A Bayesian updating framework for calibrating the hydrological parameters of road networks using taxi GPS data

Xiangfu Kong[1,2], Jiawen Yang[2], Ke Xu[3], Bo Dong[1], Shan Jiang[4,5]

[1] Research Center for AI Social Governance, Zhejiang Lab, Hangzhou, Zhejiang 311100, China
[2] Shenzhen Graduate School, Peking University, Shenzhen, Guangdong 518055, China
[3] Zhejiang Development and Planning Institute, Hangzhou, Zhejiang 310012, China
[4] Department of Urban and Environmental Policy and Planning, Tufts University, Medford, MA 02155, USA
[5] Department of Civil and Environmental Engineering, Tufts University, Medford, MA 02155, USA

Correspondence to: Xiangfu Kong (kongxf@zhejianglab.com)

Abstract. Hydrological parameters should pass through a careful calibration procedure before aiding decision-~~-~~making. However, ~~great~~ significant ~~difficulties~~ difficulty is ~~are~~ encountered when applying existing calibration methods to regions in which~~where~~ runoff data are inadequate. To achieve accurate ~~fill the gap of~~ hydrological calibration for ~~the~~ ungauged road networks, we propose~~d~~ a Bayesian updating framework that~~to~~ calibrates hydrological parameters based on taxi GPS data. Hydrological parameters were~~are~~ calibrated by adjusting their values such that the runoff generated by ~~the~~ acceptable parameter sets corresponds to ~~could yield~~ the road disruption periods during which no taxi points are observed. The proposed method was~~is~~ validated on ~~through~~ 10 flood-prone roads in Shenzhen~~,~~ and the results revealed~~s~~ that the trends of runoff could be correctly predicted for 8 ~~out~~ of 10 roads. This study ~~shows~~ demonstrates that the integration of hydrological models and taxi GPS data can provide ~~suggests~~ viable alternative measures for ~~the~~ model calibration to derive~~, and provides~~ actionable insights for flood hazard mitigation.

1 Introduction

In the context ~~Under the background~~ of climate change and increased urbanization, flooding poses ~~is posing~~ far-reaching threats to ~~the~~ urban road networks of coastal metropolises ~~metropolis~~ (Balistrocchi et al., 2020). In Australia, ~~around~~ approximately 53% of flood-related drowning deaths were the result of vehicles driving into flood~~ed~~ waters. Additionally, i~~I~~ndirect losses caused by flooding~~,~~ such as cancelled commutes, mandatory detours, and travel time delays~~,~~ ~~even~~ often outweigh direct losses (Kasmalkar et al., 2020). Quantifying the impact of flood~~ing~~ exposure requires the prediction of surface runoff over ~~the~~ roads and ~~computation of~~ road disruptions induced by ~~the~~ runoff, which are critical for the implementation of ~~to~~ flood~~ing~~ mitigation, traffic resilience improvement, and ~~risk~~ early warning systems.

Public concerns ~~about~~ regarding road flooding hazards have created ~~the pressing need~~pressure to develop fine-grained and accurate models for hydrological simulation. H~~The h~~ydrological model~~l~~ing is ~~, as~~ a ~~quite~~ relatively well-established theory that can~~,~~ provide~~s~~ ~~an~~ approximations of ~~the~~ real-world hydrological system~~,~~ and has been widely used in many road-related studies (Versini et al., 2010; Yin et al., 2016; Safaei-Moghadam et al., 2022). ~~As the~~ Because hydrological model~~l~~ing is subject to uncertainty~~, which~~ that arises from the over-simplified~~false~~ reflection of ~~the~~ hydrological systems, ~~the~~ initial and boundary conditions, and ~~the~~ lack of true knowledge, parameters for~~of~~ hydrological models ~~should~~ must be carefully calibrated ~~be carefully calibrated~~ prior to their~~before~~ ~~applying~~ application to ~~solve~~ practical problems, so that ~~the~~ models can ~~is capable of~~ closely match~~ing~~ the historical ~~evidences~~ trends (Gupta et al., 1998). As~~n~~ un-calibrated models are~~is~~ indefensible and sterile, ~~so~~ very few models documented in the literature have been applied without ~~any calibrations~~a calibration procedure (Beven, 2012).

~~During~~ Over the ~~last~~ past four decades, numerous studies have been conducted on ~~devoted to~~ the development of calibration methods. Methodologies for~~of~~ model calibration range from ~~the~~ simple trial-and-error methods ~~, which~~ that adjust~~s~~ one parameter value in each ~~turn~~ iteration until the differences between predicted and observed values are~~is~~ satisfactory~~,~~ to ~~the~~ Bayesian updating framework~~, which~~ that reject~~ed~~ the concept of a ~~idea that there is~~ single correct solution. To a great extent, ~~No matter what kinds of methods, hydrological models are basically calibrated based on the runoff data alone (Dembélé et al., 2020), so t~~the success of model calibration is ~~, to a great extent,~~ dominated by the availability of field-observed runoff data. However, runoff data are generally only gathered at a ~~only~~ few sites~~,~~ and some cities ~~even~~ never measure~~d~~ runoff data in ~~the~~ built-up regions (Gebremedhin et al., 2020). Al~~Even~~ though ~~the~~ runoff data ~~could~~ can be effectively collected by administration departments in some cities, they have~~d~~ no motivation~~s~~ to share these data ~~to~~ with the public. For example, among China's top 10 largest cities[1], only Shenzhen has shared runoff-related data on ~~the~~ an open data platform. F~~As f~~or ~~the~~ model calibration ~~for~~ at the road scale, ~~the~~ runoff data are even more difficult to acquire~~,~~ because ~~a~~ road networks are~~is~~ far denser than ~~a~~ river networks and flood gauges are only ~~located~~ installed in a few flood-prone roads ~~considering~~ based on their high measurement cost, leaving most ~~of~~ roads ~~ungaged~~ungauged. As pointed out by Beven (2012, p:55), "the ungauged catchment problem is one of the real challenges for hydrological model~~l~~ers."

This~~e~~ lack of hydrological data has prompted researchers to seek ~~extra~~ additional data sources to support flood-related decision ~~-~~making. ~~In response to this need, big data, owing to~~Based on the advancement of mobile telecommunication technologies, big data are emerging as alternative sources of information for coping with flood risks (Paul et al., 2018; Li et al., 2018; Gebremedhin et al., 2020). Citizens can voluntarily or passively ~~acting~~ act as human sensors to generate georeferenced data to improve flood monitoring. ~~Typical~~ Many studies have leveraged~~involve the use of~~ crowdsourced~~ing~~ social media data ~~(Brouwer and Eilander, 2017; Li et al., 2018)~~ (Brouwer and Eilander, 2017; Sadler et al., 2018; Zahura et al., 2020), mobile phone data (Yabe et al., 2018; Balistrocchi et al., 2020), and taxi GPS data (She et al., 2019; Kong et al., 2022). However, most ~~of~~ previous works have concentrated on using big data either for flood mapping or mining spatiotemporal patterns (Restrepo-Estrada et al., 2018), and~~-~~ parameter calibration for ungauged roads based on big data remains an open problem~~it remains an open question of how to calibrate parameters based on big data for ungauged roads~~.

[1] Rank by the resident population in 2021.

This study ~~differentiates~~ extends ~~from~~ our previous study (Kong et al., 2022) by going ~~one~~ a step further than simply recognizing ~~the~~ flooded~~ing~~ roads. We propose a calibration method for road-related hydrological parameters ~~using~~ based on ~~the~~ taxi GPS data. Many studies have shown that vehicle-related information during ~~the~~ rainfall, ~~such as~~including vehicle volume, speed, and trajectory information, is useful~~critical~~ for~~to~~ flooded~~ing~~ road detection (Zhang et al., 2019; Qi et al., 2020; Yao et al., 2020). When a road segment is inundated by ~~the~~ heavy rainfall, the vehicle volume may ~~present~~ exhibit a sharp or gradual drop~~,~~ depending on the intensity of the rainfall event. Conversely, an~~the~~ abnormal drop in~~of~~ vehicle volume during the rainfall may imply~~ies~~ that ~~the~~ a road ~~may~~ has experienced ~~some~~ rainfall-induced inundation~~s~~. This motivates~~d~~ us to use~~ a~~ traffic-related data sources to calibrate hydrological parameters. In this study, we developed a transformation process ~~which~~ that converts ~~the~~ rainfall time series data into ~~the~~ a time series of ~~the~~ probabilities~~y~~ that no taxis will drive ~~through~~ on ~~the~~ a road (no-taxi-passing probability hereafter) for ~~every~~ a given hydrological parameter set.~~, and~~ We then assigned a probability to every parameter set by integrating the no-taxi-passing probability with ~~the~~ observed taxi GPS data. We ~~not only~~ outlined a generalized taxi-data-driven calibration framework and ~~but also realize~~implemented ~~the~~ a framework with specific hydrological and transportation models.

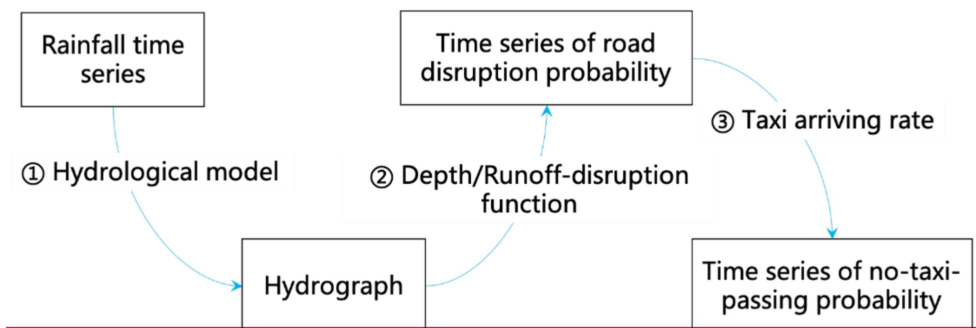
2 Methodology

2.1 ~~A~~ Bayesian updating procedure

Observed data are not always as informative as expected and may be inconsistent with other data sources;~~,~~ ~~so~~ hydrologists ~~usually~~ typically adopt the Bayesian framework to update hydrological parameters, which provides a generalized formalism that integrates prior probability representing ~~the~~ prior knowledge with ~~the~~ likelihood that reflects how ~~well~~ accurately ~~the presumed~~a model can reproduce ~~the~~ observations to form ~~the~~ a posterior probability. Suppose we have several hydrological models, each with different sets of parameters. Then,~~,~~ the purpose of the Bayesian updating procedure adopted in this study is to assign a posterior probability to every hydrological parameter set as new taxi data become available.

Two components are critical for ~~the~~ this Bayesian updating procedure. One is the prior probability~~,~~ and the other is the likelihood function. Regarding~~For~~ the prior probability, for their famous calibration model called generalized likelihood uncertainty estimation, Beven and Binley (1992) stated ~~in their famous calibration model, generalized likelihood uncertainty estimation (GLUE),~~ that all ~~the~~ parameter combinations are considered equally probable before ~~extra~~ additional information is introduced. After the first update~~ing~~, the prior probability of each updating ~~run~~ iteration ~~could~~ can be replaced by the posterior probability of the latest updating iteration~~run~~. L~~The l~~ikelihood, which is~~as~~ a measurement of how well ~~the~~ a given model conforms to the observed taxi behavio~~u~~r, is not as easy to compute as the prior probability~~,~~ because the parameter set to be estimated is hydrology-~~-~~related, whereas~~ile~~ the observed evidence is taxi-~~based~~related. Therefore, ~~question then arises as~~ we must determine how to construct a taxi-based proxy whose probability ~~equals~~is equal to ~~to that of~~ the associated hydrological parameter and construct a function enabling the transformation from ~~the~~ hydrological parameters to ~~the~~ taxi-related proxies~~y~~.

The ~~selected~~ proxy selected in this study was~~is~~ the time series of ~~probability that no taxi drives through the road in a given time interval (short for~~ the no-taxi-passing probability~~)~~. Figure 1 ~~illustrates~~ presents a generalized procedure for~~of~~ converting ~~the~~ a ~~precipitation~~ rainfall ~~process~~ time series into ~~the~~ a time series of no-taxi-passing probabilities~~y~~ for each hydrological parameter. Thise procedure consists of three steps. First, a hydrological model is used to convert ~~the~~ a rainfall time series in~~process~~ to ~~the~~ a hydrograph~~, which is a graph showing runoff with respect to time past a specific point~~. Second, a runoff-disruption function that ~~, which~~ relates ~~the~~ runoff to the probability that ~~the~~ a road is blocked~~,~~ is used to ~~transition~~ transform the hydrograph into ~~the~~ a time series of road disruption probabilities~~y~~. Third, the~~a~~ taxi ~~arriving~~ arrival rate is combined with the time series of road disruption probabilities~~y~~ to ~~yield~~ derive ~~the~~ a time series of no-taxi-passing probabilities~~y~~. ~~Note that t~~The hydrological model and ~~the~~ taxi arrival~~arriving~~ rate are considered to be unique for every road and are invariable within a short period, ~~while~~ whereas the runoff-disruption function is identical for all roads.



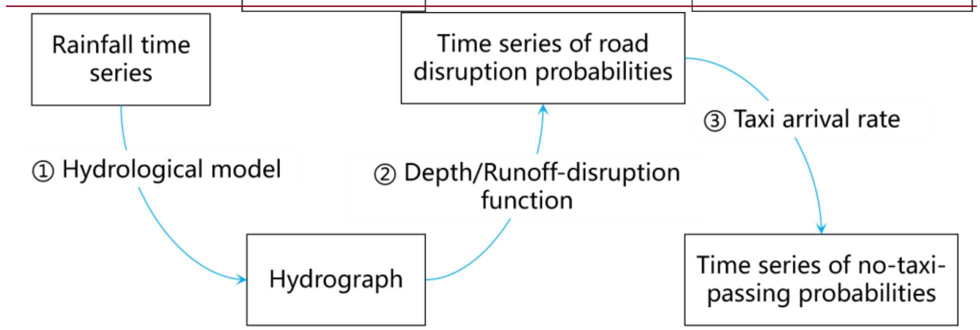


Figure 1 ~~A~~ gGeneralized procedure for~~of~~ converting ~~the~~ a rainfall time series into ~~the~~ a time series of no-taxi-passing probabilities~~y~~.

Integrating thise three-step process with the Bayesian equation enables us to compute the posterior probability of ~~the~~ a parameter set based on ~~the~~ taxi data. For a specific road, suppose there are N hydrological parameter sets to be estimated. ~~As~~ Because the runoff-disruption function and ~~the~~ taxi arrival~~arriving~~ rate are assumed to be ~~invariant~~ fixed for the road, we can construct a composite function converting the ith parameter set, which is denoted as $\boldsymbol{\theta}^{(i)}$ ~~$\theta^{(I)}$~~, into ~~the~~ a time series of no-taxi-passing probabilities~~y~~, which is denoted as $\boldsymbol{\Omega}^{(i)}$. Therefore, the probability of $\boldsymbol{\theta}^{(i)}$ ~~to be~~being optimal is equal~~s~~ to the probability of $\boldsymbol{\Omega}^{(i)}$ being~~to be~~ true, which can be expressed as follows:

$$P\left(\boldsymbol{\theta}^{(i)}\right) = P\left(\boldsymbol{\Omega}^{(i)}\right) \tag{1}$$

where $P\left(\boldsymbol{\theta}^{(i)}\right)$ and $P\left(\boldsymbol{\Omega}^{(i)}\right)$ are the prior probabilit~~y~~ies of $\boldsymbol{\theta}^{(i)}$ and $\boldsymbol{\Omega}^{(i)}$ respectively. As taxi observations become available, $P\left(\boldsymbol{\theta}^{(i)}\right)$ (or $P\left(\boldsymbol{\Omega}^{(i)}\right)$) can be updated using the Bayes ~~T~~theorem as:

$$\cancel{P(\theta^{(i)} \mid X) = P(\Omega^{(i)} \mid X) \propto P(\Omega^{(i)})\mathscr{L}(X \mid \Omega^{(i)})} \quad \cancel{(2)}$$

$$P\left(\boldsymbol{\theta}^{(i)} \middle| \boldsymbol{X}\right) = P\left(\boldsymbol{\Omega}^{(i)} \middle| \boldsymbol{X}\right) \propto P\left(\boldsymbol{\theta}^{(i)}\right)\mathcal{L}\left(\boldsymbol{X} \middle| \boldsymbol{\theta}^{(i)}\right) \tag{2}$$

where $\boldsymbol{X}$ is the taxi observation, and $P\left(\boldsymbol{\theta}^{(i)}|\boldsymbol{X}\right)$ and $P\left(\boldsymbol{\Omega}^{(i)}|\boldsymbol{X}\right)$ are the posterior probabilit~~y~~ies of $\boldsymbol{\theta}^{(i)}$ and $\boldsymbol{\Omega}^{(i)}$~~, $\Omega^{(i)}$~~ respectively ~~under the~~ conditional on the~~of~~ taxi observation~~, X~~. ~~The $\mathscr{L}(X|\Omega^{(i)})$~~ $\mathcal{L}\left(\boldsymbol{X}\middle|\boldsymbol{\theta}^{(i)}\right)$ is the likelihood of X given ~~$\Omega^{(i)}$~~ $\boldsymbol{\theta}^{(i)}$. The optimal parameter set is ~~the one~~ that which ~~derives~~ yields the $\Omega^{(i)}$ that most closely~~best~~ fits the observed taxi data.

