Supplementary Materials

A Novel GIS-SWMM-ABM Approach for Flood Risk Assessment in Data-Scarce Urban Drainage Systems

Shakeel Ahmad [1], Haifeng Jia [1,*], Anam Ashraf [1,2], Dingkun Yin [1], Zhengxia Chen [1], Rasheed Ahmed [3] and Muhammad Israr [4]

[1] School of Environment, Tsinghua University, Beijing 100084, China; ahmad18@mails.tsinghua.edu.cn (S.A.); anamashraf@xtbg.ac.cn (A.A.); ydk19@mails.tsinghua.edu.cn (D.Y.); chenzhengxia@tsinghua.edu.cn (Z.C.)
[2] Center for Integrative Conservation, Landscape Ecology Laboratory, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Mengla 666303, China
[3] Department of Soil and Climate Sciences, The University of Haripur, Haripur 22620, Pakistan; rasheedahmad@uoh.edu.pk
[4] Institute of Development Studies, The University of Agriculture Peshawar, Peshawar 25130, Pakistan; misrar@aup.edu.pk
* Correspondence: jhf@tsinghua.edu.cn

Table S1. SWMM model input parameters.

S. No.	SWMM Model Input Parameter	Description
1	Rainfall	This represents the rainfall hyetograph (intensity–duration–frequency curve) used in the SWMM simulation. It is used as an input as a user-defined hyetograph or selected from a library of predefined options.
2	Dstore-imperv (mm)	This represents the initial depression storage depth on impervious surfaces within the sub-catchment. It defines how much rainfall is stored on impervious surfaces before it begins to run off.
3	Area of sub-catchments (ha)	This specifies the total area of the sub-catchment in hectares. It is used to calculate the runoff volume and flow rates.
4	Dstore-perv (mm)	This represents the initial depression storage depth on pervious surfaces within the sub-catchment. Like Dstore-imperv, it defines the initial storage capacity of pervious surfaces before infiltration occurs.
5	Width	This defines the representative width of the sub-catchment, used in runoff calculations.
6	% Zero-imperv	This represents the percentage of the sub-catchment area considered "fully impervious," meaning no infiltration occurs in these areas.
7	Slope (%)	This specifies the average slope of the sub-catchment, which influences the runoff velocity and flow routing.
8	Depression storage	This defines the total depression storage depth across the sub-catchment area, including impervious and pervious surfaces. It represents the initial storage capacity before any runoff or infiltration occurs.

9	% Impervious	This represents the total percentage of the sub-catchment area considered impervious.
10	Invert elevation (nodes)	This specifies each node's invert elevation (bottom elevation). Nodes represent points in the drainage system where water can collect or flow through.
11	N-Imperv	This represents Manning's roughness coefficient for flow on impervious surfaces within the sub-catchment. This coefficient affects how quickly water flows over these surfaces.
12	Conduits (length, type)	This defines the characteristics of each conduit (pipe or channel), including its length and type (e.g., circular, rectangular).
13	N-perv	This represents Manning's roughness coefficient for flow on pervious surfaces within the sub-catchment. Similar to N-imperv, it affects the flow velocity on these surfaces.
14	Infiltration model (modified Green–Ampt)	This specifies that the modified Green–Ampt infiltration model will simulate infiltration in the sub-catchment. This model requires additional parameters listed below.
	a. Suction head (mm)	The depth of water below the soil surface at which infiltration stops.
	b. Conductivity (mm/hr)	The rate at which water can infiltrate the soil when the soil moisture deficit is zero.
	c. Initial soil moisture deficit (fraction)	The fraction of the soil potential storage capacity initially filled with water.

