (7) further simplifies it as:

$$Q_{0.051} = \frac{\left[P - 21.606 \left(\frac{100}{CN_{0.2}} - 1 \right)^{1.063} \right]^2}{\left[P + 402.547 \left(\frac{100}{CN_{0.2}} - 1 \right)^{1.063} \right]} \qquad (21)$$

Equation (21) re-expressed the runoff model in term of P and $CN_{0.2}$ and subjects to the constraint.

$P > 21.606 \left(\frac{100}{CN_{0.2}} - 1 \right)^{1.063}$ else $Q_v = 0$ on the 3D model

$CN_{0.2}$ = Conventional SCS tabulated curve number

$Q_{0.051}$ = Runoff depth (mm) of $\lambda = 0.051$

Equation (8) is the re-expression of Equation (2) in term of P and $CN_{0.2}$.

$$Q_{0.2} = \frac{\left[P - 50.8 \left(\frac{100}{CN_{0.2}} - 1 \right) \right]^2}{\left[P + 203.2 \left(\frac{100}{CN_{0.2}} - 1 \right) \right]} \tag{22}$$

It subjects to the constraint $P > 50.8 \left(\frac{100}{CN_{0.2}} - 1 \right)$ else $Q_v = 0$.

Equation (8) or Equation (2) represents the un-calibrated SCS CN model. The runoff depth prediction differences between Equations (8) and (21) were collectively quantified by Equation (22) of which the 3D runoff difference model (Section 3.7 and Figure 2) was constructed with. Equation (22) also quantifies type II errors from Equation (2) (existing SCS model) if it is not calibrated for runoff prediction in Peninsula Malaysia.

3.5. Outer Boundary Equation

As per Section 2.7, the calibrated new λ value (0.051) is less than 0.2; thus, its model's constraint can be adopted to represent the runoff indifference boundary where runoff has not been initiated. Therefore, Equation (22) is also subject to the constraint, $P > 0.051$ $S_{0.051}$ or $P > 21.606 \left(\frac{100}{CN_{0.2}} - 1 \right)^{1.063}$ else $Q_v = 0$. Equation (19) can be substituted into 11 to preserve the conventional curve number ($CN_{0.2}$) through following the steps.

(a)

Figure 2. *Cont.*

Runoff Difference between Eq. 2 (or 8) and 21 (new λ = 0.051) for Peninsula Malaysia
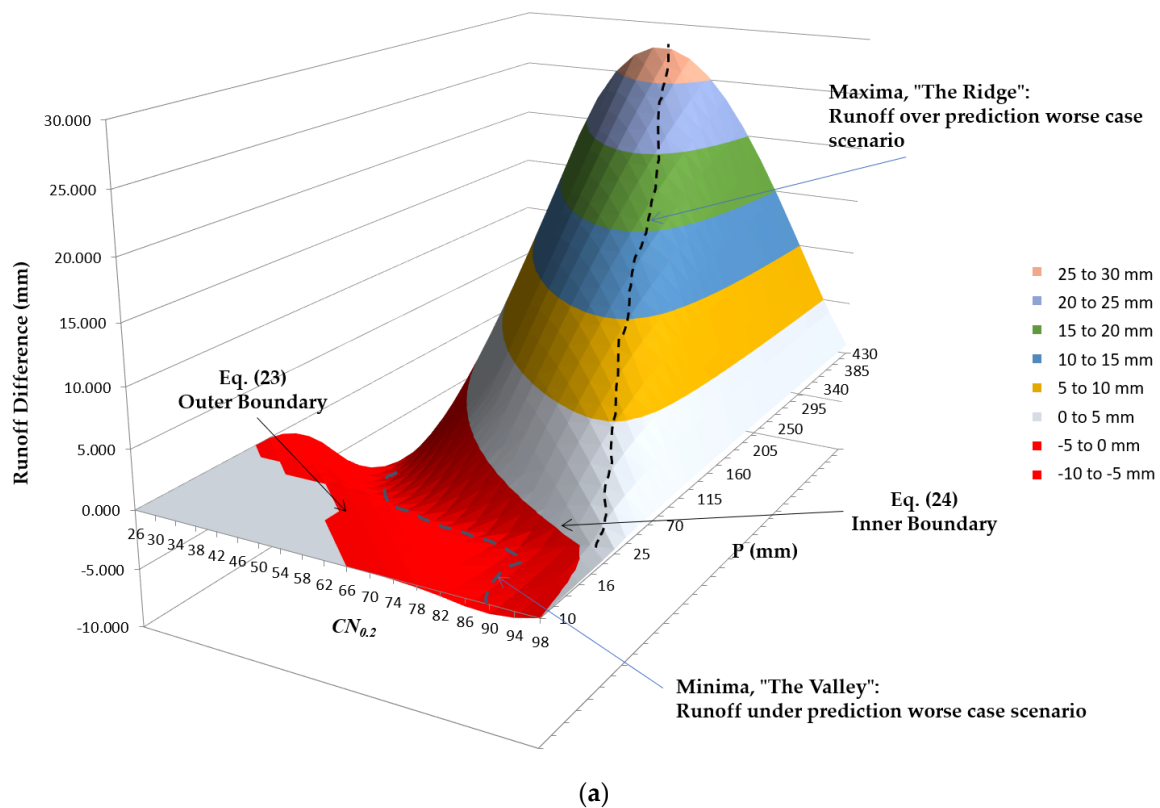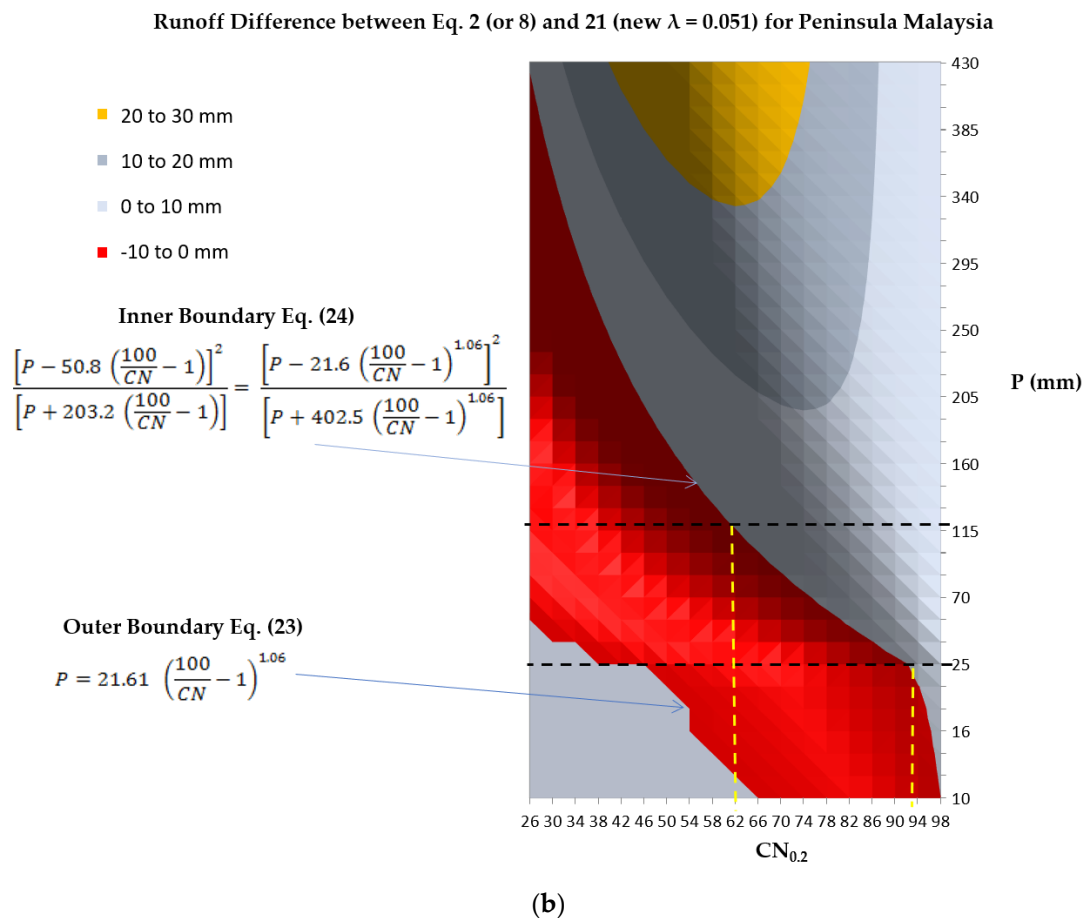
Figure 2. (**a**) The 3D runoff difference model (between Equations (2) and (21)) of Peninsula Malaysia with DID HP 27 dataset for Type II error assessment. (**b**) Top view of the 3D runoff difference model for Peninsula Malaysia with DID HP 27 dataset.