~~The solution of~~ Solving Eq. (2) involves the calculation of ~~$P(\Omega^{(i)})$~~ $P\left(\boldsymbol{\theta}^{(i)}\right)$ and ~~$\mathscr{L}(X|\Omega^{(i)})$~~ $\mathcal{L}\left(\boldsymbol{X}\middle|\boldsymbol{\theta}^{(i)}\right)$. ~~According to Eq.(1), $P(\Omega^{(i)})$ can be replaced with $P(\theta^{(i)})$, which is the prior probability of parameter sets.~~ The derivation of $P\left(\boldsymbol{\theta}^{(i)}\right)$ depends on prior knowledge regarding the parameter distribution, which is typically obtained using traditional hydrological methods. However, this prerequisite knowledge may not always be readily accessible based on limited data availability. In such cases, Beven and Binley (1992) suggested that ~~prior to the introduction of any quantitative and qualitative information,~~ any parameter set combination ~~should~~ could be considered to be equally likely. This impli~~ed~~es that the parameter set is drawn from a uniform distribution as follows:

$$\cancel{P(\Omega^{(i)}) = P(\theta^{(i)}) = 1/N} \quad \cancel{(3)}$$

$$P\left(\boldsymbol{\theta}^{(i)}\right) = 1/N \tag{3}$$

In this study, we compared the effects of two types of prior parameter distributions, namely a uniform distribution and a distribution derived from digital elevation model (DEM) data, on the resulting posterior distributions.

~~Next,~~ $\mathcal{L}\left(\boldsymbol{X}\middle|\boldsymbol{\theta}^{(i)}\right)$~~$\mathscr{L}(X|\Omega^{(i)})$~~, which is~~as~~ a likelihood function, describes the joint probability of the observed taxi data~~,~~ $\boldsymbol{X}$~~,~~ as a function of the chosen ~~$\Omega^{(i)}$~~ $\theta^{(i)}$. Consider a rainfall event that is ~~broken~~ divided into T 5 min intervals. From the taxi data, we ~~obtained~~ can obtain a sequence of taxi-related observations, which are denoted as $\boldsymbol{X} = \{x\cancel{X}_1, x\cancel{X}_2, \dots, x\cancel{X}_T\}$, where $x\cancel{X}_t = 1$ if the observed road has~~is observed with~~ at least one taxi pass~~ing~~ during~~by~~ ~~in~~ the tth ~~5 min~~ interval, and $x\cancel{X}_t = 0$ otherwise. ~~The~~ $\boldsymbol{\Omega}^{(i)} = \left\{\omega_1^{(i)}, \omega_2^{(i)}, \dots, \omega_T^{(i)}\right\}$ is also a T-dimensional vector, where $\omega_t^{(i)}$ is the no-taxi-passing probability ~~at~~ in the tth ~~5 min~~interval ~~taking~~ with $\boldsymbol{\theta}^{(i)}$ as the parameter set. Note that $\boldsymbol{\Omega}^{(i)}$ is only determined by the chosen hydrological parameter and ~~the~~ rainfall ~~process~~time series, and is not measured from ~~the~~ observed data. Considering that the arrival of taxis is independent of ~~with respect to~~ time, $\mathcal{L}\left(\boldsymbol{X}\middle|\boldsymbol{\theta}^{(i)}\right)$~~$\mathscr{L}(X|\Omega^{(i)})$~~ can be formulated as:

$$\cancel{\mathscr{L}(X|\Omega^{(i)}) = \prod_{t=1}^{T}\left(1-\omega_t^{(i)}\right)^{X_t}\left(\omega_t^{(i)}\right)^{1-X_t}} \quad \cancel{(4)}$$

$$\mathcal{L}\left(\boldsymbol{X}|\boldsymbol{\theta}^{(i)}\right) = \mathcal{L}\left(\boldsymbol{X}|\boldsymbol{\Omega}^{(i)}\right) = \prod_{t=1}^{T}\left(1-\omega_t^{(i)}\right)^{x_t}\left(\omega_t^{(i)}\right)^{1-x_t} \qquad (4)$$

By substituting ~~Eq.(3) and~~ Eq. (4) ~~back~~ into Eq. (2), the ~~Eq.(5) is~~ following equation can be obtained:

$$\cancel{P(\theta^{(i)}|X) \propto \frac{1}{N}\prod_{t=1}^{T}(1-\omega_t^{(i)})^{X_t}(\omega_t^{(i)})^{1-X_t} \qquad (5)}$$

$$P\left(\boldsymbol{\theta}^{(i)}|\boldsymbol{X}\right) \propto P\left(\boldsymbol{\theta}^{(i)}\right)\prod_{t=1}^{T}\left(1-\omega_t^{(i)}\right)^{x_t}\left(\omega_t^{(i)}\right)^{1-x_t} \qquad (5)$$

Equation (5) is the proposed Bayesian updating model for~~to~~ calibrating~~e~~ ~~the~~ hydrological parameters based on ~~the~~ taxi data, where $\boldsymbol{X}$ ~~could~~ can be directly measured and ~~$\Omega^{(i)}$~~ $\omega_t^{(i)}$ is calculated through the three-step process ~~shown~~ illustrated in Fig. 1, which will be discussed in detail in the ~~next~~ following section. Having ~~chosen~~ selected an updating model, the optimal~~um~~ parameter for one period of observations may not be optimal for another period. ~~As~~ Because the model may have ~~a~~ continuing inputs of new taxi observations, the posterior probability for $\boldsymbol{\theta}^{(i)}$ should be updated as new evidence becomes available. For the second update, the posterior probability from the first observation becomes the prior probability for the second observation~~,~~ and the posterior probability for $\boldsymbol{\theta}^{(i)}$ is recursively updated as:

$$P\left(\boldsymbol{\theta}^{(i)}|\boldsymbol{X_2}\right) \propto \mathcal{L}\left(\boldsymbol{X_2}|\boldsymbol{\theta}^{(i)}\right)P\left(\boldsymbol{\theta}^{(i)}|\boldsymbol{X_1}\right) \qquad (6)$$

where $\boldsymbol{X_1}$ and $\boldsymbol{X_2}$ are the first and the second taxi observations.

2.2 Instantiation ~~Instantization~~ of the three-step procedure

Section 2.1 presented~~s~~ a generalized three-step procedure ~~which~~ for converting~~s~~ ~~the~~ a rainfall time series into ~~the~~ a time series of no-taxi-passing probabilities~~y~~. In this section, we specialize this~~e~~ process by integrating existing theories with our model. The t~~T~~hree conceptualized steps ~~shown~~ illustrated in Fig. 1 ~~are~~ were replaced~~substituted~~ with three ~~more~~ concrete sub-models. First~~ly~~, a Soil Conservation Service (SCS) unit hydrograph was~~is~~ used to ~~turn~~ convert ~~the~~ rainfall excess into ~~the~~ a hydrograph of the target road. Second~~ly~~, an empirical runoff-disruption function based on ~~, whose~~ data extracted from various experimental, observational, and model~~l~~ing studies~~, is~~ was applied to convert the hydrograph into ~~the~~ a time series of ~~the~~ road disruption probabilities~~y~~. Third~~ly~~, a Poisson distribution~~,~~ representing the distribution of taxi arrival~~arriving~~ rate~~, is~~ was combined with the road disruption probability time series to ~~yield~~ derive ~~the~~ a no-taxi-passing probability time series.

Step 1: Converting rainfall into runoff based on the SCS unit hydrograph

Not all rainfall ~~will~~ produces runoff because ~~storage from~~ soil~~s~~ storage can absorb a certain amount of rain~~light shower~~. ~~While~~ However, in ~~the~~ urbanized areas, only a small proportion of rainfall infiltrates ~~into~~ the soil or is retained on the land surface, leaving most ~~of them~~rain to flow across ~~the~~ urban surfaces and ~~becomes~~ become ~~the~~ direct runoff. The rainfall that ~~yields~~ becomes ~~the~~ direct runoff is ~~termed~~ referred to as rainfall excess. The Natural Resources Conservation Service

(NRCS)[2] developed a method to estimate ~~the~~ rainfall excess based on ~~the~~ soil types and land uses using the following curve number equation:

$$P_e = \begin{cases} (P_a - 0.2S)/(P_a + 0.8S) & P_a > 0.2S \\ 0 & P_a \le 0.2S \end{cases} \tag{7}$$

where P_e is the accumulated rainfall excess in cm, P_a is the accumulated rainfall in cm, and S is the potential retention after runoff begins, which is defined as a function of the curve number as follows:

$$S = 2.54 \times (1000/CN - 10) \tag{8}$$

where CN is the curve number. For urban and residential land, the curve number varies from ~~65~~ 40 to ~~85~~95~~,~~ depending on the impervious area~~s~~ (Natural Resources Conservation Service, 2010a). Because prior knowledge on the CN is unavailable, it was considered as a calibrated parameter in this study.~~For sake of brevity, the curve number will not be regarded as a parameter to be calibrated in this study but as a given parameter with the value of 85.~~

~~Next, t~~The rainfall excess derived using~~by~~ Eq. (7) was~~is~~ inputted into the unit hydrograph to ~~produce~~ derive the runoff. The unit hydrograph is a commonly used rainfall-runoff model that converts rainfall excess into a temporal distribution of direct runoff~~converts rainfall excess to direct runoff~~. First proposed by Sherman in 1932, the unit hydrograph is defined as the hydrograph resulting from one unit of rainfall excess distributed uniformly over a catchment area. It assumes ~~the~~ that rainfall is uniform over the catchment area and that runoff increases linearly with ~~the~~ rainfall excess. Although ~~under most conditions~~ these assumptions cannot be perfectly satisfied under most conditions, the results obtained from the unit hydrograph are generally acceptable for most practical ~~uses~~cases. The model~~,~~ ~~was~~ originally designed for larger watersheds, ~~but it~~ has been found to be applicable to some catchment~~s~~ areas less than 5,000 m^2 in size (Chow et al., 1988).

The unit hydrograph ~~applies~~ is only applicable to~~for the~~ watershed areas where ~~the~~ runoff data are~~were~~ measured. The paucity of ~~the~~ runoff data motivated~~sparkled~~ the ~~idea~~ development of the synthetic unit hydrograph (SUH) concept. The term “synthetic” in SUH ~~denotes the~~ refers to a unit hydrograph derived from watershed characteristics, rather than empirical rainfall-runoff relationships. In this study, we utilized the SCS unit hydrograph, which is a dimensionless SUH proposed by the NRCS. For the dimensionless SUH, the discharge (i.e., y-axis) is expressed as the ratio of discharge q to ~~the~~ peak discharge q_p and the time (i.e., x-axis) is expressed as the ratio of time t to ~~the~~ peak time t_p. Therefore, the SCS unit hydrograph~~, rigorously speaking,~~ is not exactly an SUH itself, but is a useful tool for constructing an SUH.

The shape of an SCS unit hydrograph is ~~totally~~ entirely determined by the peak rate factor. A standard value of 2.08 for the peak rate factor is recommended and commonly used by the NRCS (Fig. 2). To construct an SUH from ~~the~~ an SCS unit hydrograph, the x-axis of the SCS unit hydrograph is multiplied by t_p and the y-axis is multiplied by q_p. The values of q_p and t_p are functions of the catchment area and ~~the~~ time of concentration as follows:

$$t_p = 0.6 t_C + D/2 \tag{9}$$

$$q_p = 2.08 A / t_p \tag{10}$$

[2] The NRCS ~~used to be~~ was originally called the US Soil Conservation Service ~~(SCS)~~.

where t_c is the time of concentration in hours, A is the catchment area in km^2, and D is the duration of unit rainfall excess in hours, which ~~i~~was set to ~~1/12 h~~ one-twelfth of an hour (i.e., 5 min) in this study. Notably ~~As can be seen~~, the catchment area and time of concentration are required to construct an SUH~~,~~ and they are the other two hydrological parameters ~~we would~~ that should be calibrated based on ~~the~~ taxi data. Although numerous tools and theories have been developed for estimating catchment area and time of concentration, these two parameters are still prone to significant errors, particularly in urban areas, because of challenges in accurately delineating urban catchments (Huang and Jin, 2019; Li et al., 2020). Urban catchment delineation is more complex than natural catchment delineation. Urban catchments have spatially heterogeneous surface cover types, which change with city development and construction, and modify runoff parameters (Goodwin et al., 2009). High densities of residential and commercial buildings obstruct flow paths and alter flow directions of storm water runoff, complicating rainfall-runoff and overland flow processes in urban areas (Ji and Qiuwen, 2015). Additionally, accurate urban catchment delineation necessitates high-resolution DEMs, which are not always available. In many Chinese cities, high-resolution DEMs are considered confidential data and are generally inaccessible to non-governmental organizations. Based on these challenges, deriving accurate catchment area and time of concentration data in urban areas is difficult in Shenzhen.

For the sake of simplicity, the peak rate factor ~~would~~ was not ~~be~~ calibrated and ~~be~~ was ~~fixed as~~ at 2.08, although some studies have ~~showed~~ indicated that it has a ~~much~~ wide~~r~~ range from 0.43 for steep terrain to 2.58 for very flat terrain (Chow et al., 1988). After t_c and A were~~are~~ chosen, an SUH ~~could~~ can be constructed~~,~~ and ~~then we~~ used it to convert ~~the~~ rainfall excess into ~~the~~ runoff by applying the discrete convolution equation. The detailed computation process of the discrete convolution equation can be found in most hydrological textbooks (e.g., ~~see~~ Chow et al., (1988)~~ pp: 211-213~~), and will not be discussed here. ~~To be clear, a~~ The graphical workflow in Fig. 3 ~~shows~~ illustrates the transformation of ~~how the~~ rainfall time series ~~is~~ data ~~transformed~~ into ~~the~~ a hydrograph for every parameter set.

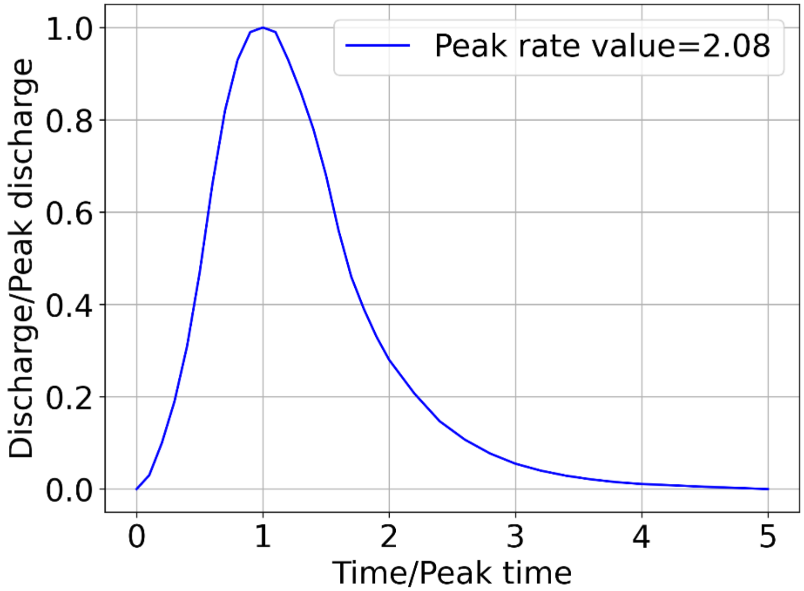


Figure 2 ~~The s~~Standard SCS unit hydrograph. Data provided by the NRCS (2007).