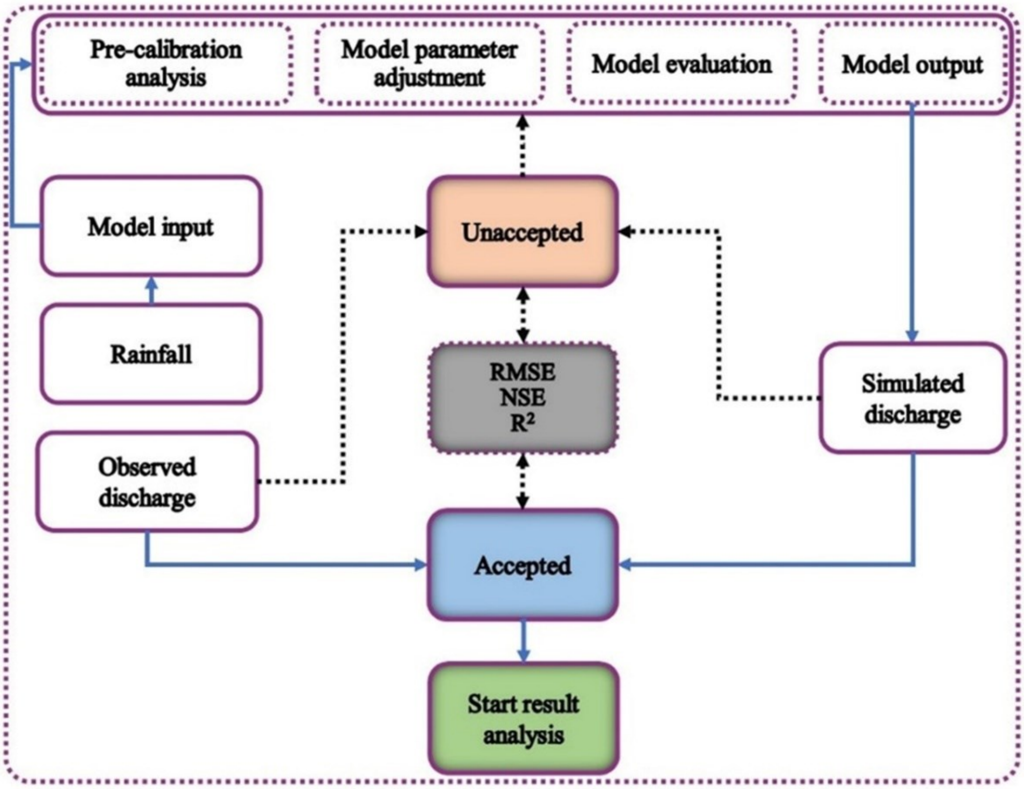

Figure S1. SWMM model calibration and validation.

Table S2. Selected events for model calibration and validation.

Date	Event No.	Event ID	Duration	Rainfall depth (mm)	
July 03, 2018	1	20180703	Minutes	32	Model calibration
August 20, 2020	2	20200820	Minutes	56	
August 13, 2020	3	20200813	Minutes	45	Model validation
July 21, 2022	4	20220721	Minutes	74	

Model Validation

The validation of the model is a crucial part of the model calibration process. It assesses the model's performance by comparing the simulated outputs to the observed data using goodness-of-fit criteria. Validation typically involves both a visual comparison of the plotted simulated and measured variables and the calculation of statistical metrics. Commonly used statistical measures to quantify the agreement between observed and simulated data include the RMSE (equation S1), NSE (equation S2), and coefficient of determination (R^2) (equation S3).

$$RMSE = \sum \sqrt{\frac{(O_i - S_i)^2}{\mathrm{n}}} \qquad \text{S1}$$

Here, O_i indicates observed values, Si indicates simulated values, and n is the number of data points.

$$NSE = 1 - \frac{\sum_{i=1}^{n}(O_i - S_i)^2}{\sum_{i=1}^{n}(O_i - O_{i\,(avg)})^2} \qquad \text{S2}$$

Here, $O_{i\,(avg)}$ is the average observed value.

$$R^2 = 1 - \frac{\sum(O_i - S_i)^2}{\sum(S_i - O_{i\,(avg)})^2} \quad \text{S3}$$

Log-Pearson Type III Distribution

The data were initially log-transformed. The mean, log of the mean (log x), standard deviation, and Skew coefficient (Gs) of the log-transformed maximum values (log xi) were calculated for each year and used to determine a statistic. The arithmetic mean were determined by Equation (S4), defined as the sum of the values ($\sum R$) divided by the number of observations (n):

$$\overline{R} = \frac{\sum R}{n} \quad \text{S4}$$

Variance (V) was determined using Equation (S5), computed as the average squared deviation of each number from its mean.

$$V = \frac{\sum[(\log(R) - avg\ (\log(R))]^2}{n-1} \quad \text{S5}$$

Here, R indicates the amount of rainfall.