Substitute λ with 0.051, Equations (7) and (19) into Equation (11) yields:

$$P = 21.606 \left(\frac{100}{CN_{0.2}} - 1 \right)^{1.063} \tag{23}$$

Equation (23) is the runoff indifference boundary equation between two runoff models. It is otherwise recognized as the "Outer Boundary" equation of the 3D runoff difference model (Figure 2a,b).

3.6. Inner Boundary Equation

When $Q_v = 0$ in Equation (22), the form can be expressed as:

$$\frac{\left[P - 50.8 \left(\frac{100}{CN_{0.2}} - 1 \right) \right]^2}{\left[P + 203.2 \left(\frac{100}{CN_{0.2}} - 1 \right) \right]} = \frac{\left[P - 21.606 \left(\frac{100}{CN_{0.2}} - 1 \right)^{1.063} \right]^2}{\left[P + 402.547 \left(\frac{100}{CN_{0.2}} - 1 \right)^{1.063} \right]} \tag{24}$$

Equation (24) is also known as the "Inner Boundary" equation of the 3D runoff difference model for Peninsula Malaysia that demarcates the runoff under-prediction and over-prediction zones between two runoff models in this study.

3.7. The Construction of the 3D Runoff Difference Model

DID HP 27 dataset consist of 227 storm events ranging from 19 mm to 420 mm. In order to analyze and quantify the runoff prediction depth difference between Equation (2)

(or Equation (8)) and 21 under multiple rainfall and $CN_{0.2}$ scenarios, rainfall depth (P) ranging from 10 mm to 430 mm across different $CN_{0.2}$ values (from 26 to 98) were entered into Equation (22) to calculate the runoff depth prediction difference that can be found in Figure 3. Those tabulated values are runoff prediction errors (or type II errors) from Equation (2) which are in line with previous studies that reported more profound error in forested watersheds represented by $CN_{0.2}$ values < 60 [2,20,22] Similarly, for Peninsula Malaysia, both runoff under and over prediction errors worsen when the value of CN0.2 reduces (Figure 3).

P (mm)	CN 26	30	34	38	42	46	50	54	58	62	67	72	73	76	82	86	90	94	98
10												-0.028	-0.042	-0.105	-0.359	-0.620	-0.698	-0.488	0.001
12											-0.017	-0.105	-0.134	-0.246	-0.629	-0.808	-0.762	-0.429	0.068
14										-0.005	-0.072	-0.230	-0.275	-0.444	-0.859	-0.933	-0.773	-0.344	0.128
16									-0.002	-0.039	-0.165	-0.401	-0.464	-0.696	-1.033	-1.007	-0.745	-0.244	0.181
18									-0.024	-0.104	-0.296	-0.615	-0.699	-0.952	-1.159	-1.039	-0.687	-0.134	0.228
20								-0.009	-0.072	-0.198	-0.461	-0.872	-0.963	-1.165	-1.245	-1.036	-0.606	-0.021	0.269
25							-0.027	-0.127	-0.302	-0.561	-1.025	-1.429	-1.474	-1.533	-1.313	-0.913	-0.340	0.264	0.354
40				-0.019	-0.151	-0.400	-0.766	-1.252	-1.816	-2.167	-2.275	-2.050	-1.968	-1.654	-0.757	-0.034	0.647	1.008	0.511
55		-0.003	-0.146	-0.475	-0.970	-1.623	-2.371	-2.929	-2.982	-2.873	-2.422	-1.685	-1.509	-0.938	0.305	1.072	1.589	1.558	0.597
70	-0.015	-0.267	-0.781	-1.516	-2.456	-3.252	-3.609	-3.328	-2.806	-1.896	-0.788	-0.553	0.162		1.499	2.149	2.391	1.965	0.652
85	-0.285	-0.942	-1.894	-3.094	-3.961	-4.284	-4.187	-3.764	-3.087	-2.219	-0.957	0.387	0.654	1.423	2.680	3.127	3.060	2.275	0.689
100	-0.890	-2.011	-3.458	-4.533	-4.910	-4.772	-4.249	-3.440	-2.427	-1.281	0.231	1.698	1.972	2.728	3.793	3.996	3.619	2.517	0.716
115	-1.817	-3.457	-4.937	-5.497	-5.395	-4.819	-3.904	-2.757	-1.466	-0.110	1.561	3.060	3.324	4.019	4.821	4.763	4.090	2.711	0.737
130	-3.056	-5.048	-6.000	-6.053	-5.491	-4.506	-3.241	-1.805	-0.292	1.214	2.966	4.423	4.665	5.263	5.761	5.440	4.491	2.871	0.753
145	-4.596	-6.292	-6.697	-6.257	-5.260	-3.902	-2.328	-0.651	1.031	2.633	4.400	5.760	5.969	6.448	6.618	6.040	4.835	3.003	0.766
160	-6.106	-7.204	-7.071	-6.159	-4.756	-3.060	-1.219	0.652	2.455	4.106	5.834	7.054	7.224	7.566	7.398	6.573	5.134	3.115	0.777
175	-7.314	-7.817	-7.162	-5.800	-4.021	-2.024	0.042	2.063	3.945	5.603	7.249	8.296	8.422	8.618	8.110	7.049	5.395	3.211	0.786
190	-8.245	-8.162	-7.002	-5.216	-3.093	-0.833	1.420	3.552	5.473	7.104	8.632	9.482	9.561	9.604	8.760	7.476	5.626	3.295	0.793
205	-8.922	-8.264	-6.620	-4.438	-2.005	0.484	2.886	5.094	7.021	8.595	9.975	10.611	10.641	10.528	9.356	7.862	5.831	3.367	0.800
220	-9.366	-8.149	-6.043	-3.493	-0.781	1.902	4.419	6.669	8.572	10.064	11.274	11.683	11.663	11.395	9.902	8.211	6.014	3.431	0.806
235	-9.597	-7.837	-5.293	-2.404	0.555	3.399	6.000	8.263	10.116	11.505	12.527	12.701	12.631	12.207	10.405	8.529	6.178	3.488	0.811
250	-9.631	-7.348	-4.391	-1.192	1.984	4.959	7.614	9.863	11.644	12.913	13.733	13.667	13.546	12.968	10.870	8.819	6.327	3.539	0.815
265	-9.484	-6.699	-3.355	0.126	3.489	6.566	9.248	11.460	13.150	14.284	14.891	14.582	14.411	13.684	11.300	9.084	6.461	3.585	0.819
280	-9.173	-5.906	-2.201	1.535	5.056	8.208	10.893	13.047	14.629	15.617	16.003	15.451	15.231	14.356	11.698	9.329	6.584	3.626	0.823
295	-8.709	-4.983	-0.943	3.020	6.673	9.876	12.541	14.619	16.079	16.911	17.070	16.276	16.007	14.989	12.069	9.554	6.697	3.664	0.826
310	-8.105	-3.944	0.407	4.570	8.330	11.560	14.186	16.170	17.497	18.165	18.093	17.059	16.743	15.585	12.414	9.763	6.800	3.698	0.829
325	-7.374	-2.799	1.837	6.175	10.018	13.253	15.822	17.699	18.881	19.379	19.075	17.803	17.440	16.148	12.737	9.956	6.895	3.729	0.831
340	-6.524	-1.560	3.336	7.826	11.729	14.949	17.444	19.202	20.231	20.554	20.016	18.511	18.103	16.680	13.039	10.136	6.983	3.758	0.834
355	-5.567	-0.236	4.896	9.514	13.456	16.644	19.050	20.677	21.546	21.690	20.919	19.185	18.732	17.183	13.322	10.304	7.065	3.785	0.836
370	-4.510	1.164	6.510	11.233	15.194	18.332	20.637	22.123	22.826	22.790	21.786	19.827	19.331	17.659	13.588	10.461	7.141	3.809	0.838
385	-3.362	2.633	8.168	12.976	16.937	20.011	22.202	23.540	24.072	23.853	22.618	20.438	19.902	18.111	13.839	10.608	7.212	3.832	0.840
400	-2.131	4.163	9.866	14.739	18.683	21.678	23.744	24.927	25.284	24.881	23.417	21.022	20.445	18.540	14.075	10.746	7.278	3.854	0.842
415	-0.823	5.747	11.597	16.515	20.426	23.329	25.262	26.283	26.462	25.875	24.184	21.580	20.963	18.948	14.298	10.875	7.340	3.873	0.843
430	0.555	7.379	13.356	18.301	22.164	24.964	26.755	27.609	27.607	26.836	24.922	22.113	21.458	19.336	14.508	10.997	7.398	3.892	0.845

Figure 3. Runoff differences generated from Equation (22) for various rainfall (P) and Curve Number ($CN_{0.2}$) scenarios. Note: 1 mm = 1 million liters runoff volume in a 1 km^2 area.