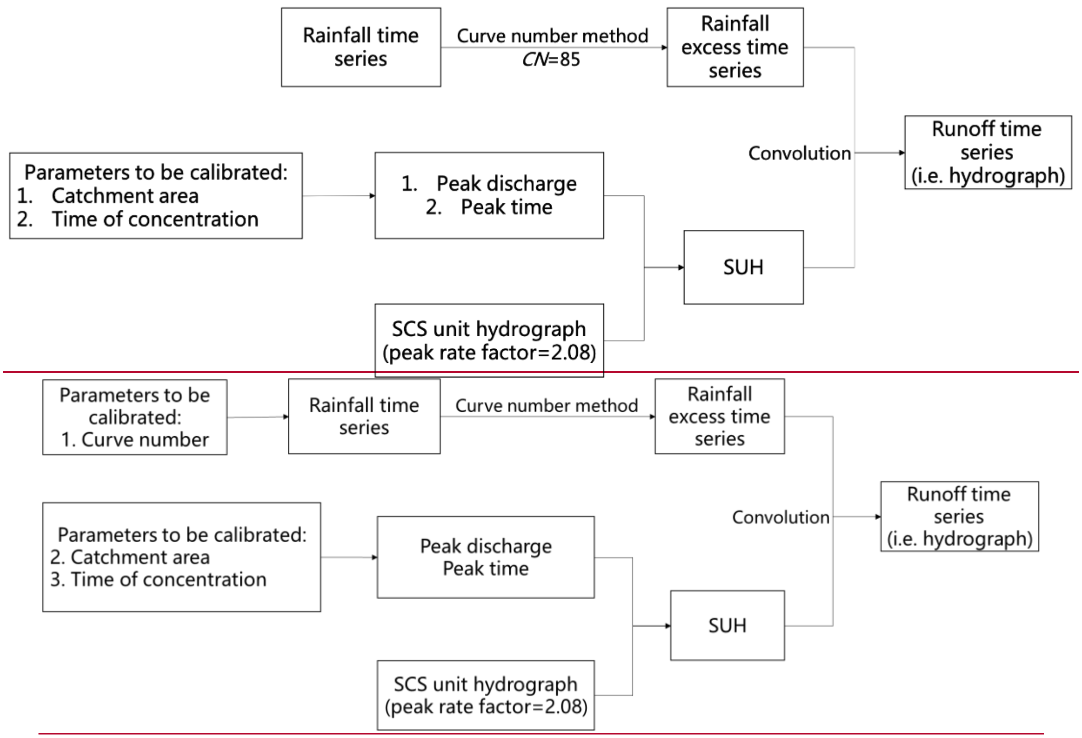


Figure 3 Workflow of the SCS unit hydrograph ~~to~~ for converting rainfall into runoff.

Step 2: Derivatione of the road disruption probability using the runoff-disruption function

The goal of Step 2 is to convert the hydrograph generated in Step 1 into ~~the~~ a time series of road disruption probabilities~~y~~, or more specifically, the probability that a taxi driver chooses to turn their car when arriving at a flooded road. Most models in the literature assume that a road is either open or closed, which ~~usually~~ does not correspond to the empirical evidence that many drivers may take risks to drive on inundated roads~~along the rood even though it is inundated~~. ~~In order~~ ~~t~~To transition from a binary view of a flooded road being considered “open” or “closed,”~~,~~ Pregnolato et al. (2017) proposed ~~to~~ the use of a curve that relates the depth of floodwater to a reduction in vehicle speed to ~~show~~ indicate the probability of road disruption.~~,~~ ~~and such~~ This idea was~~is~~ soon ~~followed~~ adopted by Contreras-Jara et al. (2018) and Nieto et al. (2021).

A driver will turn around when he believes that the flow rate is too risky for their ~~overcomes the~~ vehicle configuration. From this perspective, the road disruption probability is ~~equals to~~equal to the probability that ~~the~~ vehicle performance is ~~lower~~ less than the flow rate perceived by a driver. However, it is ~~a~~ difficult ~~task~~ to quantify the factors that influence~~common belief of what guide people’s~~ willingness of people to drive through a flooded ~~waterway~~roadway, and ~~is also~~ impossible~~difficult~~ to obtain the precise knowledge ~~of~~ regarding all taxi-flood intersections. Alternatively, to ensure ~~the~~ vehicle stability in flood flows, guidelines are ~~usually~~ typically recommended based on the limiting value~~s~~ of depth times velocity.~~,~~ ~~and m~~Many ~~studies~~ researchers have ~~carried out~~conducted laboratory testing on the stability of different ~~kinds~~ types of vehicle models exposed to different combinations of depth and velocity (Merz and Thieken, 2009; Shah et al., 2018). As s~~S~~uggested by Pregnolato et al. (2017), we constructed ~~the~~ our runoff-disruption function by integrating data from ~~reviewed~~ the literature~~s~~ and ~~some~~ authoritative guidelines. In this study, the road disruption probability was~~is~~ defined as the probability that the product of flow velocity and flow depth ~~is higher~~ was greater than the stability limits extracted~~ing~~ from the literature~~existing studies~~, which are ~~shown~~ listed in Table 1 and plotted in Fig. 4. The expression of the fitting curve is:

$$y = [1 + \exp(-16.6(x - 0.48)^2)]^{-1} \tag{11}$$

where x is the product of flow velocity and flow depth, and y is the disruption probability. According to Eq. (11), a road has a disruption probability of 50% when the product of flow velocity and flow depth is 0.47 $m^2\ s^{-1}$~~,~~ and is totally disrupted when the product is ~~higher~~ greater than 0.80 $m^2\ s^{-1}$. By ~~A~~applying the fitting curve, we can easily convert the flood runoff into the disruption probability as follows:

$$P(Disrupt)_t^{(i)} = \left[1 + \exp\left(-16.6\left(q_t^{(i)}/W - 0.48\right)^2\right)\right]^{-1} \tag{12}$$

where $P(Disrupt)_t^{(i)}$ and $q_t^{(i)}$ are the road disruption probability and discharge ~~at~~ in the tth interval~~5 min~~ derived ~~by~~ from the hydrological model with the parameter set $\theta^{(i)}$, respectively, and W is the road width.

Table 1 Guidelines recommended ~~by~~ in the existing literature~~s~~.

Reference	Vehicle type	Feature	Recommended limits for vehicle stability ($m^2\ s^{-1}$)
Shah et al. (2018)	Volkswagen Scirocco	Flow direction =0°	velocity×depth<0.014
Al-Qadami et al. (2022)	Perodua Viva	Ground clearance =0.18 m	velocity×depth<0.39

Calculated according to Kramer et al. (2016)	VW Golf III	Not mention~~/~~	velocity×depth<0.42
Shand et al. (2016)	Large passenger	Ground clearance >0.12 m	velocity×depth<0.45
Martínez-Gomariz et al. (2017)	Mini Cooper	Ground clearance =0.12 m	velocity×depth<0.49
Martínez-Gomariz et al. (2017)	BMW i3	Ground clearance =0.10 m	velocity×depth<0.49
Martínez-Gomariz et al. (2017)	BMW 650	Ground clearance =0.08 m	velocity×depth <0.50
Martínez-Gomariz et al. (2017)	Mercedes GLA	Ground clearance =0.17 m	velocity×depth <0.59
Moore and Power (2002)	All but very small cars	Not mention~~/~~	velocity×depth <0.60
Calculated according to Xia et al. (2014)	Honda Accord	Not mention~~/~~	velocity×depth <0.65

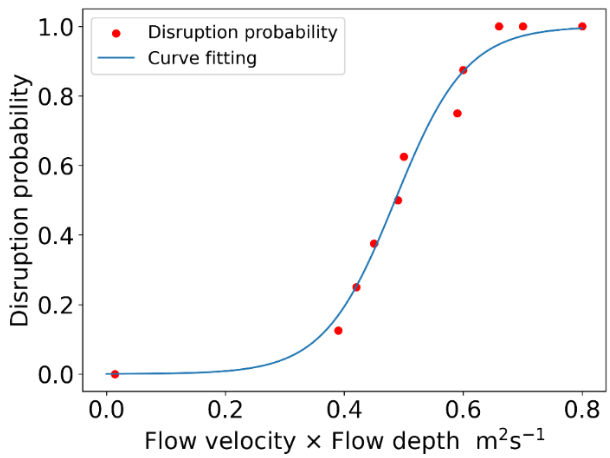


Figure 4 Empirical runoff-disruption function derived from the existing literature.~~s~~

Step 3: Derivation~~e~~ of the time series of no-taxi-passing probabilities~~y~~

A road ~~has~~ is considered to have no taxis passing ~~by~~ in a fixed time ~~step~~ interval if the road has no taxis ~~visiting~~ arriving or if every taxi that arrives ~~at~~ the road turns around.~~,~~ ~~so~~ Therefore, the no-taxi-passing probability can be ~~inferred~~ calculated ~~by~~ using the following equation:

$$\cancel{\omega_t^{(i)} = \sum_{n=0}^{\infty} P(Arrival\_taxi = n) \times \left(P(Disrupt)_t^{(i)}\right)^n \quad (13)}$$

$$\omega_t^{(i)} = \sum_{n=0}^{\infty} P(Arrival\_taxi = n)_t \times \left(P(Disrupt)_t^{(i)}\right)^n \tag{13}$$

where $\omega_t^{(i)}$ is the no-taxi-passing probability in the tth interval ~~5 min~~, and ~~$P(Arrival\_taxi=n)$~~ $P(Arrival\_taxi = n)$~~$(Arrival_t axi = n)$~~$_t$ is the probability that n taxis arrive~~s~~ at the road segment during~~in~~ the tth ~~5 min~~interval. Equation (13) indicates that if every taxi that arrives at the road segment makes a turn because of the flooded ~~waterway~~roadway, then the taxi volume ~~of~~ on the road will be zero. In this study, $P(Arrival\_taxi = n)$~~$(Arrival_t axi = n)$~~$_t$ ~~$P(Arrival\_taxi=n)$~~ was~~is~~ assumed to follow the Poisson distribution:

$$\cancel{P(Arrival\_taxi=n)=e^{-\lambda_t}\lambda^n/n! \quad (14)}$$

$$P(Arrival\_taxi = n)_t = \mathrm{e}^{-\lambda_t}{\lambda_t}^n/n! \tag{14}$$

where ~~λ~~λ_t is the average number of taxis arriving at the road during the tth interval. By ~~S~~substitut~~e~~ing Eq. (14) in~~to~~ Eq. (13), we ~~derive~~obtain:

$$\cancel{\omega_t^{(i)}=\sum_{n=0}^{\infty}(e^{-\lambda}\lambda^n/n!)\times\left(P(Disrupt)_t^{(i)}\right)^n \quad (15)}$$

$$\omega_t^{(i)} = \sum_{n=0}^{\infty}\left(\mathrm{e}^{-\lambda_t}{\lambda_t}^n/n!\right) \times \left(P(Disrupt)_t^{(i)}\right)^n \tag{15}$$

By ~~A~~applying $e^x = \sum_{n=0}^{\infty} x^n/n!$, Eq. (15) can be ~~further~~ converted into:

$$\cancel{\omega_t^{(i)}=e^{-\lambda}\sum_{n=0}^{\infty}\left(P(Disrupt)_t^{(i)}\lambda\right)^n/n! = \exp\left(\lambda\left(P(Disrupt)_t^{(i)}-1\right)\right) \quad (16)}$$

$$\omega_t^{(i)} = \mathrm{e}^{-\lambda_t}\sum_{n=0}^{\infty}\left(P(Disrupt)_t^{(i)}\lambda_t\right)^n/n! = \exp\left(\lambda_t\left(P(Disrupt)_t^{(i)} - 1\right)\right) \tag{16}$$

Equation (16) indicates that $\omega_t^{(i)}$ is ~~totally~~ entirely determined by ~~λ~~λ_t and $P(Disrupt)_t^{(i)}$. ~~Since~~ Because $P(Disrupt)_t^{(i)}$ is ~~given~~ obtained ~~through~~ from Step 2, what is left to determine is the value of ~~λ~~λ_t. The value of λ_t fluctuates according to the time of day, indicating higher taxi volume during congested periods and lower volume during non-congested periods. Therefore, we calculate λ_t by averaging the taxi volume during the tth interval to account for time-of-day variations. It should be noted that ~~A~~as the intensity of rain increases ~~gets heavier~~, experienced taxi drivers will avoid flood-prone roads in advance, ~~which~~ meanings ~~s~~ that strictly speaking, ~~λ~~λ_t~~, strictly speaking,~~ is a decreasing function of rainfall intensity. However, fitting the rainfall-~~λ~~λ_t curve requires ~~substantial~~ many taxi GPS trajectories to inspect the route choices of taxi drivers under heavy rain, which is ~~currently unfeasible in~~outside the scope of this study. ~~We assume that λ is a constant quantity which keeps unchanged with respect to rainfall. The value of λ can be calculated by averaging all 5 min taxi volume using the historical taxi GPS data.~~ Therefore, we assumed that λ_t was independent of rainfall.

~~Finally,~~ Table 2 lists all the sub-models and parameters used in~~of~~ the three-step process. The core principle of the three-step process ~~i~~was to calculate the time series of no-taxi-passing probabilit~~y~~ies, $\Omega^{(i)}$~~,~~ given~~for~~ each parameter set~~,~~ $\theta^{(i)}$. ~~As~~Because the best choice of ~~a~~ model is often data~~-~~specific, it is ~~probable~~ likely that the model combination proposed in this study is not optimal for other ~~studies~~scenarios. To apply the proposed calibration method in practic~~al use~~e, ~~one must specify~~

the sub-models for~~in~~ the three-step process must be specified according to the available data, prior knowledge, and accuracy requirements.

Table 2 Specific sub-models and parameters used in~~of~~ the three-step process.

Purpose of ~~the~~ each step	Specific model	Parameter	Source of parameters
Step 1: Convert ~~the~~ rainfall data into ~~the~~ a hydrograph	Curve number equation	1. Curve number	Parameters to be calibrated ~~Existing literature~~
	SCS unit hydrograph	~~2. peak rate factor~~2. Catchment area 3. Time of concentration	~~Existing literature~~Parameters to be calibrated
		~~3. Catchment area~~ ~~4. Time of concentration~~4. Peak rate factor	~~Parameters to be calibrated~~Existing literature
Step 2: Convert the hydrograph into ~~the~~ a time series of disruption probabilities~~y~~	Empirical runoff-disruption function	5. Limit of product of flow velocity and depth	Existing literature~~s~~
Step 3: Convert the time series of disruption probabilities~~y~~ into ~~the~~ a time series of no-taxi probabilities~~y~~	Taxi arrival rate following~~s~~ ~~the~~ Poisson distribution	6. Average taxi volumes in 5 min periods	Taxi GPS data

3 ~~A w~~Working example

The method outlined above was tested on ~~an~~ the intersection of ~~located in~~ Xinzhou ~~R~~road and Hongli ~~R~~road in Shenzhen, which is recognized as a ~~waterlogging~~ flood-prone point by the Water Authority of Shenzhen Municipality. Recall that the parameters to be calibrated are the curve number *CN*, catchment area~~,~~ *A*, and time of concentration~~,~~ t_C. The parameter spaces for *CN*, *A*, and t_C. are determined by DEMs and other prior knowledge, which will be discussed in Section 4. ~~The range of parameters should be wide enough to encompass most possible values. After several rounds of testing, the maximum value for *A* is set as 0.5 km², and the maximum value for t_c is 5 h. Optimal parameter sets for most roads would fall into the region enclosed by the maximum parameter sets.~~ Table 3 ~~shows~~ presents the details ~~information~~ of the parameter sets to be calibrated, which ~~totally~~ form ~~3,000~~8 × 20 × 30 = 4,800 possible combinations. For ease of exposition, we assume that all parameters are uniformly distributed.

Table 3 Detailed information ~~of~~ on parameter sets to be calibrated.

Parameter	Annotation	Minimum	Maximum	Incremental	Number of possible parameter values
Curve number	*CN*	40	75	5	8
Catchment area	*A*	0.1 km² ~~0.01 km²~~	0.29 km² ~~0.5 km²~~	0.01 km²	20~~50~~
Time of concentration	t_C	0.75 h ~~1/12 h~~	3.2 h ~~5 h~~	1/12 h	30~~60~~

~~The t~~Taxi GPS data collected during two storm events that occurred~~ing~~ on ~~9~~ May 9, 2015 and ~~23~~ May 23, 2015 ~~were~~ are used to calibrate the parameter sets ~~of~~ for the target intersection. Rainfall time series data and taxi observations ~~under~~ during ~~two~~ these two storms are ~~shown~~ presented in Fig. 5. Each taxi observation contains two time series:~~.~~ ~~One is~~ the time series of ~~5 min~~ taxi volumes at 5 min intervals~~,~~ and ~~the other is the 5 min~~ time series of road statuses at 5 min intervals.~~,~~ ~~which is~~These were derived from the taxi volumes~~,~~ with ~~the~~ a value of one~~to be 1~~ if the taxi volume was~~is~~ ~~higher~~ greater than ~~0~~ zero and a value of zero~~0~~ if the taxi volume ~~is~~ was ~~0~~zero.