Standard deviations were determined using Equation (S.6), which represents the SD, calculated as the square root of variance as present in Equation (S5).

$$SD = \sqrt{\frac{\sum[(\log(R) - avg\ (\log(R))]^2}{n-1}} \quad \text{S.6}$$

Finally, the Skew coefficient was determined using Equation (S.7), showing the Skew coefficient (G_s).

$$G_s = \frac{n * \sum[(\log(R) - avg\ (\log(R))]^3}{(n-1)(n-2)\ SD\ [(\log(R))]^3} \quad \text{S.7}$$

First, we created a spreadsheet to solve for Gs and arrange the rainfall data in descending order. At this point, we used Excel functions to calculate standard deviations, variances, and skew coefficients. We created a second column containing the log of each rainfall to calculate the average rainfall and average log for the second column (Avg(log)). The third column consisted of the log R – avg (log R)^2 for each row, then we calculated the sum of the third columns. The fourth column consisted of [log R – avg (log R)]^3. We then calculated the sum for the third column, as displayed in Table S3.

In the LPT-III method, the frequency of rainfall depth R_d (mm) for any rainfall duration with a specified return period (TP) is computed using Equation (S.7) [75].

$$Log10\ (R) = mean\left(log10(R)\right) + K(RP, Gs) * SD(log10(R)) \quad \text{S.8}$$

We used the frequency factor (K) table [76] and skew coefficient to find the value of K for 2, 5, 10, and 25-year return periods, which were obtained from [46]. By knowing the skewness coefficient and the recurrence interval, the K values for the LPT-III distribution were extracted. The antilog of Equation (S.7) provides the estimated extreme value for the given return period. Table S4 shows the computed rainfall depth values for different return periods using LPT-III distribution.

Table S3. Annual maximum rainfall (*R*) and calculated statistical variables for the study area (duration = 24 hr).

Year	R (mm)	R*= LogR	$(R^*-R^*_{avg})^2$	$(R^*-R^*_{avg})^3$
2001	187.96	2.27	0	0
2002	144.27	2.16	0.03	-0.01
2003	150.11	2.18	0.02	0
2004	363.73	2.56	0.05	0.01
2005	377.70	2.58	0.06	0.01
2006	262.64	2.42	0.01	0
2007	149.86	2.18	0.02	0
2008	286.77	2.46	0.02	0
2009	181.86	2.26	0.01	0
2010	224.03	2.35	0.00	0
2011	152.91	2.18	0.02	0
2012	100.84	2.00	0.11	-0.04
2013	149.10	2.17	0.03	0
2014	363.22	2.56	0.05	0.01
2015	204.22	2.31	0	0
2016	196.60	2.29	0	0
2017	173.99	2.24	0.01	0
2018	525.78	2.72	0.15	0.06
2019	246.63	2.39	0	0
2020	338.58	2.53	0.04	0.01
2021	163.83	2.21	0.01	0
2022	188.21	2.27	0	0
	A-mean	**R^*_{avg}**	**Sum**	**Sum**
	10.61	2.33	0.65	0.05
	V	0.03		
	SD	0.18		
	G_s	0.47		

Note: R (mm) is the annual maximum rainfall, R* is the logarithm of rainfall, A-mean is the average mean of R (mm), R^*_{avg} is the average of the R*values, V is the variance, SD is the standard deviation, and G_s is the Skew coefficient for the LPT-III method.

Table S4. Rainfall depths for different return periods based on log Pearson type III.

Return period (RP) (years)	Frequency factor (K)	Mean Log10 (R)	Rainfall depth (R_d) (mm)
2	-0.08	2.32	208.18
5	0.81	2.47	298.39
10	1.32	2.56	367
25	1.9	2.67	464.49