Red zone cells in Figure 3 are where Equation (2) under-predicted runoff amount against Equation (21). On the other hand, the white zone cells are where Equation (2) over-predicted runoff amount. The empty cells on the upper left corner of the figure are where I_a has not been fulfilled yet to initiate any runoff amount. Collectively, Figure 3 can also be presented as a 3D model as seen in Figure 2a,b. Equations (23) and (24) represent boundary lines as indicated on the 3D model, respectively. SCS practitioners can refer to Figure 3 to perform runoff prediction correction on Equation (2).

For areas in Peninsula Malaysia with $CN_{0.2}$ value from 67 to 76 (marked by the dash line), the existing SCS model underpredicts runoff amount as indicated in red zone when rainfall depth of a storm is < 70 or 85 mm. SCS model tends to overpredict runoff amount after 85 mm and its overprediction tendency worsens toward larger storm events as indicated in white zone. Without model calibration, the SCS model worst runoff underprediction within these areas happens at $CN_{0.2}$ = 67 area at rainfall depth of 55 mm, the model underpredicted runoff amount by 2.4 million liters in 1 km^2 area while it nearly overpredicted runoff amount by 25 million liters when the storm depth reaches 430 mm in Peninsula Malaysia. Blind adoption of the existing SCS CN model is likely to over-predict runoff amount when the rainfall depth of a storm event is larger than 85 mm in Peninsula Malaysia. As such, any past study or engineering projects based upon the return period concept of rainfall amount below 70 mm might be under-designed.

3.8. Soft Computing and Data Mining of the 3D Runoff Difference Model

Even though the 3D runoff difference model can be expressed using the closed form Equation (22), it is not easy to obtain the minimum (global minima) or maximum (global maxima) runoff depth difference equations. However, with the 3D runoff difference model as a visual aid accompanied by soft computing techniques, the data mining of this vital information becomes attainable.

The minimum and maximum runoff depth prediction errors across multiple P and $CN_{0.2}$ scenarios between the two runoff models can be extracted from Figure 3. The statistically significant equations can then be determined using the SPSS to formulate the worst under and over-estimated runoff prediction error equations from Equation (2) or Equation (8) against Equation (21).

The data mining process extracts all the minimum and maximum runoff prediction differences (bold numbers, highlighted in red and yellow color, respectively in Figure 4) according to each rainfall depth scenarios (in row).

P (mm)	CN 26	30	34	38	42	46	50	54	58	62	67	72	73	76	82	86	90	94	98
10												-0.028	-0.042	-0.105	-0.359	-0.620	**-0.698**	-0.488	**0.001**
12											-0.017	-0.105	-0.134	-0.246	-0.629	**-0.808**	-0.762	-0.429	**0.068**
14										-0.005	-0.072	-0.230	-0.275	-0.444	-0.859	**-0.933**	-0.773	-0.344	**0.128**
16									-0.002	-0.039	-0.165	-0.401	-0.464	-0.696	**-1.033**	-1.007	-0.745	-0.244	**0.181**
18									-0.024	-0.104	-0.296	-0.615	-0.699	-0.952	**-1.159**	-1.039	-0.687	-0.134	**0.228**
20								-0.009	-0.072	-0.198	-0.461	-0.872	-0.963	-1.165	**-1.245**	-1.036	-0.606	-0.021	**0.269**
25							-0.027	-0.127	-0.302	-0.561	-1.025	-1.429	-1.474	**-1.533**	-1.313	-0.913	-0.340	0.264	**0.354**
40				-0.019	-0.151	-0.400	-0.766	-1.252	-1.816	-2.167	**-2.275**	-2.050	-1.968	-1.654	-0.757	-0.034	0.647	**1.008**	0.511
55		-0.003	-0.146	-0.475	-0.970	-1.623	-2.371	-2.829	**-2.982**	-2.873	-2.422	-1.685	-1.509	-0.938	0.305	1.072	**1.589**	1.558	0.597
70	-0.015	-0.267	-0.781	-1.516	-2.456	-3.252	-3.609	**-3.613**	-3.328	-2.806	-1.896	-0.788	-0.553	0.162	1.499	2.149	**2.391**	1.965	0.652
85	-0.285	-0.942	-1.894	-3.094	-3.961	**-4.284**	-4.187	-3.764	-3.087	-2.219	-0.957	0.387	0.654	1.423	2.680	**3.127**	3.060	2.275	0.689
100	-0.890	-2.011	-3.458	-4.533	**-4.910**	-4.772	-4.249	-3.440	-2.427	-1.281	0.231	1.698	1.972	2.728	3.793	**3.996**	3.619	2.517	0.716
115	-1.817	-3.457	-4.937	**-5.497**	-5.395	-4.819	-3.904	-2.757	-1.466	-0.110	1.561	3.060	3.324	4.019	**4.821**	4.763	4.090	2.711	0.737
130	-3.056	-5.048	-6.000	**-6.053**	-5.491	-4.506	-3.241	-1.805	-0.292	1.214	2.966	4.423	4.665	5.263	**5.761**	5.440	4.491	2.871	0.753
145	-4.596	-6.292	**-6.697**	-6.257	-5.260	-3.902	-2.328	-0.651	1.031	2.633	4.400	5.760	5.969	6.448	**6.618**	6.040	4.835	3.003	0.766
160	-6.106	**-7.204**	-7.071	-6.159	-4.756	-3.060	-1.219	0.652	2.455	4.106	5.834	7.054	7.224	**7.566**	7.398	6.573	5.134	3.115	0.777
175	-7.314	**-7.817**	-7.162	-5.800	-4.021	-2.024	0.042	2.063	3.945	5.603	7.249	8.296	8.422	**8.618**	8.110	7.049	5.395	3.211	0.786
190	**-8.245**	-8.162	-7.002	-5.216	-3.093	-0.833	1.420	3.552	5.473	7.104	8.632	9.482	9.561	**9.604**	8.760	7.476	5.626	3.295	0.793
205	**-8.922**	-8.264	-6.620	-4.438	-2.005	0.484	2.886	5.094	7.021	8.595	9.975	10.611	**10.641**	10.528	9.356	7.862	5.831	3.367	0.800
220	**-9.366**	-8.149	-6.043	-3.493	-0.781	1.902	4.419	6.669	8.572	10.064	11.274	**11.683**	11.663	11.395	9.902	8.211	6.014	3.431	0.806
235	**-9.597**	-7.837	-5.293	-2.404	0.555	3.399	6.000	8.263	10.116	11.505	12.527	**12.701**	12.631	12.207	10.405	8.529	6.178	3.488	0.811
250	**-9.631**	-7.348	-4.391	-1.192	1.984	4.959	7.614	9.863	11.644	12.913	**13.733**	13.667	13.546	12.968	10.870	8.819	6.327	3.539	0.815
265	**-9.484**	-6.699	-3.355	0.126	3.489	6.566	9.248	11.460	13.150	14.284	**14.891**	14.582	14.411	13.684	11.300	9.084	6.461	3.585	0.819
280	**-9.173**	-5.906	-2.201	1.535	5.056	8.208	10.893	13.047	14.629	15.617	**16.003**	15.451	15.231	14.356	11.698	9.329	6.584	3.626	0.823
295	**-8.709**	-4.983	-0.943	3.020	6.673	9.876	12.541	14.619	16.079	16.911	**17.070**	16.276	16.007	14.989	12.069	9.554	6.697	3.664	0.826
310	**-8.105**	-3.944	0.407	4.570	8.330	11.560	14.186	16.170	17.497	**18.165**	18.093	17.059	16.743	15.585	12.414	9.763	6.800	3.698	0.829
325	**-7.374**	-2.799	1.837	6.175	10.018	13.253	15.822	17.699	18.881	**19.379**	19.075	17.803	17.440	16.148	12.737	9.956	6.895	3.729	0.831
340	**-6.524**	-1.560	3.336	7.826	11.729	14.949	17.444	19.202	20.231	**20.554**	20.016	18.511	18.103	16.680	13.039	10.136	6.983	3.758	0.834
355	**-5.567**	-0.236	4.896	9.514	13.456	16.644	19.050	20.677	21.546	**21.690**	20.919	19.185	18.732	17.183	13.322	10.304	7.065	3.785	0.836
370	**-4.510**	1.164	6.510	11.233	15.194	18.332	20.637	22.123	**22.826**	22.790	21.786	19.827	19.331	17.659	13.588	10.461	7.141	3.809	0.838
385	**-3.362**	2.633	8.168	12.976	16.937	20.011	22.202	23.540	**24.072**	23.853	22.618	20.438	19.902	18.111	13.839	10.608	7.212	3.832	0.840
400	**-2.131**	4.163	9.866	14.739	18.683	21.678	23.744	24.927	**25.284**	24.881	23.417	21.022	20.445	18.540	14.075	10.746	7.278	3.854	0.842
415	**-0.823**	5.747	11.597	16.515	20.426	23.329	25.262	26.283	**26.462**	25.875	24.184	21.580	20.963	18.948	14.298	10.875	7.340	3.873	0.843
430	**0.555**	7.379	13.356	18.301	22.164	24.964	26.755	**27.609**	27.607	26.836	24.922	22.113	21.458	19.336	14.508	10.997	7.398	3.892	0.845