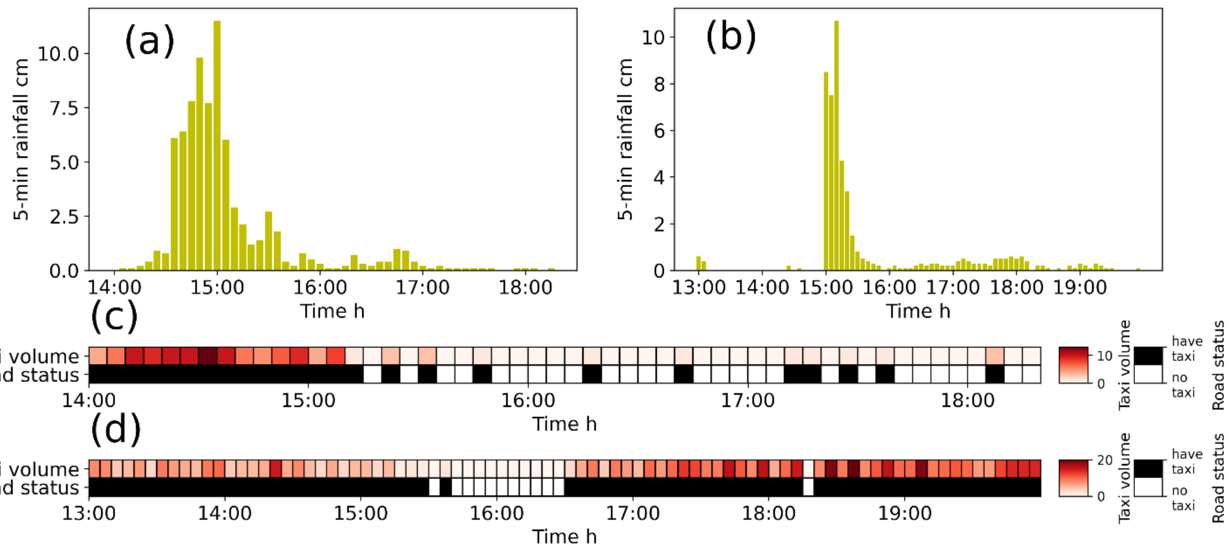


Figure 5 Rainfall and taxi observations used to calibrate ~~the~~ hydrological parameters:~~.~~ **(a)** ~~5 min r~~Rainfall time series in 5 min intervals~~on~~ on May 9, 2015,~~9 May 2015.~~ **(b)** ~~5 min r~~rainfall time series in 5 min intervals on ~~23~~ May 23, 2015,~~.~~ **(c)** ~~T~~taxi observations on ~~9~~ May 9, 2015,~~.~~ and **(d)** ~~T~~taxi observations on ~~23~~ May 23, 2015.

Given the rainfall on ~~9~~ May 9, 2015, we ~~should~~ must calculate the time series of no-taxi-passing probabilities ~~possibility~~ for each parameter combination. Because there are 4,800 parameter sets, we can derive 4,800 possible time series of no-taxi-passing probabilities. For simplicity, we only present the 3,120th parameter set (i.e., *CN* = 65, *A* = 0.2 km^2, and t_C = 2.75 h) as an example to demonstrate the working of the proposed method. According to the three-step process, the first step is to convert the original rainfall in~~to~~ ~~the~~ rainfall excess using the curve number method given *CN* = 65 (Fig. 6(a)). Then, ~~for each combination of *A* and t_c, we construct a SUH. As there are 3,000 parameter sets, we can construct 3,000 different SUHs. For simplicity, we only chose the 1,170th parameter set, i.e. *A* =0.2 km^2 and t_c =2.75 h, as examples to show the calibration works. Using Eq.(9) and Eq.(10), the~~ we calculated the peak discharge q_p and peak time t_p using Eqs. (9) and (10)~~can be calculated as~~:

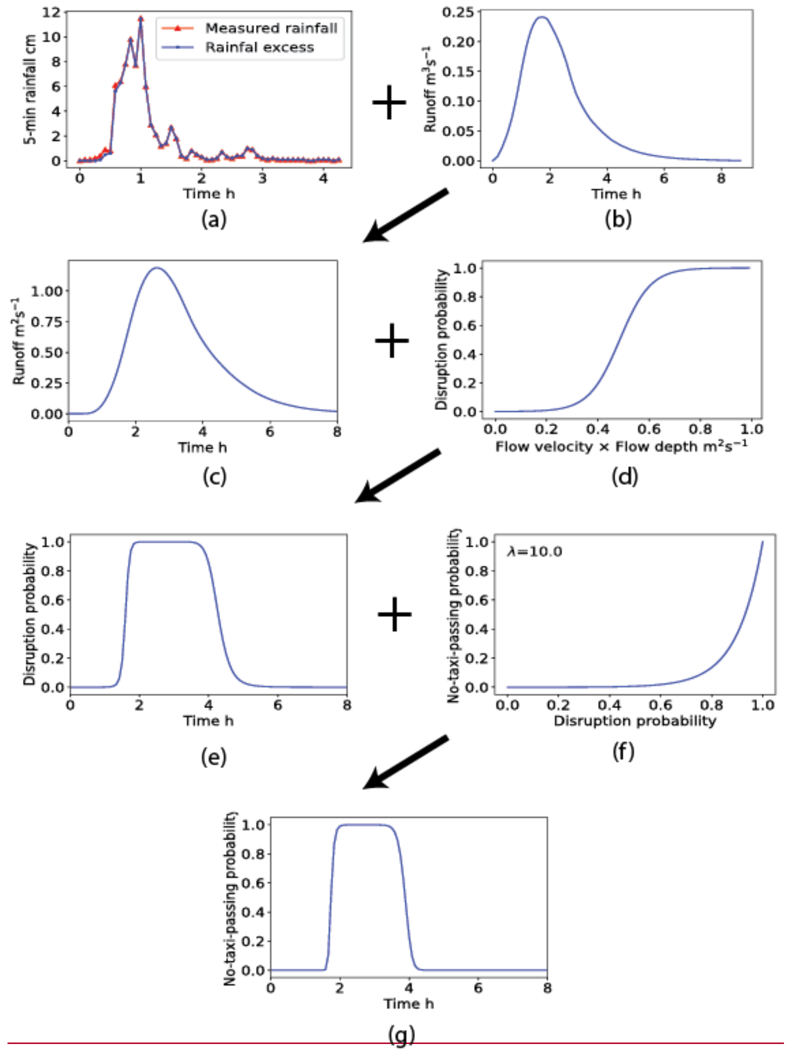
$$t_p = 0.6 \times 2.75 + \frac{1}{2 \times 12} \approx 1.69\ h$$

$$q_p = 2.08 \times \frac{0.2}{1.69} \approx 0.24\ \mathrm{m^3 s^{-1}}$$

The SUH ~~i~~was derived ~~by~~ through multiplication~~multiplied~~ by t_p on the *x*-axis and ~~by~~ q_p on the *y*-axis of the standard SCS unit hydrograph (Fig. 6(b)). Next, the rainfall excess ~~shown~~ presented in Fig. 6(a) ~~i~~was combined with the derived SUH to ~~yield~~ obtain ~~the~~ a hydrograph through ~~the~~ convolution (Fig. 6(c)).

In the second step, the runoff ~~i~~was transformed into ~~the~~ a time series of road disruption probabilities~~y~~ based on the runoff-disruption function (Fig. 6(d)). ~~Note that t~~The runoff-disruption function takes the product~~ion~~ of water depth and velocity (in ~~the~~ units of m^2 s^{-1}) as inputs. Therefore, the original runoff (in ~~the~~ units of m^3 s^{-1}) ~~produced by~~derived in the first step should be divided by the road width before inputting it into the runoff-disruption function.

In the third step, the time series of road disruption probabilities~~y~~ (Fig.6(e)) ~~i~~was converted to ~~that of~~ no-taxi-passing probabilities~~y~~ using Eq. (16) (Fig. 6(f)). ~~According to the historical taxi GPS data, the average number of taxis arriving at the road, λ, is 10.0 taxi per 5 min.~~ The average number of taxis during the flooding period is presented in Fig. (6f), and t~~T~~he derived time series of no-taxi-passing probabilities~~possibility~~ is ~~shown~~ presented in Fig. 6(g).



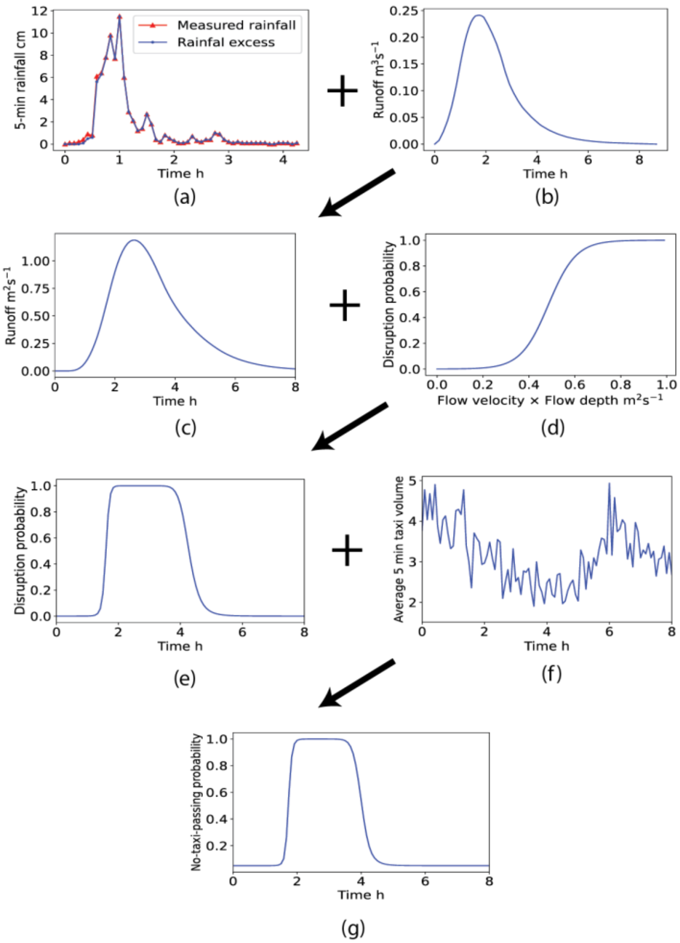


Figure 6 ~~An e~~Example transformation of ~~how the~~ a rainfall time series ~~is transformed~~ into ~~the~~ no-taxi-passing probabilities~~y~~ using the three-step procedure for the ~~1170th~~ 3,120th parameter set:~~.~~ **(a)** ~~T~~time series of rainfall and rainfall excess,~~.~~ **(b)** SUH constructed using the 3,120th~~1170th~~ parameter set,~~.~~ **(c)** ~~D~~derived runoff,~~.~~ **(d)** ~~E~~empirical runoff-disruption function,~~.~~ **(e)** ~~D~~derived time series of disruption probabilities~~y.~~, **(f)** ~~Disruption no-taxi-passing probability function~~ average taxi volume in 5 min intervals,~~.~~ and **(g)** ~~D~~derived no-taxi-passing probabilities~~y~~.

After the time series of no-taxi-passing probabilities~~y~~ for every parameter set were~~is~~ derived, ~~we can calculate~~ the degree of belief that a given parameter set is optimal was calculated by integrating it with the taxi observations on ~~9~~ May 9, 2015. According to Eq. (5), the posterior probability of the ~~1,170~~3,120th parameter set is calculated as:

$$\cancel{P(\theta^{(1170)}|X) \propto P(\Omega^{(1170)})\mathscr{L}(X \mid \Omega^{(1170)}) = \frac{1}{3000}\prod_{t=1}^{T}(1-\omega_t^{(1170)})^{X_t}(\omega_t^{(1170)})^{1-X_t}}$$

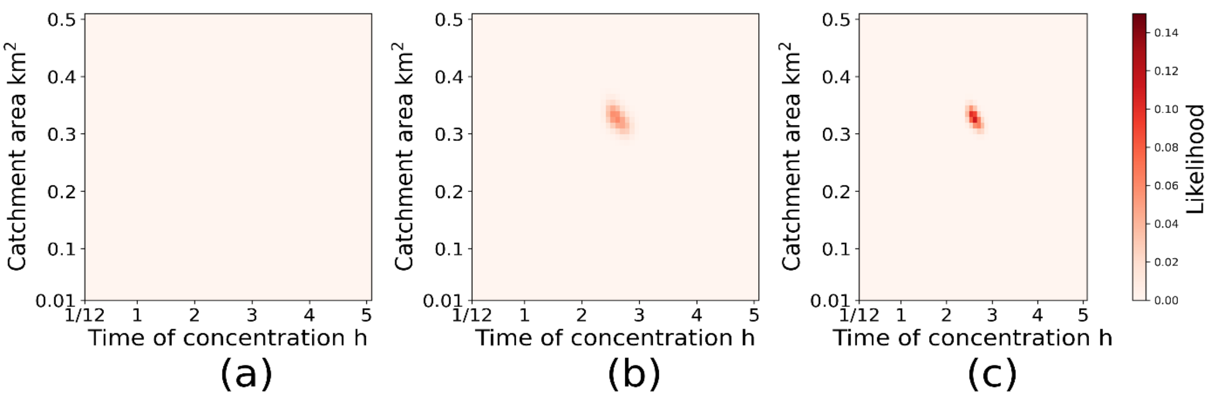
$$P\left(\boldsymbol{\theta}^{(3120)}|\boldsymbol{X}\right) \propto L\left(\boldsymbol{X}|\boldsymbol{\theta}^{(3120)}\right)P\left(\boldsymbol{\theta}^{(3120)}\right) = 1/4800 \times \prod_{t=1}^{T}\left(1-\omega_t^{(3120)}\right)^{x_t}\left(\omega_t^{(3120)}\right)^{1-x_t}$$

where ~~$\mathscr{L}(\theta^{(1170)}|X)$~~ $\mathcal{L}\left(\boldsymbol{\theta}^{(\mathbf{3120})}\middle|\boldsymbol{X}\right)$ is the ~~likelihood~~ posterior distribution of probabilities that the ~~1,170~~3,120th parameter set is optimal ~~conditioning~~ conditional on $\boldsymbol{X}$, which represents~~is~~ the taxi observations on ~~9~~ May 9, 2015 ~~shown~~ presented in Fig. 5(c). ~~The $P(\Omega^{(1170)})$~~ $P\left(\boldsymbol{\theta}^{(\mathbf{3120})}\right)$ is the prior probability of the ~~1,170~~3,120th parameter set ~~to~~ being optimal~~,~~ and its value~~s~~ is ~~1/3000~~1 / 4,800 because there are ~~3,000~~4,800 possible combinations.

By ~~F~~following ~~the above~~this process, we can calculate the posterior probabiliti~~y~~es for every parameter set. ~~Furthermore~~Additionally, the posterior probability distribution of a parameter set ~~could~~ can be updated using the taxi observations and rainfall data on ~~23~~ May 23, 2015 as:

$$P\left(\boldsymbol{\theta}^{(i)}\middle|\boldsymbol{X}_2\right) \propto \mathcal{L}\left(\boldsymbol{X}_2\middle|\boldsymbol{\theta}^{(i)}\right)P\left(\boldsymbol{\theta}^{(i)}\middle|\boldsymbol{X}_1\right)$$

where $P\left(\boldsymbol{\theta}^{(i)}\middle|\boldsymbol{X}_1\right)$ is the original posterior probability distribution ~~calculated~~ calibrated based on the storm on ~~9~~ May 9, 2015, and $P\left(\boldsymbol{\theta}^{(i)}\middle|\boldsymbol{X}_2\right)$ is the updated posterior distribution after the data of storm from~~on 23~~ May 23, 2015 are added. Fig. 7 illustrate~~d~~s the evolution of the probability distribution with the availability of ~~more~~ additional taxi data. The first row in Fig. 7 represents the prior joint distribution of hydrological parameter sets, and the second and third rows represent the posterior distribution after each round of updating. The posterior distribution dominates the uniform prior distribution after the first updat~~ing~~e, and the distribution is refined slightly ~~a little bit~~ after the second updat~~ing~~e.