Figure 4. Soft computing, data mining of minimum and maximum runoff depth difference of each rainfall class (in row). Note: 1 mm = 1 million liters runoff volume in a 1 km^2 area.

Two statistically significant and best correlation equations were identified through SPSS regression modelling as:

$$\text{Min Qv} = 5.14 \times 10^{-5}\, P^2 - 0.052\, P - 0.222 \tag{25}$$

$$\text{Max Qv} = 5.14 \times 10^{-5}\, P^2 + 0.045\, P - 0.734 \tag{26}$$

where Min Q_v represents worse under-predicted runoff scenarios while Max Q_v represents the maximum over-predicted runoff scenarios. Equation (25) has an R^2-adj of 0.999, standard error of 0.037 and $p < 0.001$ while Equation (26) has an R^2-adj of 0.999, standard error of 0.191 and $p < 0.001$. Given a specific rainfall depth, the worst under-estimated and over-estimated runoff prediction errors of Equation (2) or Equation (8) due to a specific rainfall depth can be estimated by Equations (25) and (26), respectively.

It is also possible to employ soft computing technique to derive similar runoff prediction error equations in term of curve number. From Figure 5, the minimum and maximum

runoff prediction differences can be extracted as per their respective curve number (in column) which induced the runoff difference (bold numbers, highlighted in red and yellow color, respectively in Figure 5).

Two statistically significant and best correlation equations from SPSS regression modelling results are:

$$\text{Min } Q_v = 2.594 - (329.896/CN_{0.2}) \tag{27}$$

$$\text{Max } Q_v = 2.2 \times 10^{-4}\, CN_{0.2}{}^3 - 0.061\, CN_{0.2}{}^2 + 4.77\, CN_{0.2} - 86.519 \tag{28}$$

where Min Q_v, Max Q_v and $CN_{0.2}$ have been defined earlier. Equation (27) has an R^2-adj of 0.992, standard error of 0.242 and $p < 0.001$ while Equation (28) has an R^2-adj of 0.999, standard error of 0.255 and $p < 0.001$. Given a specific curve number, the worst under-estimated and over-estimated runoff prediction errors of Equation (2) or Equation (8) due to a specific $CN_{0.2}$ area can be estimated with Equations (27) and (28), respectively.

P (mm)	CN 26	30	34	38	42	46	50	54	58	62	67	72	73	76	82	86	90	94	98
10												-0.028	-0.042	-0.105	-0.359	-0.620	-0.698	**-0.488**	0.001
12											-0.017	-0.105	-0.134	-0.246	-0.629	-0.808	-0.762	-0.429	0.068
14										-0.005	-0.072	-0.230	-0.275	-0.444	-0.859	-0.933	**-0.773**	-0.344	0.128
16									-0.002	-0.039	-0.165	-0.401	-0.464	-0.696	-1.033	-1.007	-0.745	-0.244	0.181
18									-0.024	-0.104	-0.296	-0.615	-0.699	-0.952	-1.159	**-1.039**	-0.687	-0.134	0.228
20									-0.009	-0.072	-0.198	-0.461	-0.963	-1.165	-1.245	-1.036	-0.606	-0.021	0.269
25							-0.027	-0.127	-0.302	-0.561	-1.025	-1.429	-1.474	-1.533	**-1.313**	-0.913	-0.340	0.264	0.354
40				-0.019	-0.151	-0.400	-0.766	-1.252	-1.816	-2.167	-2.275	**-2.050**	**-1.968**	**-1.654**	-0.757	-0.034	0.647	1.008	0.511
55		-0.003	-0.146	-0.475	-0.970	-1.623	-2.371	-2.829	-2.982	**-2.873**	**-2.422**	-1.685	-1.509	-0.938	0.305	1.072	1.589	1.558	0.597
70	-0.015	-0.267	-0.781	-1.516	-2.456	-3.252	-3.609	-3.613	**-3.328**	-2.806	-1.896	-0.788	-0.553	0.162	1.499	2.149	2.391	1.965	0.652
85	-0.285	-0.942	-1.894	-3.094	-3.961	-4.284	-4.187	**-3.764**	-3.087	-2.219	-0.957	0.387	0.654	1.423	2.680	3.127	3.060	2.275	0.689
100	-0.890	-2.011	-3.458	-4.533	-4.910	-4.772	**-4.249**	-3.440	-2.427	-1.281	0.231	1.698	1.972	2.728	3.793	3.996	3.619	2.517	0.716
115	-1.817	-3.457	-4.937	-5.497	-5.395	**-4.819**	-3.904	-2.757	-1.466	-0.110	1.561	3.060	3.324	4.019	4.821	4.763	4.090	2.711	0.737
130	-3.056	-5.048	-6.000	-6.053	**-5.491**	-4.506	-3.241	-1.805	-0.292	1.214	2.966	4.423	4.665	5.263	5.761	5.440	4.491	2.871	0.753
145	-4.596	-6.292	-6.697	**-6.257**	-5.260	-3.902	-2.328	-0.651	1.031	2.633	4.400	5.760	5.969	6.448	6.618	6.040	4.835	3.003	0.766
160	-6.106	-7.204	-7.071	-6.159	-4.756	-3.060	-1.219	0.652	2.455	4.106	5.834	7.054	7.224	7.566	7.398	6.573	5.134	3.115	0.777
175	-7.314	-7.817	**-7.162**	-5.800	-4.021	-2.024	0.042	2.063	3.945	5.603	7.249	8.296	8.422	8.618	8.110	7.049	5.395	3.211	0.786
190	-8.245	-8.162	-7.002	-5.216	-3.093	-0.833	1.420	3.552	5.473	7.104	8.632	9.482	9.561	9.604	8.760	7.476	5.626	3.295	0.793
205	-8.922	**-8.264**	-6.620	-4.438	-2.005	0.484	2.886	5.094	7.021	8.595	9.975	10.611	10.641	10.528	9.356	7.862	5.831	3.367	0.800
220	-9.366	-8.149	-6.043	-3.493	-0.781	1.902	4.419	6.669	8.572	10.064	11.274	11.683	11.663	11.395	9.902	8.211	6.014	3.431	0.806
235	-9.597	-7.837	-5.293	-2.404	0.555	3.399	6.000	8.263	10.116	11.505	12.527	12.701	12.631	12.207	10.405	8.529	6.178	3.488	0.811
250	**-9.631**	-7.348	-4.391	-1.192	1.984	4.959	7.614	9.863	11.644	12.913	13.733	13.667	13.546	12.968	10.870	8.819	6.327	3.539	0.815
265	-9.484	-6.699	-3.355	0.126	3.489	6.566	9.248	11.460	13.150	14.284	14.891	14.582	14.411	13.684	11.300	9.084	6.461	3.585	0.819
280	-9.173	-5.906	-2.201	1.535	5.056	8.208	10.893	13.047	14.629	15.617	16.003	15.451	15.231	14.356	11.698	9.329	6.584	3.626	0.823
295	-8.709	-4.983	-0.943	3.020	6.673	9.876	12.541	14.619	16.079	16.911	17.070	16.276	16.007	14.989	12.069	9.554	6.697	3.664	0.826
310	-8.105	-3.944	0.407	4.570	8.330	11.560	14.186	16.170	17.497	18.165	18.093	17.059	16.743	15.585	12.414	9.763	6.800	3.698	0.829
325	-7.374	-2.799	1.837	6.175	10.018	13.253	15.822	17.699	18.881	19.379	19.075	17.803	17.440	16.148	12.737	9.956	6.895	3.729	0.831
340	-6.524	-1.560	3.336	7.826	11.729	14.949	17.444	19.202	20.231	20.554	20.016	18.511	18.103	16.680	13.039	10.136	6.983	3.758	0.834
355	-5.567	-0.236	4.896	9.514	13.456	16.644	19.050	20.677	21.546	21.690	20.919	19.185	18.732	17.183	13.322	10.304	7.065	3.785	0.836
370	-4.510	1.164	6.510	11.233	15.194	18.332	20.637	22.123	22.826	22.790	21.786	19.827	19.331	17.659	13.588	10.461	7.141	3.809	0.838
385	-3.362	2.633	8.168	12.976	16.937	20.011	22.202	23.540	24.072	23.853	22.618	20.438	19.902	18.111	13.839	10.608	7.212	3.832	0.840
400	-2.131	4.163	9.866	14.739	18.683	21.678	23.744	24.927	25.284	24.881	23.417	21.022	20.445	18.540	14.075	10.746	7.278	3.854	0.842
415	-0.823	5.747	11.597	16.515	20.426	23.329	25.262	26.283	26.462	25.875	24.184	21.580	20.963	18.948	14.298	10.875	7.340	3.873	0.843
430	**0.555**	**7.379**	**13.356**	**18.301**	**22.164**	**24.964**	**26.755**	**27.609**	**27.607**	**26.836**	**24.922**	**22.113**	**21.458**	**19.336**	**14.508**	**10.997**	**7.398**	**3.892**	**0.845**