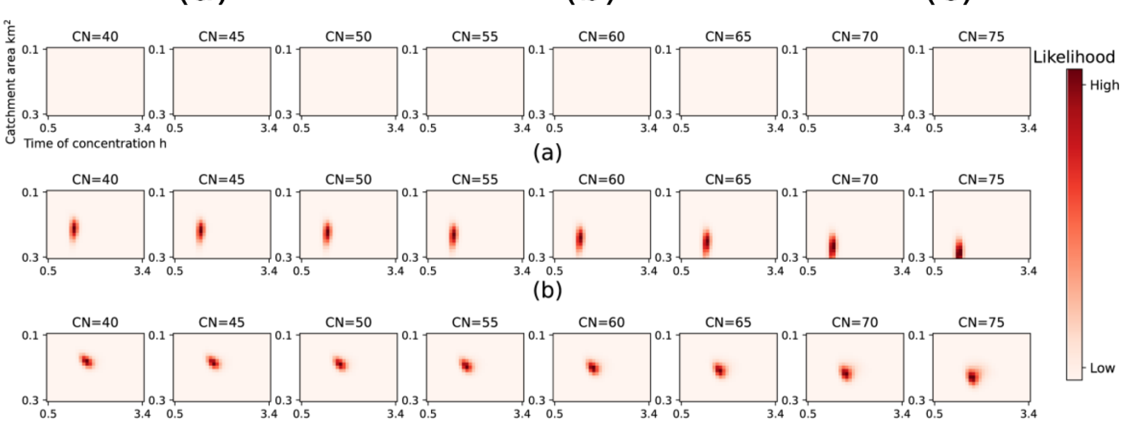


Figure 7 Evolution of the posterior probability distribution of hydrological parameter sets:~~.~~ **(a)** Prior distribution before updating,~~.~~ **(b)** Posterior distribution after the first updating,~~.~~ and **(c)** Posterior distribution after the second updating.

4 Method ~~V~~validation ~~and result~~

4.1 ~~Method validation~~Data description

The proposed method ~~i~~was validated ~~upon~~ on flood-prone roads located in Shenzhen, China, which is a coastal city frequently hit by extreme storms ~~in~~ during summer. ~~Another reason that Shenzhen is chosen is that only Shenzhen, as far as we known,~~ To the best of our knowledge, Shenzhen is the only city that has shared ~~the~~ runoff-related data ~~to~~ with the public in China. Three data sources, ~~which~~ namely ~~are~~ taxi GPS data, rainfall data, and authoritative water level data, were ~~are~~ used to validate ~~the~~ our parameter calibration method. Hydrological parameters ~~are~~ were calibrated using the first two data sources~~,~~ and the water level data acted~~s~~ as the ground truth to validate the proposed method. Taxi GPS data were~~are~~ anonymized and aggregated ~~to~~ in ~~the road every~~ 5 min intervals. ~~The r~~Rainfall data, which ~~are~~ were also collected ~~every~~ in 5 min intervals, ~~are~~ were measured at 115 gauging stations citywide~~,~~ and ~~are~~ mapped to the road network throughout Shenzhen using the ~~O~~ordinary Kriging spatial interpolation algorithm. The water level data ~~are~~ were only measured at ~~some~~ certain ~~waterlogging~~ flood-prone points~~,~~ with a dynamic sampling interval ranging from 5 min when the weather was rainy to 1 h when the weather was clear~~rainless~~. The proposed calibration method was validated by ~~checking~~ analyzing the hydrographs derived from the calibrated hydrological models ~~and~~ against the authoritative water levels for 10 selected roads. Detailed information ~~of~~ on the three data sources is provided~~are listed~~ in Table 4.

Table 4 Detailed information ~~of~~ on the three data sources.

Item	Taxi GPS data[1]	Rainfall data[1]	Water level data[2]
Source	Transport Commission of Shenzhen Municipality	Meteorological Bureau of Shenzhen Municipality	Shenzhen Municipal Government Data Open Platform[1]
Record	Taxi volume of each road	5 min accumulative rainfall	Water level
Data collection period	May 2015	2015 and 2019	2019
Data collection interval	5 min	5 min	5 min -1 h
Location	Citywide	115 rainfall gaging stations	171 flooding gaging sites

[1] The complete taxi GPS data and rainfall data are not openly accessible ~~due~~ owing to the requirements of data policy. To validate ~~the~~ our research findings, we have uploaded the necessary data ~~in~~ to Zenodo (Kong, 2022).

[2] Openly available at ~~the site~~: https://opendata.sz.gov.cn/data/dataSet/toDataDetails/29200_01403147.

The ~~T~~two storm events~~,~~ ~~occurred~~ on ~~9~~ May 9, 2015 and ~~23~~ May 23, 2015 ~~are~~ were treated as calibration events and ~~the~~ a storm ~~occurred~~ on ~~11~~ June 11, 2019, was~~is~~ retained for testing. ~~Obviously~~Clearly, there is a four~~4~~ year ~~span~~ gap between the calibration data and validation data ~~due to~~based on ~~the~~ data availability. The hydrological environments of flood-prone roads may have changed during these years, which could render the parameters calibrated based on data from 2015 inaccurate for analysis in 2019. To reduce the validation error caused by this~~the~~ time ~~difference~~gap, the roads to be validated should have been vulnerable to flooding ~~on~~ in both 2015 and 2019 so that the hydrological parameters of these roads would have a higher chance ~~to~~ of remaining unchanged. Therefore, ~~in~~ a total of 10 flood-prone roads~~,~~ ~~which~~ that were labelled as ~~flood-prone roads~~such in ~~on~~ both the List of 2015 Flood-prone Roads in Shenzhen (Water Authority of Shenzhen Municipality, 2015) and the List of 2019 Flood-prone Roads in Shenzhen (Water Authority of Shenzhen Municipality, 2019), were carefully selected (Fig. 8).

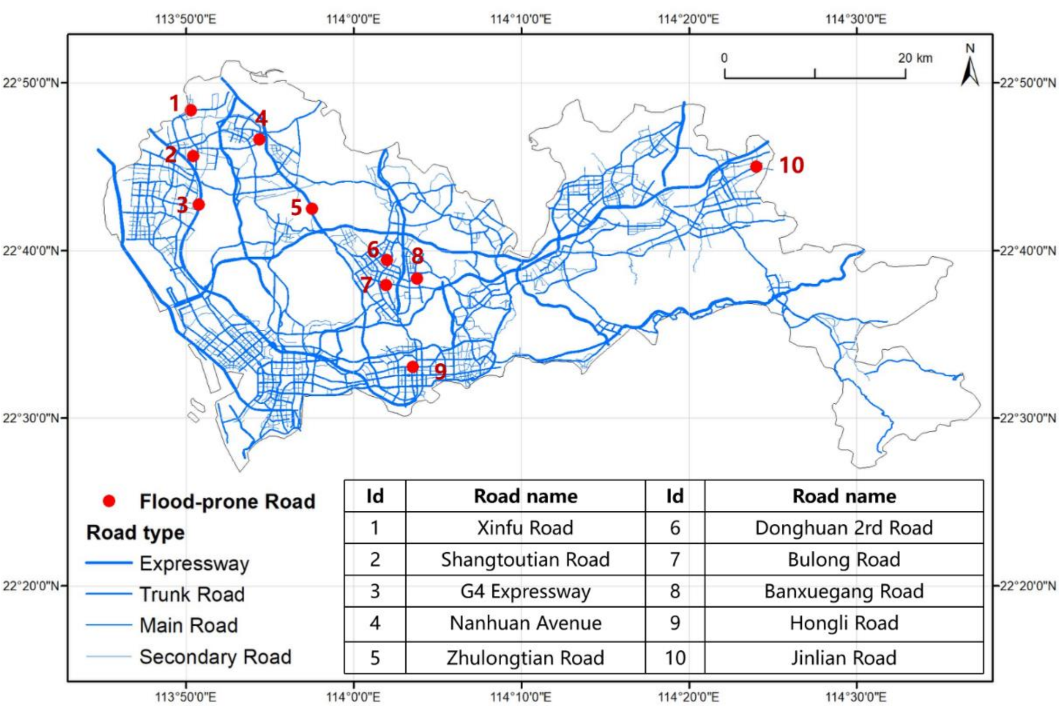
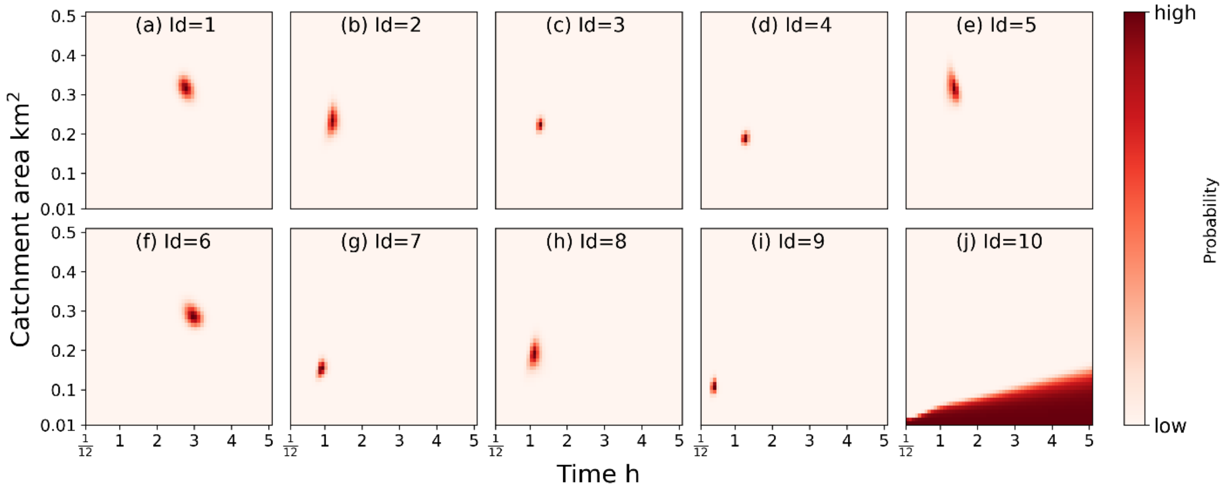


Figure 8 Spatial distribution of 10 flood-prone roads in Shenzhen.

~~Next, the posterior probability of parameter sets after calibration for the 10 roads are illustrated in Fig.9. As shown in Fig.9, the posterior probability distribution of parameter sets for most flood-prone roads are clustered around the optimal parameter set after two runs of updating, indicating that the uncertainty of parameter sets is refined to a much smaller area when taxi observations are added. It should also be noted that the posterior probability of parameter set for the Jinlian Road (Road ID=10) is evenly distributed on a triangular region (Fig.9j). By examining the taxi data of the road, we found that the taxi volume was greater than 0 for most 5-min intervals during two storms, indicating that the road was not disrupted during two storms. As hydrological parameters are calibrated by adjusting their values such that the runoff generated by the acceptable parameter sets could yield the disruption period during which no taxi points are observed, the lack of no-taxi-passing period would provide less information for calibration compared with when no-taxi-passing period is observed. This explains why the posterior probability is not refined to a small-area domain. However, we can still get some valuable information from Fig.9j. First, the catchment area for the Jinlian Road should not be too large to generate the runoff which may cause the road disruption.~~

~~Second, the catchment area is highly intercorrelated with time of concentration. As the catchment area gets larger, the time of concentration is more likely to increase so that the high runoff volume could not converge in a short time.~~



~~**Figure 9** Posterior probability distribution of hydrological parameter sets after the first updating for 10 flood-prone roads. Subplots (a)-(j) represent the probability distribution for Road 1-10.~~

4.2 Prior distributions of calibrated parameters

We introduced two types of prior distributions to demonstrate the effects of prior distributions on calibrated parameters. The first prior distribution was determined based on prior knowledge and DEMs from Shenzhen, which were obtained from ASTER GDEM V3, which is a product of NASA and Japan's Ministry of Economy, Trade, and Industry (METI) (ASTER Global Digital Elevation Map, 2023). This global DEM covers the entire land surface of the earth with a 30 m resolution, exhibiting notable improvements in horizontal and vertical accuracy while reducing anomalies compared to previous versions. We inputted the DEMs from Shenzhen into the hydrological software PCSWMM to delineate catchments and calculate the catchment area. Subsequently, we computed the time of concentration using the watershed lag method (Natural Resources Conservation Service, 2010b). As suggested by Zhang and Huang (2018), we used the average curve number for Shenzhen in 2015, which was assessed to be 60, as the estimated curve number for each road under validation.

We then constructed a discretized parameter space for the three parameters for each road as follows. For the curve number, we examined eight possible values centered on 60 with steps of five. For the catchment area, we considered 20 possible values centered on the estimated value with steps of 0.01 km^2. For the time of concentration, we considered 30 possible values centered on the estimated value with steps of 5 min. After constructing the parameter space for the parameters, we assigned a triangular prior distribution to each, which assumed the maximum probability at the estimated value and linearly decreased to zero at the parameter space boundaries, as depicted in Fig. 9.

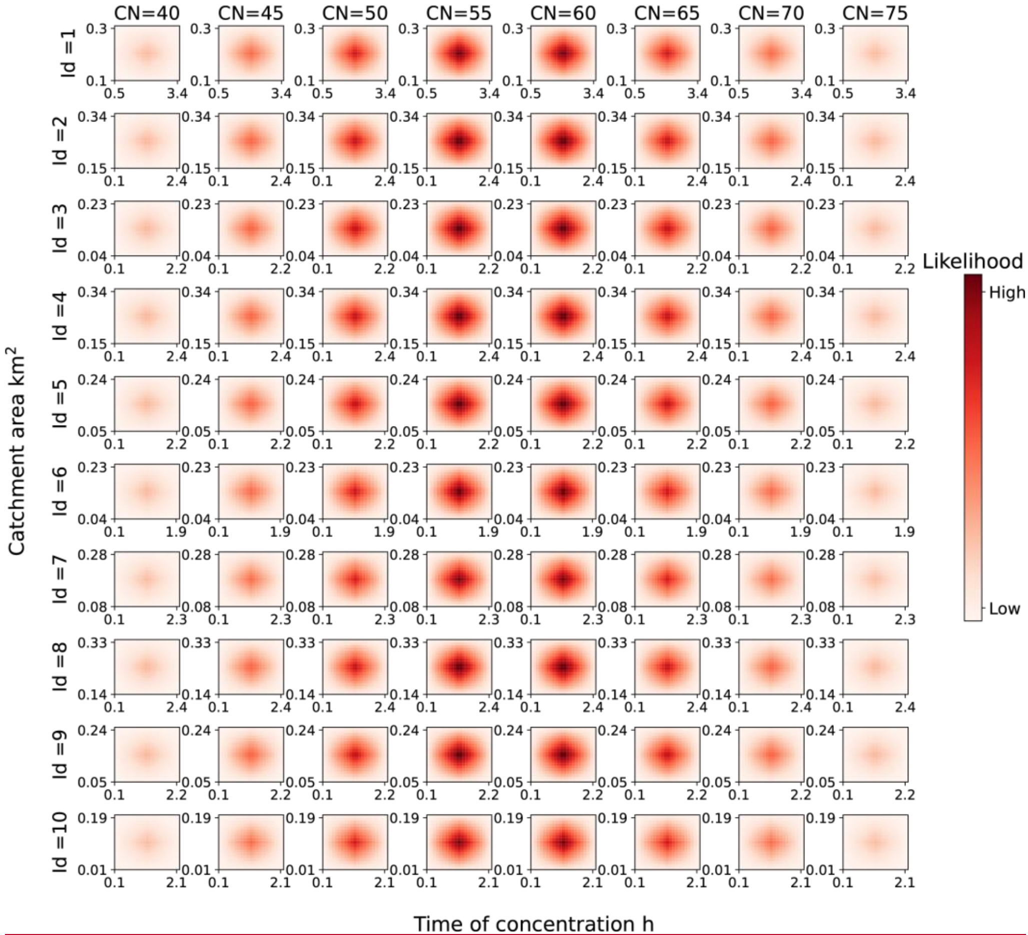


Figure 9 Prior probability distributions of hydrological parameter sets based on DEMs and other prior knowledge for 10 flood-prone roads.

The second prior distribution assumed that the three parameters all follow uniform distributions. The parameter spaces for the second prior distribution were the same as those for the first. As a result, the joint probability of each parameter set was equal to (1 / 20) × (1 / 30) × (1 / 8). To facilitate comparisons, we present the detailed information on the two types of prior distributions in Table 4.

Table 4 Detailed information of the two types of prior distributions.