Figure 5. Soft computing, data mining of minimum and maximum runoff depth difference of each $CN_{0.2}$ class (in column). Note: 1 mm = 1 million liters runoff volume in a 1 km^2 area.

The dash line on the 3D model in the valley of the red zone is described by Equations (25) and (27) while Equations (26) and (28) represent the dash line found on the ridge of the 3D runoff difference model (see Figure 2a). SCS practitioners can adopt Equations (25)–(28) to estimate the worst-case runoff prediction errors of Equation (2) when compared to the newly found λ (0.051) model in Peninsula Malaysia. On the other hand, regional or watershed specific equations can also be established by SCS practitioners for their study as proposed.

3.9. Runoff Difference Curves of the Critical Rainfall Amount

This study introduced runoff difference curves which were created with numerical analysis technique to visually present Equation (22) and to identify P_{crit}. Runoff difference curves graph combines two runoff curves (of conjugate curve numbers) into a single runoff difference curve to represent the concept of 2 previous studies [2,20,22] in another view. The graph can be plotted for specific $CN_{0.2}$ classes across multiple rainfall depth scenarios to show P_{crit} at where the curve crosses x-axis (Figure 6).
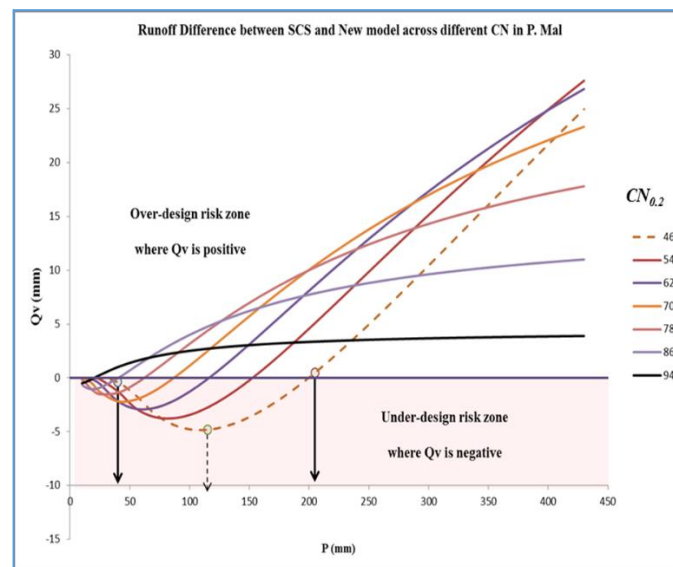
Figure 6. Runoff difference curve graph of Peninsula Malaysia. The graph was created to identify P_{crit} point(s) of different $CN_{0.2}$ classes. P_{crit} is/are the point(s) where the runoff difference curve crosses x-axis, marked by circle(s) with solid down arrow lines. The dotted down arrow line estimates the rainfall depth of maximum "under-design" risk for $CN_{0.2} = 46$. Note: When $CN_{0.2} = 46$ (dash line curve), Equation (29) solved $P_{crit} = 199.6$ mm (right bold down arrow). Equation (23) calculated the outer boundary is at $P = 25.6$ mm while the lower P_{crit} value = 45.2 mm (left bold down arrow). In conclusion, for $CN_{0.2} = 46$, Equation (2) under predicts runoff amount from any rainfall depth >25.6 mm until 199.6 mm (P_{crit}) and over predicts runoff amount for any rainfall depths >199.6 mm when compared to Equation (21).

Runoff difference curve can be used as a visual aid to identify the P_{crit} amount where the curve intersects the x-axis (when $Q_v = 0$). Possible true solution(s) as initial guess(es) of the trial and error process from the curve can be visually identified rather than guessing an arbitrary starting point for numerical solution as proposed by previous researchers [2,20]. Equation (22) is a quadratic model that yields two potential P_{crit} solutions.

Figure 6 illustrates the use of runoff difference curves to identify the "critical rainfall amount" (P_{crit}) of several $CN_{0.2}$ scenarios. For example, at $CN_{0.2} = 46$ (dash line curve), P_{crit} is approximately 40 mm and 205 mm (eyeballed from the graph, P_{crit} points are marked by solid downwards arrow where the curve intersects the x-axis, implying that Q_v is near to 0). However, the I_a amount has not been initiated for rainfall less than 40 mm according to Figure 3 and therefore, only 205 mm was used as the original trial and error estimate to satisfy Equation (22) and solve for the final solution of P_{crit} of $CN_{0.2} = 46$.

Runoff difference curve provides a brief overview and shows that Equation (2) will under-predict runoff amount at $CN_{0.2}$ area of 46 with any rainfall depths below the P_{crit} value (around 205 mm) and becomes an over-prediction thereafter. A non-linear under-design risk is therefore exhibited in the curve, with a peak of approximately 115 mm in rainfall depth (shown as dotted downwards arrow). Runoff difference curve provides additional insight of the worst under-estimated and over-estimated runoff prediction errors due to Equation (2) of specific rainfall depth which can be estimated with Equations (25) and (26), respectively.

3.10. The Critical Rainfall Amount (P_{crit}) Closed Form Equation

Through completing the square technique, this study has successfully used Equation (22) to obtain the closed form equation of P_{crit} in terms of $CN_{0.2}$. The closed form equation can be applied to solve for the P_{crit} in any pairing runoff models with any λ values. The equation can calculate the P_{crit} amount precisely and replace the trial and error

procedure mentioned in Sections [2.11] and [2.12]. SCS practitioners can refer to the proposed method in this article to derive the specific P_{crit} equation for their studies.