Item	**Prior probability distributions based on DEM and other prior knowledge**			**Uniform distributions**		
	Number of possible values	Parameter interval	Prior marginal distribution	Number of possible values	Parameter interval	Prior marginal distribution

Curve number	8	5	Maximum probability at 60 and linearly reduces to zero at the parameter space boundaries	8	5	1/8 for each possible value
Catchment area	20	0.01	Maximum probability at the estimated value and linearly reduces to zero at the parameter space boundaries	20	0.01	1/20 for each possible value
Time of concentration	30	1/12		30	1/12	1/30 for each possible value

4.3 Posterior distributions after calibration

We first calibrated the parameters based on the prior distributions calculated according to the DEMs and other prior knowledge. The resulting posterior distributions are presented in Fig. 10. Each row in Fig. 10 represents a different road, and each column represents a curve number. Each subplot presents the joint probability distribution of the catchment area and time of concentration for a given curve number. The color intensity in Fig. 10 represents the magnitude of the probabilities. Following two iterations of updating, the posterior probability distributions for both the catchment area and time of concentration converge around the optimal parameter sets for most flood-prone roads. This demonstrates that incorporating taxi observations significantly reduces the uncertainty associated with catchment area and time of concentration. The probability typically achieves its maximum value when the curve number is either 55 or 60. Furthermore, each subplot contains a salient cluster with higher probability than other regions, suggesting that there may be multiple acceptable parameter sets.

Furthermore, the optimal catchment area under a given curve number decreases as the curve number increases, whereas the optimal time of concentration under a given curve number increases with the curve number. This is logical, because a higher curve number corresponds to increased rainfall excess under identical rainfall conditions, requiring a reduction in catchment area to maintain the runoff that best aligns with the taxi observations. Similarly, an increase in the time of concentration diminishes the peak runoff produced by the additional runoff generated by a higher curve number, thereby preserving the optimal runoff status.

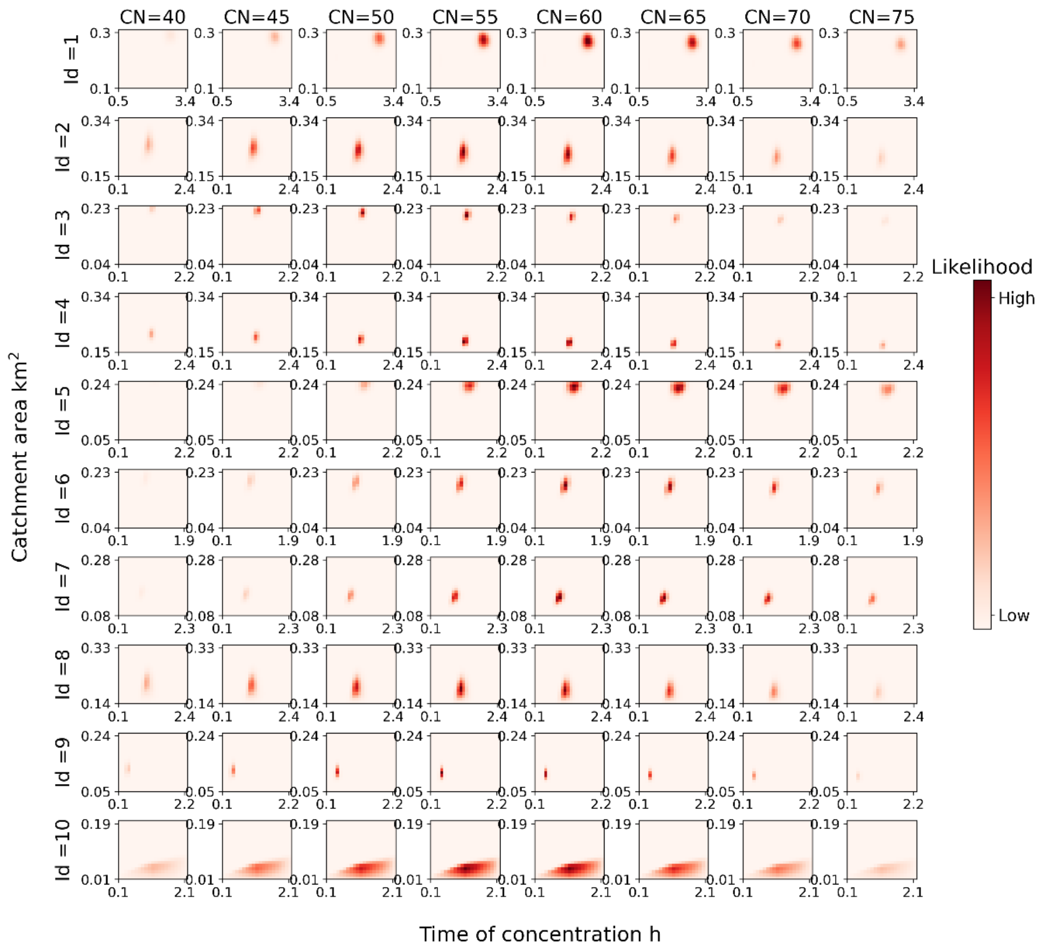


Figure 10 Posterior probability distributions of hydrological parameter sets for 10 flood-prone roads after calibration. The prior probability distributions were derived from the DEMs and additional prior knowledge.

We also present the marginal distributions of the three parameters for 10 roads before and after calibration in Fig. 11. In Fig. 11, the marginal posterior distributions of the curve number appear relatively similar to the marginal prior distributions. It seems that the proposed method employing taxi data provides limited information regarding the distribution of curve numbers compared to the catchment area and time of concentration. This outcome may be a result of the range and discretization granularity of the parameter spaces. Catchment area and time of concentration encompass 20 and 30 possible values, respectively, whereas the curve number has only 8 potential values. The smaller parameter space of the curve number reduces the search space, and its impact on the no-taxi-passing probability is comparatively lower than that of the catchment area and time of concentration.

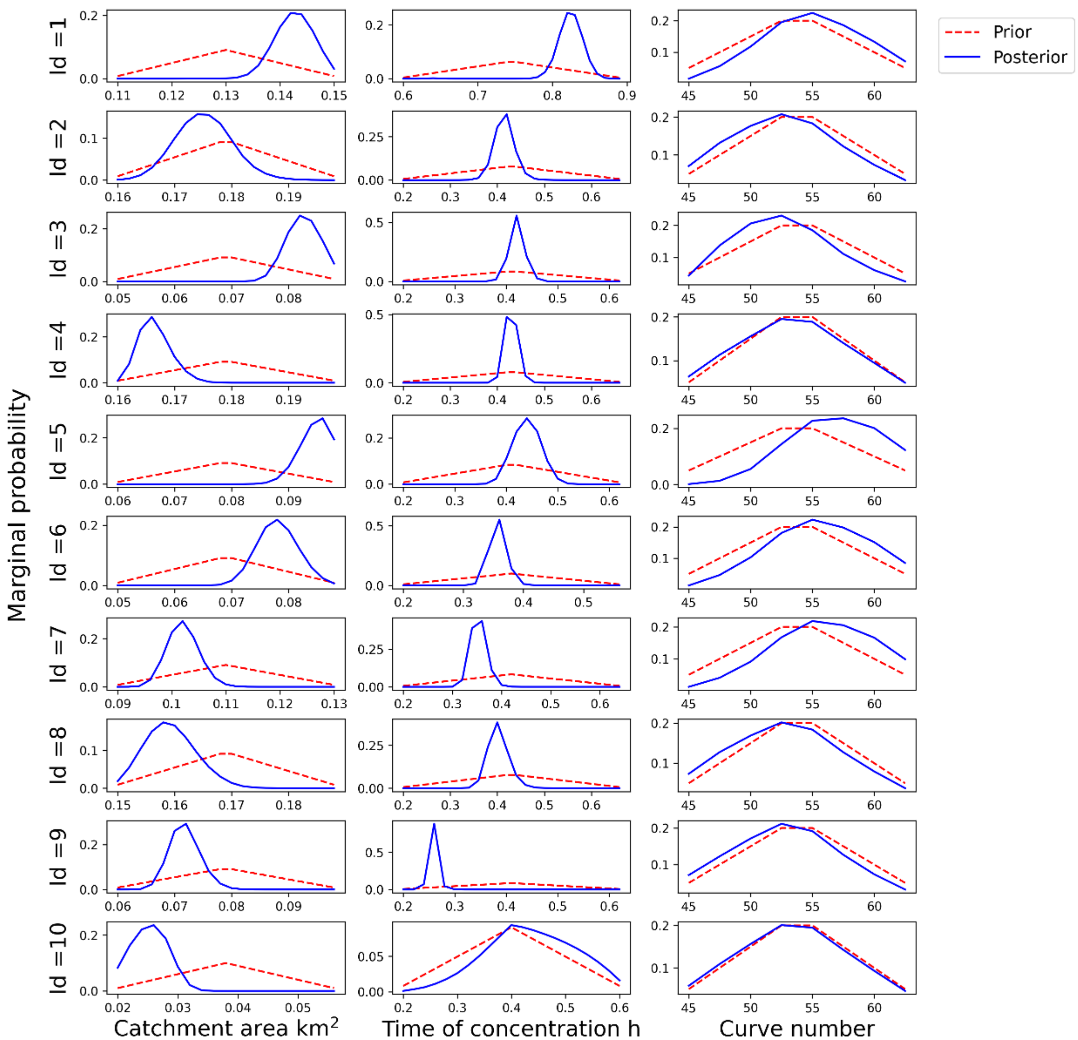


Figure 11 Marginal prior and posterior probability distributions of the curve number for 10 flood-prone roads.

For example, for road ID = 6, the optimal parameter set consists of a catchment area of 0.19 km², time of concentration of 0.9 h, and curve number of 55. To investigate the effects of these parameters on the hydrograph and time series of no-taxi-passing probabilities, we held two parameters constant at their optimal values and observed the impact of changing the third parameter. Our findings are presented in Fig. 12. One can see that when the catchment area varies from 0.04 to 0.23 km², the maximum no-taxi-passing probability increases from 20% to 100% and the duration for which the no-taxi-passing probability exceeds 0.5 increases from 0.0 to 1.3 h. Similarly, when the time of concentration fluctuates from 0.1 to 1.9 h, the peak time of the no-taxi-passing probability varies from 0.5 to 1.8 h. In contrast, when the curve number varies from 40 to 75, the maximum no-taxi-passing probability is fixed at 100%, the duration for which the no-taxi-passing probability exceeds 0.5 extends from 1.1 to 1.3 h, and the peak time of the no-taxi-passing probability remains fixed at the 1.1 h. These small fluctuations in the time series of no-taxi-passing probabilities are representative of why the distribution of curve numbers remains relatively stable after calibration compared to the catchment area and time of concentration.

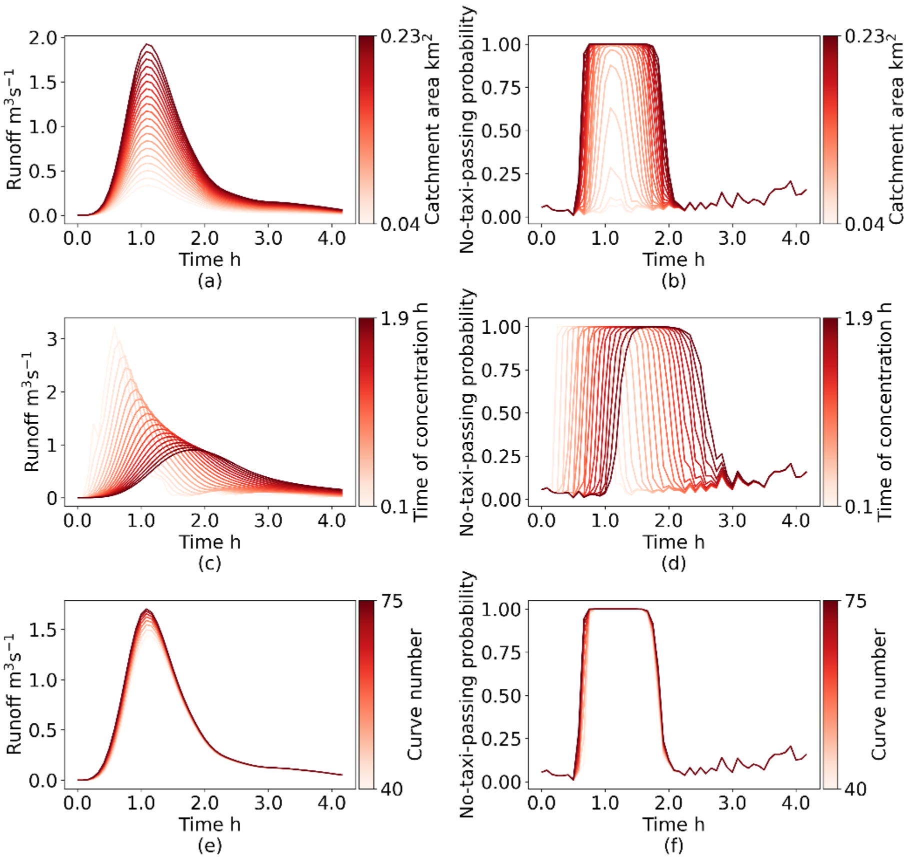


Figure 12 Impacts of three parameters on the variation of the time series of runoff and no-taxi-passing probabilities: (a) catchment area conditional on runoff, (b) catchment area conditional on the no-taxi-passing probability, (c) time of concentration conditional on runoff, (d) time of concentration conditional on the no-taxi-passing probability, (e) curve number conditional on runoff, and (f) curve number conditional on the no-taxi-passing probability.

The posterior distributions calibrated based on the uniform prior distribution are presented in Fig. 13. When comparing two posterior distributions derived from two prior distributions, it is clear that the posterior distributions of the catchment area and time of concentration are very similar, indicating that the impact of prior distributions on these parameters rapidly diminishes after taxi-related knowledge is added. As stated by Beven and Binley (1992 pp: 286), "as soon as information is added in terms of comparisons between observed and predicted responses then, if this information has value, the distribution of calculated likelihood values should dominate the uniform prior distribution when uncertainty estimates are recalculated."

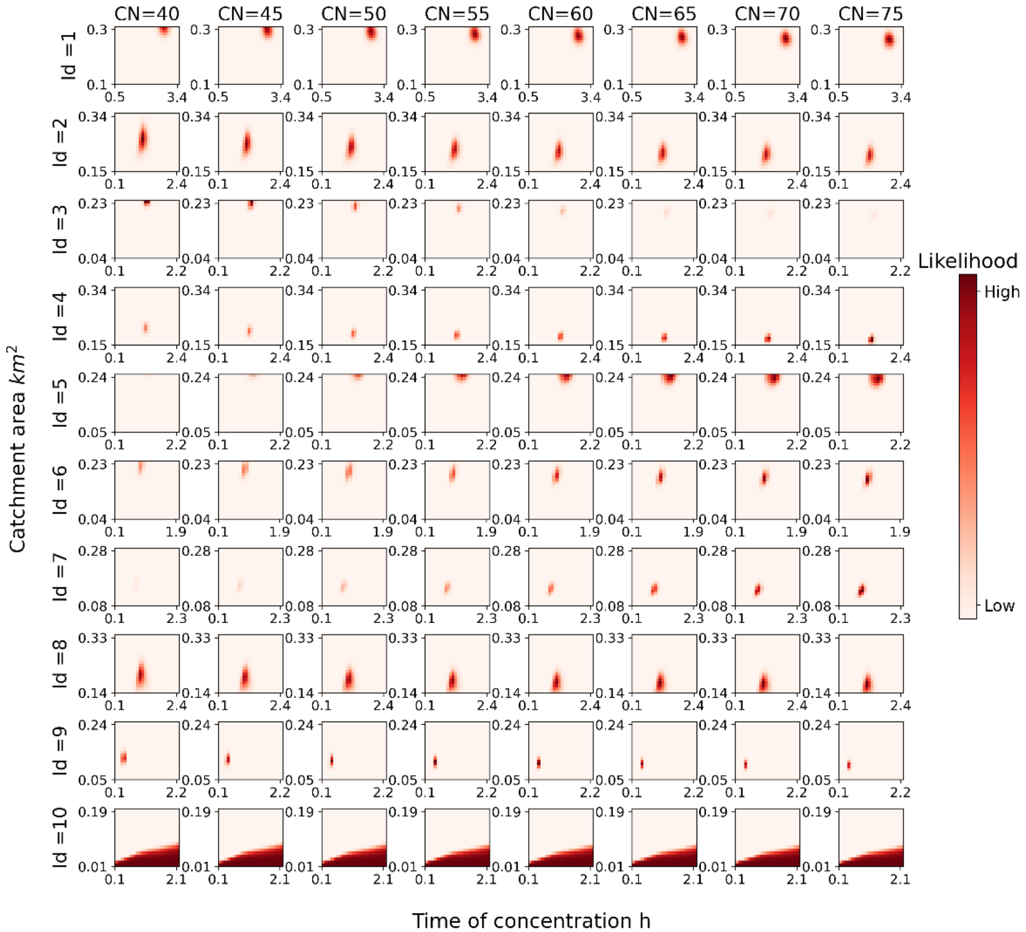


Figure 13 Posterior probability distributions of hydrological parameter sets for 10 flood-prone roads after calibration. The prior probability distributions were derived from a uniform distribution.