The derivation of the closed form equation of the critical rainfall depth (P_{crit}) from this study is shown below. From Equation (22),

$$Q_v = \frac{\left[P - 50.8\left(\frac{100}{CN_{0.2}} - 1\right)\right]^2}{\left[P + 203.2\left(\frac{100}{CN_{0.2}} - 1\right)\right]} - \frac{\left[P - 21.606\left(\frac{100}{CN_{0.2}} - 1\right)^{1.063}\right]^2}{\left[P + 402.547\left(\frac{100}{CN_{0.2}} - 1\right)^{1.063}\right]}$$

$$\text{Let}: A = 21.606\left(\frac{100}{CN_{0.2}} - 1\right)^{1.063}$$

$$\text{Let}: B = 50.8\left(\frac{100}{CN_{0.2}} - 1\right)$$

When $Q_v = 0$ (Runoff indifferent between 2 models), substitute A and B and solve for P (P_{crit}).

$$\frac{[P - B]^2}{[P + 4B]} = \frac{[P - A]^2}{[P + 18.631A]}$$

After grouping and simplifying, P (**P_{crit}**) can be solved via quadratic form as below:

$$a = 4B - 2A + 2B - 18.631A$$

$$b = A^2 - 8AB - B^2 + 2(18.631)AB$$

$$c = 4BA^2 - 18.631AB^2$$

$$P_{crit} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{29}$$

P_{crit} = Critical rainfall depth (mm)
$CN_{0.2}$ = Conventional curve number of a watershed

Equation (29) is a quadratic model that yields two potential P_{crit} solutions. The outer boundary (Equation (23)) can be used as checkpoint to determine if the lower P_{crit} value is a valid solution because any rainfall depths beyond the outer boundary will start to yield runoff difference between the two models after fulfilling the I_a requirement. The lower P_{crit} value is usually discarded due to its proximity to (or less than) the outer boundary.

If the P_{crit} value < the P value of Equation (23) (outer boundary equation), the I_a is yet to be fulfilled thus it is impossible to have any runoff or runoff difference amount. Runoff difference curves graph is also an effective visual aid to supplement the P_{crit} closed-form equation (refer to Figure [6] example).

Results from several derived formulae were compiled in Table [3] to provide another quick overview of the P_{crit} for Peninsula Malaysia across multiple $CN_{0.2}$ scenarios. According to the DID HP 27 dataset, the lowest calculated $CN_{0.2}$ is 48.8; hence, column A tabulates $CN_{0.2}$ range from 47 to 99 to cover the entire possible $CN_{0.2}$ scenario in Peninsula Malaysia. Column B and D were calculated using Equation (6), column C used Equation (20) and column E used Equation (29). Column F calculated $CN_{0.2}$ percentage change into $CN_{0.051}$.

Column A and E can be used to construct another P_{crit} overview curve across multiple $CN_{0.2}$ scenarios (Figure [7]) with a statistically significant equation regressed via SPSS as:

$$P_{crit} = -245.4 \ln(CN_{0.2}) + 1132.6 \tag{30}$$

Equation (30) has an R^2-adj of 0.997, standard error of 3.047 and $p < 0.001$. Given $CN_{0.2}$ value of a watershed, the corresponding P_{crit} value can be estimated with Equation (30). Equation (2) under predicts runoff amount at any rainfall depths below the P_{crit} overview curve in Figure [7] and vice versa. Figure [7] is also in line with the research outcome reported by [2] that Equation (2) had the tendency to under-estimate runoff amount in rural and forested watersheds as $CN_{0.2}$ decreases.

Table 3. Conjugate $CN_{0.051}$ and P_{crit} for Peninsula Malaysia.

(A)	(B)	(C)	(D)	(E)	(F)
$CN_{0.2}$	$S_{0.2}$	CN_λ (0.051)	$S_{0.051}$	P_{crit} (mm)	%
99	2.57	98.76	3.20	7.38	0.2%
97	7.86	96.02	10.52	12.65	1.0%
95	13.37	93.20	18.52	17.83	1.9%
93	19.12	90.36	27.10	22.86	2.8%
91	25.12	87.52	36.23	27.85	3.8%
89	31.39	84.69	45.92	32.85	4.8%
87	37.95	81.89	56.18	37.91	5.9%
85	44.82	79.11	67.05	43.05	6.9%
83	52.02	76.38	78.57	48.31	8.0%
81	59.58	73.68	90.75	53.71	9.0%
79	67.52	71.02	103.67	59.26	10.1%
77	75.87	68.40	117.35	65.00	11.2%
75	84.67	65.83	131.88	70.94	12.2%
73	93.95	63.30	147.29	77.11	13.3%
71	103.75	60.81	163.68	83.53	14.4%
69	114.12	58.37	181.14	90.22	15.4%
67	125.10	55.98	199.74	97.22	16.4%
65	136.77	53.63	219.60	104.56	17.5%
63	149.18	51.33	240.84	112.26	18.5%
61	162.39	49.07	263.60	120.38	19.6%
59	176.51	46.86	288.03	128.94	20.6%
57	191.61	44.69	314.31	138.00	21.6%
55	207.82	42.57	342.65	147.61	22.6%
53	225.25	40.49	373.28	157.83	23.6%
51	244.04	38.46	406.49	168.74	24.6%
49	264.37	36.46	442.59	180.41	25.6%
47	286.43	34.51	481.96	192.95	26.6%
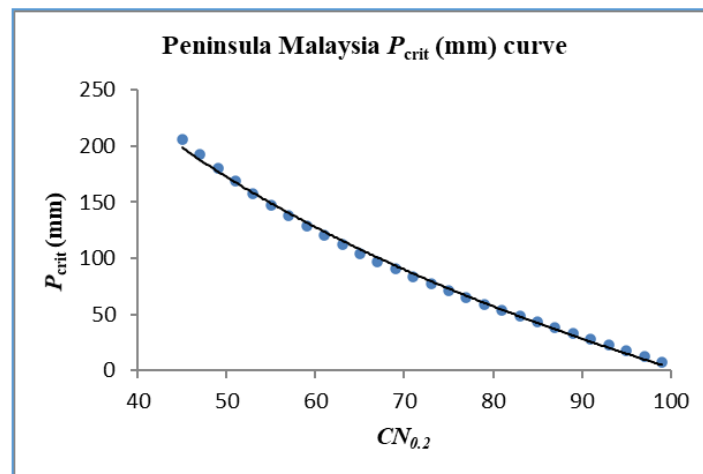
Figure 7. P_{crit} overview curve for Peninsula Malaysia. Equation (2) under predicts runoff amount for any rainfall depths below the curve at respective CN0.2 area. The underprediction tendency worsens as $CN_{0.2}$ value decreases.

Using the same concept as presented in **??** and Section 3.10, the closed form P_{crit} can also be derived to verify previous study results where the optimum λ value was identified as 0.05 in the USA. The correlation between S_λ and $S_{0.2}$ is best represented by $S_{0.05} = 1.33 S_{0.2}^{1.15}$ [2,20,22]. It is noteworthy to mention that US researchers used inches in their dataset; hence, Equation (18) (CN formula, SI version) needs to be converted and $CN_\lambda = \frac{1000}{S_\lambda + 10}$ should be used instead. The closed form P_{crit} equation can be derived with

the same method as proposed in Section 3.10 to verify their published P_{crit} (inches) values (Table 4) in USA [2,22].

Table 4. The P_{crit} (inches) values with its corresponding $CN_{0.2}$ and $CN_{0.05}$ values for runoff prediction studies in USA (Modified from [2,22]).

Conjugate Curve Numbers and P_{crit} Values				
$CN_{0.2}$	$S_{0.2}$ (in)	$CN_{0.05}$	$S_{0.05}$ (in)	P_{crit} (in)
100	0	100	0	-
95	0.526	94.02	0.636	2.44
90	1.111	86.95	1.501	1.72
85	1.765	79.64	2.556	1.95
80	2.5	72.39	3.815	2.27
75	3.333	65.31	5.311	2.63
70	4.286	58.51	7.091	3.05
65*	5.385	52.03	9.219	**3.52 (4.51)***
60	6.667	45.9	11.785	4.04
55	8.182	40.14	14.915	4.64
50**	10	34.74	18.787	**5.33 (5.35)****
45	12.222	29.71	23.663	6.15
40	15	25.03	29.947	7.13
35	18.571	20.71	38.285	8.35

Note: (4.51)* old value for $CN_{0.2}$ = 65. (5.35)** old value for $CN_{0.2}$ = 50.

The closed form P_{crit} equation verified all P_{crit} values in Table 4 except for $CN_{0.2}$ = 50** and 65*. For $CN_{0.2}$ = 50**, the calculated P_{crit} using the closed form equation method is 5.33 inches (instead of 5.35 inches)**. The variance to the published value is about 0.5 mm. However, for $CN_{0.2}$ = 65*, the calculated P_{crit} is 3.52 inches (instead of 4.51 inches)*, which is much lower than the published value by about 25 mm.

Verification of Table 4 Pcrit values prove that the P_{crit} closed form equation can be used to calculate the exact P_{crit} value for any comparing SCS CN models for SCS practitioners. The success in the closed form equation derivation narrows the study gap from previous work. It can be adopted to replace the trial and error technique used by previous researchers [2,20,22].

3.11. Critical Curve Number (CN_{crit})

Equation (29) will yield two possible CN_{crit} solutions (when Q_v = 0 in Equation (22)). Although it is possible for those CN_{crit} values to exist, all values must be verified. Potential CN_{crit} solution(s) as the initial guess(es) to the trial and error process to satisfy Equation (22) can be identified when visually aided by runoff difference curves.

For an example, when rainfall = 100 mm (dash line curve in Figure 8), potential CN_{crit} value is about 66 (marked by bold solid down arrows where the curve intersects with the x-axis or Q_v = 0). Other possible CN_{crit} value were discarded because the dash line curve intersects the x-axis at the left end at $CN_{0.2}$ around 22 and 99 on the right end, those values remain as a theoretical $CN_{0.2}$ value only.

3.12. Asymptotic Curve Number of Peninsula Malaysia

According to the AFM (Section 2.10), the DID HP 27 dataset resembles the standard behavior pattern (Figure 9) and thus Equation (16) was adopted to derive CN_∞ as the best representative $CN_{0.2}$ value for the dataset. Through least square fitting method under AFM, the fitting parameter k was identified to be 40.79 and CN_∞ = 67.77. When rounded to the closest positive integer, CN_∞ = $CN_{0.2}$ = 68.
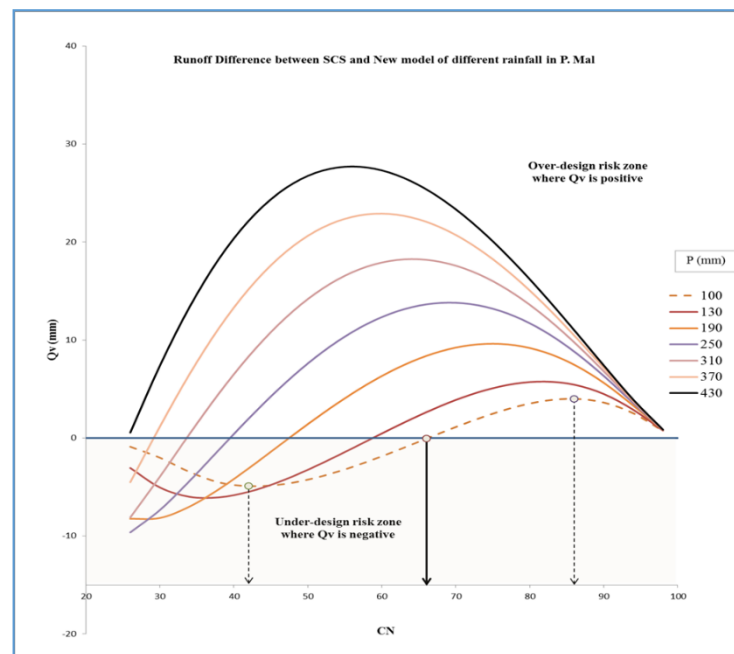
Figure 8. Runoff difference curves between Equation (2) or Equation (8) and (21). CN_{crit} is the point that the runoff difference curve intersects the x-axis, marked by circle with solid down arrows lines. The dotted down arrow lines estimate the rainfall depth of maximum "under and over-design" risk for P = 100 mm, respectively. Note: when rainfall = 100 mm (dash line curve), runoff difference curve also suggests that the return period design base on rainfall depth of 100 mm is likely to cause under-design risk (negative Q_v) in watersheds where $CN_{0.2}$ value(s) is (are) <66, meanwhile incurring over-design risk (positive Q_v) in $CN_{0.2}$ values >66. Estimated worst under-design risk (marked with dotted down arrows) occurs around $CN_{0.2}$ = 42 while the worst over-design risk at about 86. The worst under and over-estimated runoff prediction errors due to Equation (2) of those $CN_{0.2}$ area can be estimated with Equations (27) and (28), respectively.
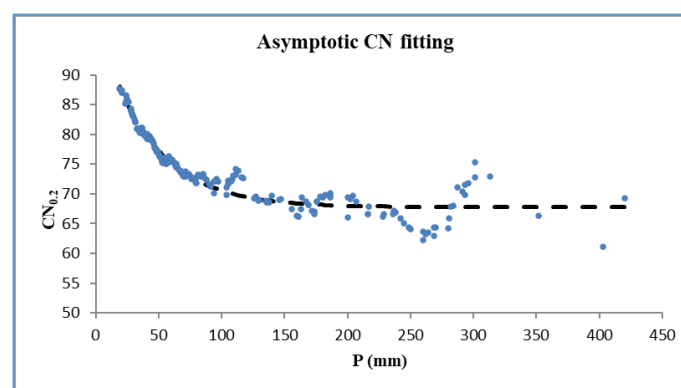
Figure 9. Asymptotic CN fitting of the dataset. For standard behavior pattern, CN_{∞} is the point where a near to stable state of $CN_{0.2}$ fits to the higher rainfall depths.

The AFM CN∞ result is in proximity to the equivalent $CN_{0.2}$ value of 72.58 which was derived in Section 3.2, whereas CN_{∞} = 68 also falls within the 99% $CN_{0.2}$ confidence interval of this study. This proves that the proposed SCS CN model calibration methodology in this article is capable to produce results that are in line with other method introduced by previous study.

Using Equation (18), the calculated $S_{0.2}$ value of the AFM CN is 120.78 mm and I_a = 0.20 × 120.78 mm = 24.16 mm. These numbers are used in formulating the SCS runoff model with Equation (1) for benchmarking (Table 5).

Table 5. Asymptotic CN fitting method (AFM) and new λ runoff model's residual analyses comparison with descriptive and inferential statistics at alpha = 0.01 level.

	AFM Model	New λ Model
λ value	0.20	0.051
E	0.910	0.919
RSS	69,933	62,926
Residual Standard Deviation	17.083	16.556
Residual Standard Deviation: BCa 99% CI	[14.200, 19.552]	[13.875, 18.898]
Residual Skewness	0.401	−0.098
Mean Residual:	−4.188	−2.079
Mean Residual: BCa 99% CI	[−6.953, −1.035]	[−4.814, 0.920]
Residual: Range	96.89	101.45
Residual Variance	291.822	274.091
Residual Variance: BCa 99% CI	[201.207, 382.593]	[192.434, 358.014]

The newly calibrated λ model has lower *RSS* with higher *E* index compared to the runoff model formulated with the Asymptotic CN value. The models' residual skewness is near to zero, thus the mean residual value can act as an indicator for the predictive model's accuracy. The new λ model has lower mean residual with 99% confidence interval range which spans across zero, indicating its capability to achieve zero (residual) runoff prediction error. On the other hand, the AFM model tends to under-predict runoff volumes since their mean residual confidence interval range is within negative value range. The descriptive statistics indicates that the AFM model has a lower residual range. However, the standard deviation and variance in the model's residual are lower in the new λ model with smaller confidence interval ranges. Hence, the new λ model has higher stability and reliability for the dataset of this study.

AFM model faced another issue, whereby the calculated I_a value (24.16 mm) is larger than nearly 3.10% (seven recorded rainfall events) of the DID HP 27 dataset. According to the runoff constraint defined by SCS (as stated in Section 2.1) any rainfall depths < I_a value would not initiate any runoff; hence, AFM model failed to comply with the SCS constraint for those seven P-Q data pairs. On the other hand, New λ model does not have this issue.