4.4 Validation results

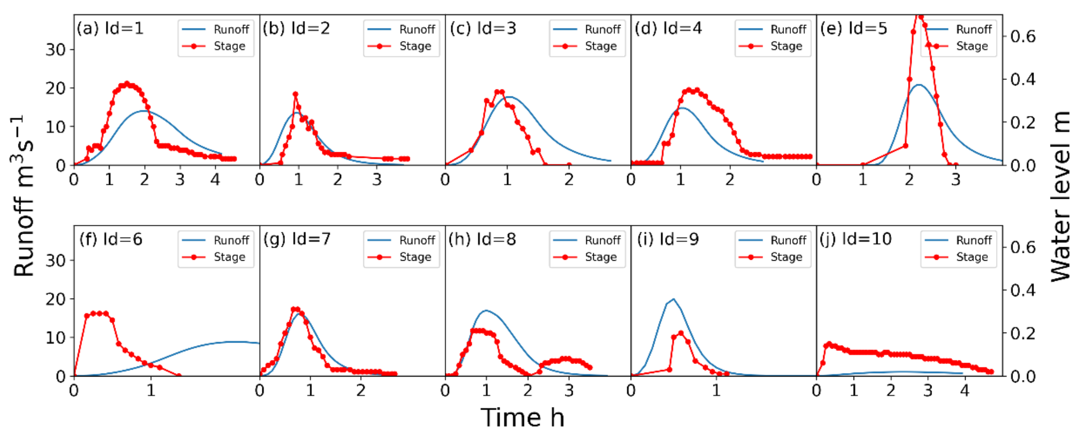
After the parameter sets were calibrated, they were combined with ~~the~~ an SCS unit hydrograph to construct ~~the~~ an SUH, which ~~were~~ was ~~further~~ combined with the rainfall data ~~occurring~~ from ~~on 11~~ June 11, 2019 to produce the predicted hydrograph. ~~As~~ Because the posterior probability associated with each parameter set can be regarded as a fuzzy measure reflecting the degree of belief that the parameter set is true, the weighted runoff~~s~~ values for each parameter set were summed to ~~produce~~ calculate the final predicted runoff:

$$Q = \sum_{i=1}^{N} P\left(\theta^{(i)} \middle| X\right) Q^{(i)} \tag{17}$$

~~where~~ Here, Q is the final predicted runoff, $Q^{(i)}$ is the simulated runoff derived from the ith parameter set, and $P\left(\theta^{(i)}|X\right)$ is the posterior probability of the ith parameter set, which ~~acting~~ acts as ~~the~~ a weight.

The output of the calibrated hydrological model is runoff (with ~~the~~ units of m^3 s^{-1}), whereas the validation data are~~is~~ water level data (with ~~the~~ units of m). Because~~As~~ the calibration data and validation data ~~arise~~ came from multiple sources and have different units, conventional error-based statistics such as the mean absolute error ~~(MAE)~~were~~are~~ not suitable for~~in~~ this study. ~~Most often, t~~The discharge of a stream is rarely measured directly. Instead, streamflow is typically determined by converting measured water depth (i.e., water stage) into discharge through a rating curve, which provides a functional relationship between the water stage and discharge at a specified point (Le Coz et al., 2014). Inspired by the application of the rating curve, we validated ~~the~~ our method by ~~developing the rating curve for every road, and then estimate~~ estimating the goodness-of-fit between the water level which was measured in the field and the corresponding runoff which was predicted based on the proposed calibration method.~~of those rating curves.~~ A higher goodness of fit indicates synchronous trends between the runoff and water level, which indirectly demonstrates the feasibility of the proposed method.

Because the posterior distributions derived from the two types of prior distributions were very similar, we only considered the posterior distribution calibrated based on prior distributions derived from DEMs and other prior knowledge for validation. Comparisons between the observed water depth and ~~the~~ simulated runoff for 10 selected roads are ~~shown~~ presented in Fig. 14~~0~~, and ~~rating~~ corresponding ~~curves~~scatter plots ~~constructed by fitting the runoff-stage scatter plot~~ are ~~shown~~ presented in Fig. 15~~1~~. We use the Pearson correlation coefficient, which measures the linear correlation between two variables, as ~~the~~ a goodness-of-fit indicator. ~~The result shows~~One can see that 8 of 10 roads ~~have rating curves with~~ are characterized by significant positive Pearson coefficients, indicating that the runoff and water have similar and consistent variation ~~process~~trends.



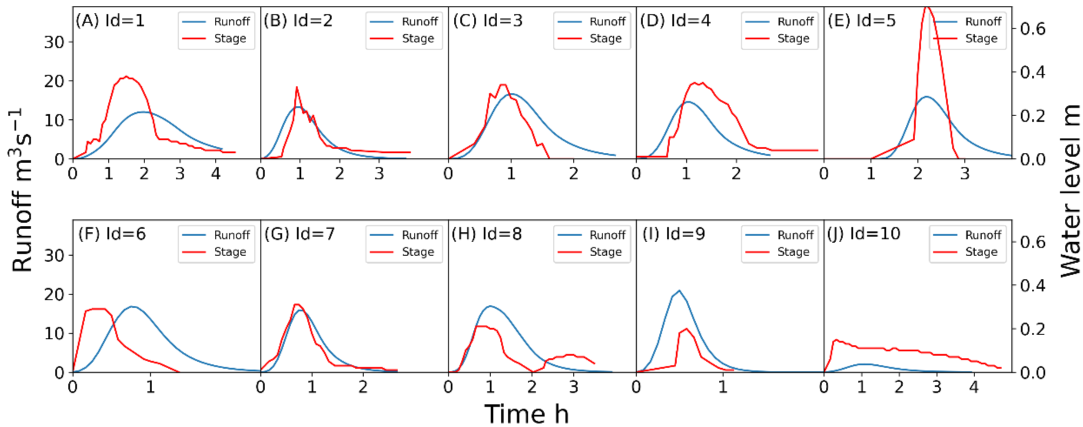
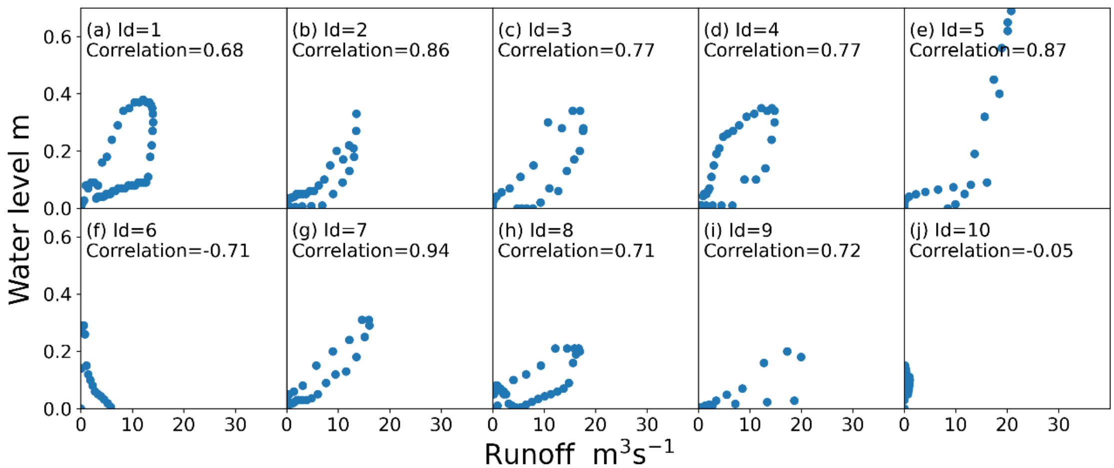


Figure ~~10~~ 14 Comparisons between the observed water depth and ~~the~~ simulated runoff for r~~R~~oads 1 to~~-~~10. The maximum value is 30 m^3 s^{-1} ~~of~~ on the left *y*-axis (i.e., runoff) and 0.6 m ~~of~~ on the right *y*-axis (i.e., stage) for ~~every~~ each subplot.



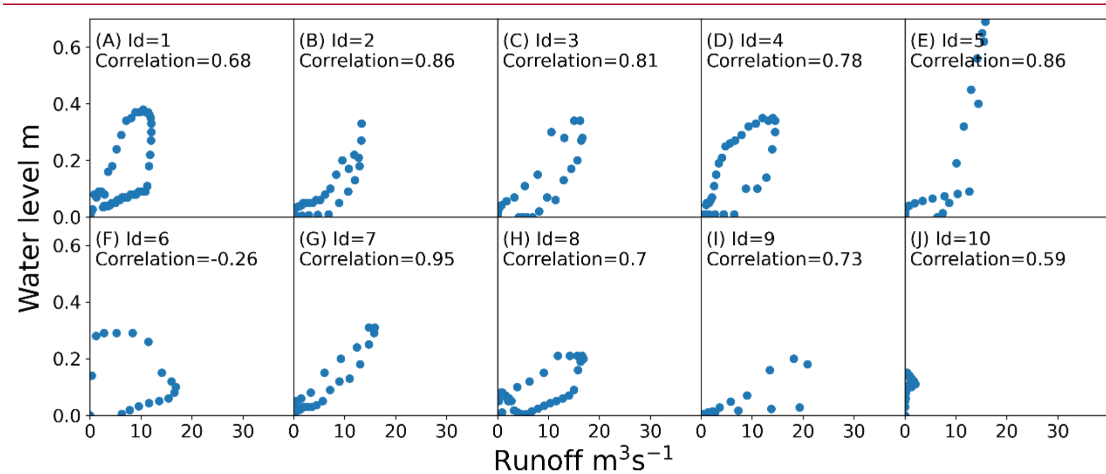


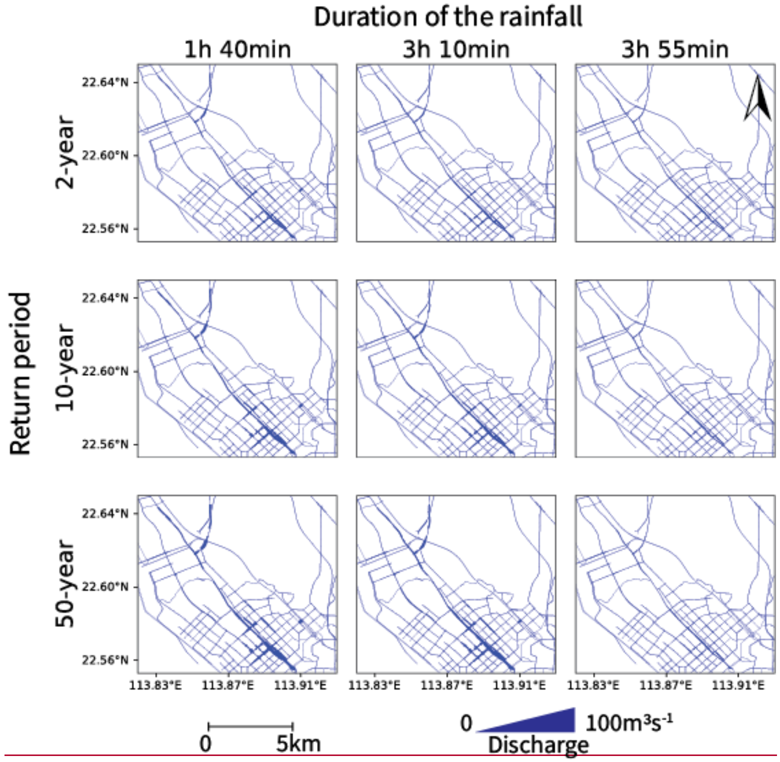
Figure ~~11~~ 15 Scatter plots ~~(a-j)~~ of the observed water depth and the simulated runoff for r~~R~~oads 1-10.

It is ~~worth ~~noteworthy~~ing~~ that goodness-~~-~~of-~~-~~fit ~~solely ~~simply describes the degree of correlation between the observed and simulated data, and may contain validation bias. As suggested by Legates and McCabe (1999), correlation-based statistics are~~is~~ insensitive to additive and proportional differences between ~~the ~~simulations and observations. Therefore, the fitting of a rating curve is only a partial~~reveals part of the~~ validation~~ truth,~~ and the usefulness of the proposed calibration method ~~needs~~ requires further ~~inspection~~analysis.

~~4.2 Application of the method to plot flooding maps in Shenzhen~~

~~Based on the proposed calibration method, we simulated how the road network experiences flooding for different rainfall return periods. Three storm events of different return periods (*T* = 2, 10, and 50 years) were designed according to the Rainfall Intensity Formula of Shenzhen (Meteorological Bureau of Shenzhen Municipality, 2015). Each storm lasts 3 hours, with an accumulative rainfall amount of 159 mm, 230 mm, and 283 mm for the 2, 10, and 50 year return period. Hydrological parameters of high-level flood-prone roads, including expressway, main road, and secondary road, are calibrated using the taxi data on 9 May 2015 and 23 May 2015. The flood-prone roads are identified based on the algorithm proposed in our previous studies (Kong et al., 2022). The road discharges under different rainfall return periods are simulated by inputting the designed rainfall to the calibrated hydrologic models. As an example, Fig.12 shows the spatio-temporal evolution of simulated discharge of parts of the road network, which locate in Baoan District, Shenzhen, for different return periods. With the return periods rising from 2 year to 50 year, the average peak discharge for flood-prone roads increases by 80.6%, with the value from 13.9 $m^3 \cdot s^{-1}$ to 25.1 $m^3 \cdot s^{-1}$. Inputting the simulated runoff to the empirical runoff-disruption function, expressed in Eq.(12), the time series of disruption probability for every road could be derived[3]. To facilitate discussion, we temporarily define the disruption period as the time when disruption probability is higher than 0.5. The average disruption period for flood-prone roads increases from 1.67 h to 3.15 h as the return period increases from 2-year to 50 year.~~

~~[3] Original runoff should be divided by the road width before inputting to the empirical runoff-disruption function.~~



~~**Figure 12** Spatio-temporal evolution of simulated runoff for different return periods in Baoan District, Shenzhen.~~

5 Discussion

Three main points ~~are worth discussing~~ about the proposed calibration method are worthy of further discussion. The first is that~~,~~ although the presented validation results support the use of taxi GPS data to calibrate hydrological parameters for poorly gauged road networks, the proposed method is more applicable to ~~the~~ roads ~~which~~ that are~~is~~ frequently visited by taxis. Uncertainty increases as the taxi volume ~~of~~ on a road decreases. A road is considered to be passable when at least one taxi GPS point~~s~~ is~~are~~ observed during ~~the~~ a time interval, ~~while~~ but we cannot assert that ~~the~~ a road is disrupted when the taxi volume is zero. When a road with frequent taxi~~s~~ traffic~~frequently passing by~~ is observed with no taxi GPS points during ~~the~~ a storm, it is highly probable that the road is disrupted by ~~the~~ flooding, which provides relatively reliable information for parameter calibration. Conversely, when a road with ~~few~~ little taxi~~s~~ ~~visiting~~ traffic has no taxi points during ~~the~~ a storm, there is a ~~great~~ relatively ~~chance~~ high likelihood that the road remains passable and is simply exhibiting its typical trend of~~just has~~ no taxis~~ as usual~~. Therefore, ~~The~~the proposed calibration method ~~thus~~ becomes relatively unreliable ~~considering that~~ when ~~the~~ a “no-taxi-passing period” is no longer a good proxy for~~of~~ the “disruption period~~” for the taxi-data-sparse road~~ on a road with sparse taxi data. To compensate for ~~the~~ a shortage of ~~the~~ taxi GPS data, ~~extra~~ additional data sources~~,~~ such as ride-hailing data and bus data~~,~~ should be incorporated in ~~the~~ future work.