4. Conclusions

This article presented the methodology to perform the SCS CN model calibration under the guide of inferential statistics with regional rainfall-runoff data. The study honed the runoff prediction accuracy of a popular rainfall-runoff model and based on its mathematical framework to develop engineering applications. Key highlights are as below:

1. The methodology to reassess the validity of a popular runoff model was presented. Under this study, the existing SCS runoff model is invalid for runoff modelling (alpha = 0.01), and therefore the model must be calibrated. λ = 0.051 (99% CI ranges from 0.034, 0.051) and $CN_{0.2}$ = 72.58 (99% CI ranges from 67 to 76) are the calibrated results for runoff prediction in Peninsula Malaysia according to the dataset of this study. Within these $CN_{0.2}$ areas, SCS model underpredicts runoff amount when rainfall depth of a storm is <70 to 85 mm and its overprediction tendency worsens toward larger storm events if it is not calibrated. The SCS CN model underpredicted runoff amount the most (2.4 million L/km^2 area) at $CN_{0.2}$ = 67 area and rainfall depth of 55 mm while it nearly overpredicted runoff amount by 25 million L/km^2 area when the storm depth reaches 430 mm in Peninsula Malaysia.

2. The closed form equation of the "Critical Rainfall Amount (P_{crit})" was solved (Sections 2.12 and 3.10) to narrow the research gap. Figure 6 example illustrated its use and past publication errors were detected (Table 4). The "Critical Curve Number (CN_{crit})" concept and the use of the runoff difference curves graph were also introduced in this article (Sections 2.11–2.13 and Sections 3.9–3.11) with demonstrated applications shown in Figures 6–8.

3. The 3D runoff difference model (Figure 2a,b) was created with Equation (22) to assess the runoff prediction results of the existing SCS CN model and its type II errors. Equations (25)–(28) to estimate the worst-case runoff prediction errors of the SCS CN model when it is not calibrated with $\lambda = 0.051$ for runoff predictions in Peninsula Malaysia. Any past study or engineering projects using this model and based upon the return period concept of rainfall amount below 70 mm might be under-designed while the model has over-design risk when a storm depth is larger than 85 mm. SCS practitioners are encouraged to refer to the general formulae (Equations (10)–(12)) and proposed methods in this article to derive the specific model and equations for their studies. Equation (2) must be validated with rainfall-runoff dataset prior to its adoption for runoff prediction in any part of the world.

4. Authors cautioned that there are several limitations of the proposed methodology. Minimum sample size should be at least 100 observations while the alpha level setting for Null assessment is pending upon research need. BCa should be used instead of bootstrapping and the choice of the statistical software must come with the option to provide confidence interval for median value to cater for model calibration need when the dataset is skewed. Runoff error analyses beyond the confidence interval or dataset limit may not be meaningful for interpretations.

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.; Woodward, D.E.; Van Mullem, J. *Curve Number Hydrology: State of the Practice*; ASCE: Reston, VA, USA, 2009; p. 106.
3. Hawkins, R.H. Curve Number Method: Time to Think Anew? *J. Hydrol. Eng.* **2014**, *19*, 1059. [CrossRef]
4. 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* **2019**, *12*, 60. [CrossRef]
5. National Engineering Handbook; Part 630 Hydrology, Chapter 10, Figure 10.1: USDA, NRCS. 2004. Available online: https://directives.sc.egov.usda.gov/OpenNonWebContent.aspx?content=17752.wba (accessed on 30 March 2021).
6. Ross, C.W.; Prihodko, L.; Anchang, J.; Kumar, S.; Ji, W.J.; Hanan, N.P. HYSOGs250m, global gridded hydrologic soil groups for curve-number-based runoff modeling. *Sci. Data* **2018**, *5*, 180091. [CrossRef] [PubMed]
7. Jaafar, H.H.; Ahmad, F.A.; El Beyrouthy, N. GCN250, new global gridded curve numbers for hydrologic modeling and design. *Sci. Data* **2019**, *6*, 1–9. [CrossRef]
8. Chen, Y.; Wang, Y.; Zhang, Y.; Luan, Q.; Chen, X. Flash floods, land-use change, and risk dynamics in mountainous tourist areas: A case study of the Yesanpo Scenic Area, Beijing, China. *Int. J. Disaster Risk Reduct.* **2020**, *50*, 101873. [CrossRef]
9. Park, K.; Won, J.-H. Analysis on distribution characteristics of building use with risk zone classification based on urban flood risk assessment. *Int. J. Disaster Risk Reduct.* **2019**, *38*, 101192. [CrossRef]
10. Sun, R.; Gong, Z.; Gao, G.; Shah, A.A. Comparative analysis of Multi-Criteria Decision-Making methods for flood disaster risk in the Yangtze River Delta. *Int. J. Disaster Risk Reduct.* **2020**, *51*, 101768. [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* **2020**, *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. 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]

14. 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]

15. Lal, M.; Mishra, S.K.; Pandey, A.; Pandey, R.P.; Meena, P.K.; Chaudhary, A.; Jha, R.K.; Shreevastava, A.K.; Kumar, Y. Evaluation of the Soil Conservation Service curve number methodology using data from agricultural plots. *Hydrogeol. J.* **2017**, *25*, 151–167. [CrossRef]

16. Fidal, J.; Kjeldsen, T. Accounting for soil moisture in rainfall-runoff modelling of urban areas. *J. Hydrol.* **2020**, *589*, 125122. [CrossRef]

17. 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.* **2021**, *35*. [CrossRef]

18. 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. Process.* **2017**, *31*, 2226–2243. [CrossRef]

19. 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.* **2017**, *77*, 662–669. [CrossRef]

20. Jiang, R. *Investigation of Runoff Curve Number, Initial Abstraction Ratio*; University of Arizona: Tucson, AZ, USA, 2001.

21. DID, Hydrological Procedure No. 27. Design Flood Hydrograph Estimation for Rural Catchments in Peninsula Malaysia. JPS, DID, Kuala Lum-Pur. 2010. Available online: https://www.water.gov.my/jps/resources/PDF/Hydrology%20Publication/Hydrological_Procedure_No_27_(HP_27).pdf (accessed on 30 March 2021).

22. Woodward, D.E.; Hawkins, R.H.; Jiang, R.; Hjelmfelt, J.A.T.; Van Mullem, J.A.; Quan, Q.D. Runoff Curve Number Method: Examination of the Initial Abstraction Ratio. In Proceedings of the World Water & Environmental Resources Congress 2003; American Society of Civil Engineers (ASCE), Philadelphia, PA, USA, 23–26 June 2003; pp. 1–10.

23. 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.

24. 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]

25. 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]

26. Helsel, D.R.; Hirsch, R.M.; Ryberg, K.R.; Archfield, S.A.; Gilroy, E.J. Statistical methods in water resources. In *Techniques and Methods*; US Geological Survey: Reston, VA, USA, 2020; p. 458.

27. IBM. *IBM, SPSS Bootstrapping 21 Guide*; IBM Press: Indianapolis, IN, USA, 2012.

28. Efron, B. *Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction*; Cambridge University Press: London, UK, 2010.

29. Puth, M.-T.; Neuhäuser, M.; Ruxton, G.D. On the variety of methods for calculating confidence intervals by bootstrapping. *J. Anim. Ecol.* **2015**, *84*, 892–897. [CrossRef]

30. Schneider, L.; McCuen, R.H. Statistical Guidelines for Curve Number Generation. *J. Irrig. Drain. Eng. ASCE* **2005**, *131*, 282–290. [CrossRef]

31. Hjelmfelt, A.T. Curve-number procedure as infiltration method. *J. Irrig. Drain. Eng. ASCE* **1980**, *106*, 1107–1111.

32. Hjelmfelt, A.T. Empirical Investigation of Curve Number Technique. *J. Irrig. Drain. Eng. ASCE* **1980**, *106*, 1471–1476.

33. Hawkins, R.H. Asymptotic determination of runoff curve numbers from data. *J. Irrig. Drain. Eng. ASCE* **1993**, *119*, 334–345. [CrossRef]

34. González, Á.; Temimi, M.; Khanbilvardi, R. Adjustment to the curve number (NRCS-CN) to account for the vegetation effect on hydrological processes. *Hydrol. Sci. J.* **2015**, *60*, 591–605. [CrossRef]

35. Kowalik, T.; Walega, A. Estimation of CN Parameter for Small Agricultural Watersheds Using Asymptotic Functions. *Water* **2015**, *7*, 939–955. [CrossRef]