Second~~ly~~, the disruption of one road may cause cascading failures, ~~so that~~where the disruption is~~may be~~ rapidly propagated~~ing~~ from the inundated road to ~~the~~ adjacent non-inundated roads under the constraints of ~~the~~ road connectivity. For a road ~~which~~ that is disrupted, but not inundated by ~~the~~ a storm, the implementation of the proposed calibration method may be subject to structural errors. ~~Assume~~ Consider ~~there are~~ two connected roads called~~, namely~~ Road 1 and Road 2~~,~~ ~~which~~ that are both disrupted during a storm~~,~~ and have taxi volumes ~~of two roads are therefore~~ of zero (Fig. ~~13~~ 16). ~~The difference lies in that~~ In this case, Road 1 is disrupted by the flooding, ~~while~~ whereas Road 2 is only disrupted ~~due to~~ because it is connected~~ing~~ to ~~the disrupted road, i.e.~~ Road 1. If taxi data are~~is~~ the only data ~~source~~ used for calibration, then the posterior distributions of the hydrological parameters for Road 1 and Road 2 ~~should~~ will be identical after calibration, because the sequences of taxi volume are identical for both roads. ~~Clearly~~However, we know that the hydrological parameters ~~for~~ of these two roads ~~could~~ are not ~~be~~ the same, because only one road is flooded~~otherwise Road 1 and Road 2 should be both flooded~~. Just like we cannot simply treat the "no-taxi-passing period" as the "disruption period", we cannot confuse the "disruption period" with the "flooded period." In ~~the~~ future work, an algorithm that facilitates ~~enabling to~~ distinguishing the flooding-induced disruption ~~and the~~from connectivity-induced disruption should be developed.

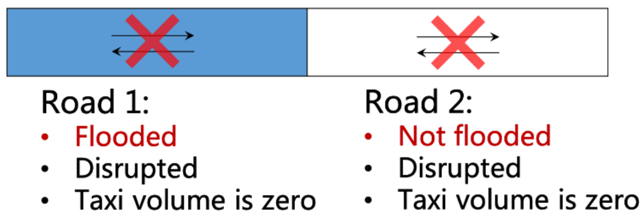


Figure ~~13~~ 16 ~~A g~~Graphical representation ~~to show~~highlighting the difference between the "disruption period" and flooded period~~the "no-taxi-passing period."~~

Third~~ly~~, the ~~specific~~ proposed three-step process, which consists of ~~the~~ an SCS unit hydrograph, ~~the~~ empirical runoff-disruption function, and ~~the~~ Poisson distribution, is~~performs as~~ a realization of the generalized framework ~~shown~~ presented in Fig. 1. The s~~S~~ub-models used in~~of~~ the three-step process ~~are not deterministic, and~~ can be flexibly ~~substituted~~ replaced with ~~by~~ other sub-models according to ~~the needed~~ complexity requirements and data availability. For example, an alternative to the SCS unit hydrograph is the distributed hydrological model. Unlike~~Compared with~~ the SCS unit hydrograph, the distributed hydrological model partitions a watershed into physically homogeneous units and captures the complex spatial variation induced by human activity in high resolution, which may be more applicable to ~~the~~ urbanized environments, such as ~~the~~ road networks. However, considering that some critical data such as ~~including the~~ road drainage data and land use data are missing, as well as the ~~calibration procedure will become~~ extreme~~ly~~ computational~~ly intensive~~ cost associated with the distributed hydrological model, we did not ~~use the distributed hydrological model~~ adopt this model in this study. Another assumption we made in this study is that the number of taxis arriving at a road follows a Poisson distribution. By conducting the Chi-square goodness of fit test, we found that the frequency distribution of taxi volumes adheres to a Poisson distribution for more than 50% of 5 min intervals for 7 of the 10 roads presented in Fig. 8, indicating that the Poisson model appears to be a suitable assumption. However, this hypothesis may not be universally applicable, particularly in different urban contexts, where alternate distributions such as the Weibull distribution may provide a more accurate representation.

6 Conclusion

An urban flooding model requires various types of data for calibration. In this study, we proposed a Bayesian calibration framework for the hydrological parameters of ~~the~~ a road network based on ~~the~~ taxi GPS data. A three-step procedure~~,~~ consisting of a rainfall-runoff model, ~~a~~ runoff-disruption ~~model~~function, and ~~a disruption~~ no-taxi-passing probability model~~,~~ ~~enables~~ enabled us to transform ~~the~~ a given rainfall time series into ~~the~~ a time series of no-taxi-passing probabilities~~y~~ for each parameter set, which is key to ~~the~~ taxi-data-driven model calibration. The calculated no-taxi-passing probabilities~~y~~, which ~~acting~~ acted as a proxy for~~of~~ the associated hydrological parameter sets, ~~is further~~ were compared ~~with the~~to observed taxi data ~~through~~ based on the Bayes equation to assess the posterior probability distributions of the hydrological parameter sets. Three parameters, namely the curve number, catchment area, and time of concentration, were calibrated. The proposed calibration method was~~is~~ instantiated by combining ~~some~~ classical hydrological models ~~and~~ with traffic flow models~~,~~ and ~~is~~ was validated on 10 flood-prone roads in Shenzhen. The validation results ~~show~~ indicate that the trends of runoff could be correctly predicted for eight~~8~~ roads, which demonstrates the potential of calibrating hydrological parameters based on taxi GPS data ~~indicating a good performance for hydrological parameter calibration~~.

This study ~~illustrates~~ highlights the ~~great~~ potential of integrating transportation-related data with hydrological theory for the~~in~~ transportation resilience improvement and flood risk management ~~for~~ of ~~the~~ road networks. We hope that our study can provide~~s~~ a flexible calibration framework for countries ~~which~~ that ~~are short of~~have little runoff data, but rich ~~of~~ taxi data. We ~~accept~~ acknowledge that the application of the proposed method is currently limited by the heterogeneous spatial distributions of taxis citywide and ~~the~~ cascading effects of road inundation, but expect this to change with the increasing availability of vehicle data and continuous~~ly~~ optimization of model~~l~~ing approaches.

Code and data availability

The data and code used to validate the proposed method are available on~~at~~ Zenodo (https://doi.org/10.5281/zenodo.7894921~~https://doi.org/10.5281/zenodo.7294849~~).

Author contributions

JY ~~conceptualised~~ conceptualized the article and collected ~~the~~ field data. XK designed the methodology and was responsible for ~~the~~ code compilation. KX plotted the figures and revised the manuscript. BD ~~manged~~ managed the implementation of research activities. SJ discussed ~~the~~ results and contributed to ~~the~~ method validation. XK wrote the final version of the article with contributions from all co-authors.

Competing interests

The ~~contact~~ corresponding author ~~has declared~~ declares that none of the authors have~~s~~ any competing interests.

Acknowledgments

This research was supported by the National Key Research and Development Program of China (~~g~~Grant number No. 2022YFC3303100).

References

Al‐Qadami, E. H. H., Mustaffa, Z., Al‐Atroush, M. E., Martinez‐Gomariz, E., Teo, F. Y., and El‐Husseini, Y.: A numerical approach to understand the responses of passenger vehicles moving through floodwaters, J. Flood~~.~~ Risk~~.~~ Manag., 15,~~,~~ https://doi.org/10.1111/jfr3.12828, 2022.

Balistrocchi, M., Metulini, R., Carpita, M., and Ranzi, R.: Dynamic maps of human exposure to floods based on mobile phone data, Nat. Hazards~~.~~ Earth~~.~~ Syst. Sci., 20, 3485–3500, https://doi.org/10.5194/nhess-20-3485-2020, 2020.

Beven, K.: Rainfall-~~R~~runoff ~~M~~modelling: The ~~P~~primer, 2nd ~~Edition~~ed., 2nd ed., Wiley-Blackwell, 2012.

Beven, K. and Binley, A.: The future of distributed models: Model calibration and uncertainty prediction, Hydrol. Process.~~,~~, 6, 279–298,~~,~~ https://doi.org/10.1002/hyp.3360060305, 1992.

Brouwer, T.~~ and~~, Eilander, D., van Loenen, A., Booij, M. J., Wijnberg, K. M., Verkade, J. S., Wagemaker, J.: Probabilistic flood extent estimates from social media flood observations, Nat. Hazards~~.~~ Earth~~.~~ Syst. Sci., 17, 735–747, https://doi.org/10.5194/nhess-17-735-2017, 2017.

Chow, V. T., Maidment, D. R., and Mays, L. W.: Applied ~~H~~hydrology, McGraw-Hill Book Company, 1988.

Contreras-Jara, M., Echaveguren, T., Vargas Baecheler, J., Chamorro Giné, A., and de Solminihac Tampier, H.: Reliability-based estimation of traffic interruption probability due to road waterlogging, J. Adv. Transp~~ort~~.~~,~~, 2018, 1–12,~~,~~ https://doi.org/10.1155/2018/2850546, 2018.

Gebremedhin, E. T., Basco‐Carrera, L., Jonoski, A., Iliffe, M., and Winsemius, H.: Crowdsourcing and interactive modelling for urban flood management, J. Flood~~.~~ Risk~~.~~ Manag., 13,~~,~~ https://doi.org/10.1111/jfr3.12602, 2020.

Goodwin, N. R., Coops, N. C., Tooke, T. R., Christen, A., and Voogt, J. A.: Characterizing urban surface cover and structure with airborne lidar technology, Can~~adian~~. J~~ournal~~. ~~of~~ Remote Sens~~ing~~., 35, 297–309,~~,~~ https://doi.org/10.5589/m09-015, 2009.

Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, Water Resour. Res., 34, 751–763.~~,~~ https://doi.org/10.1029/97WR03495, 1998.

Huang, M. and Jin, S.: A methodology for simple 2-D inundation analysis in urban area using SWMM and GIS, Nat. Hazards, 97, 15–43.~~,~~ https://doi.org/10.1007/s11069-019-03623-2, 2019.

Ji, S. and Qiuwen, Z.: A GIS-based ~~S~~subcatchments ~~D~~division ~~A~~approach for SWMM~~, TOCIEJ~~, TOCIEJ, 9, 515–521.~~,~~ https://doi.org/10.2174/1874149501509010515, 2015.

Kasmalkar, I. G., Serafin, K. A., Miao, Y., Bick, I. A., Ortolano, L., Ouyang, D., and Suckale, J.: When floods hit the road: ~~r~~Resilience to flood-related traffic disruption in the San Francisco Bay Area and beyond, Sci. Adv., 6, eaba2423.~~,~~ https://doi.org/10.1126/sciadv.aba2423, 2020.

Kong, X.: Data and code used in the article titled "~~-~~A Bayesian updating framework for calibrating hydrological parameters of road network using taxi GPS data," Zenodo [dataset].~~,~~ https://zenodo.org/record/7294880#.Y2nLv3ZByUk, 2022.

Kong, X., Yang, J., Qiu, J., Zhang, Q., Chen, X., Wang, M., and Jiang, S.: Post-event flood mapping for road networks using taxi GPS data, J. Flood~~.~~ Risk~~.~~ Manag., 15, e12799.~~,~~ https://doi.org/10.1111/jfr3.12799, 2022.

Kramer, M., Terheiden, K., and Wieprecht, S.: Safety criteria for the trafficability of inundated roads in urban floodings, Int. J. Disast~~.~~, Risk ,~~.~~ Reduct~~.~~, 17, 77–84.~~,~~ https://doi.org/10.1016/j.ijdrr.2016.04.003, 2016.

Le Coz, J., Renard, B., Bonnifait, L., Branger, F., and Le Boursicaud, R.: Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: A Bayesian approach, J. Hydrol., 509, 573–587.~~,~~ https://doi.org/10.1016/j.jhydrol.2013.11.016, 2014.

Legates, D. R. and McCabe, G. J.: Evaluating the use of "goodness-of-fit" Measures in hydrologic and hydroclimatic model validation, Water Resour. Res., 35, 233–241.~~,~~ https://doi.org/10.1029/1998WR900018, 1999.

Li, C., Fan, Z., Wu, Z., Dai, Z., Liu, L., and Zhang, C.: Methodology of ~~S~~sub-~~C~~catchment ~~D~~division ~~C~~considering ~~L~~land ~~U~~uses and ~~F~~flow ~~D~~directions, IJGI, 9, 634.~~,~~ https://doi.org/10.3390/ijgi9110634, 2020.

Li, Z., Wang, C., Emrich, C. T., and Guo, D.: A novel approach to leveraging social media for rapid flood mapping: ~~a~~A case study of the 2015 South Carolina floods, Cartogr. Geogr. Inf. Sci., 45, 97–110.~~,~~ https://doi.org/10.1080/15230406.2016.1271356, 2018.

Martínez-Gomariz, E., Gómez, M., Russo, B., and Djordjević, S.: A new experiments-based methodology to define the stability threshold for any vehicle exposed to flooding, Urban~~.~~ Water~~.~~ J., 14, 930–939.~~,~~ https://doi.org/10.1080/1573062X.2017.1301501, 2017.

Merz, B. and Thieken, A. H.: Flood risk curves and uncertainty bounds, Nat. Hazards~~.~~, 51, 437–458.~~,~~ https://doi.org/10.1007/s11069-009-9452-6, 2009.

Moore, K. A. and Power, R. K.: Safe buffer distances for offstream earth dams, Australas~~ian.~~. J~~ournal.~~. ~~of~~ Water Resour~~ces.~~., 6, 1–15~~,~~. https://doi.org/10.1080/13241583.2002.11465206, 2002.

ASTER ~~G~~global ~~D~~digital ~~E~~elevation ~~M~~map~~:~~. https://asterweb.jpl.nasa.gov/gdem.asp, last access: 5 May 2023.

Natural Resources Conservation Service: Chapter 16: Hydrographs, in: National ~~E~~engineering ~~H~~handbook, 3, United States Department of Agriculture~~, 3~~, 2007.

Natural Resources Conservation Service~~: Hydrology National Engineering Handbook~~ Chapter 9: Hydrologic ~~S~~soil-~~C~~cover ~~C~~complexes, in: Hydrology National engineering handbook, United States Department of Agriculture, 2010a.

Natural Resources Conservation Service~~: Hydrology National Engineering Handbook~~ Chapter 15: Time of ~~C~~concentration, in: Hydrology National engineering handbook, United States Department of Agriculture, 2010b.

Nieto, N., Chamorro, A., Echaveguren, T., Sáez, E., and González, A.: Development of fragility curves for road embankments exposed to perpendicular debris flows~~, Geomat, Nat~~, Geomatics Natural Hazards and Risk, ~~Haz, Risk., 12~~12, 1560–1583~~,~~, https://doi.org/10.1080/19475705.2021.1935330, 2021.

Paul, J. D., Buytaert, W., Allen, S., Ballesteros-Cánovas, J. A., Bhusal, J., Cieslik, K., Clark, J., Dugar, S., Hannah, D. M., Stoffel, M., Dewulf, A., Dhital, M. R., Liu, W., Nayaval, J. L., Neupane, B., Schiller, A., Smith, P. J., and Supper, R.: Citizen science for hydrological risk reduction and resilience building~~: Citizen science for hydrological risk reduction and resilience building, Wires.~~, Water~~.~~, 5, e1262~~,~~, https://doi.org/10.1002/wat2.1262, 2018.

Pregnolato, M., Ford, A., Wilkinson, S. M., and Dawson, R. J.: The impact of flooding on road transport: A depth-disruption function~~, Transport. Res. D-Tr. E.~~, Transp. Res. Part D: Transp. Env., 55, 67–81,~~,~~ https://doi.org/10.1016/j.trd.2017.06.020, 2017.

Qi, Y., Zheng, Z., and Jia, D.: Exploring the spatial-temporal relationship between rainfall and traffic flow: ~~a~~A case study of ~~b~~Brisbane, Australia~~, Sustainability~~, Sustainability, 12, 5596~~,~~, https://doi.org/10.3390/su12145596, 2020.

~~Restrepo-Estrada, C., Andrade, S. C., Abe, N., Fava, M. C., Mendiondo, E. M., and Albuquerque, J. P.~~Restrepo-Estrada, C., de Andrade, S. C., Abe, N., Fava, M. C., Mendiondo, E. M., de Albuquerque, J. P.: Geo-social media as a proxy for hydrometeorological data for streamflow estimation and to improve flood monitoring, Comput. Geosci., 111, 148–158, https://doi.org/10.1016/j.cageo.2017.10.010, 2018.

Sadler, J. M., Goodall, J. L., Morsy, M. M., and Spencer, K.: Modeling urban coastal flood severity from crowd-sourced flood reports using Poisson regression and Random Forest, J~~ournal~~. ~~of~~ Hydrol~~ogy~~., 559, 43–55,~~,~~ https://doi.org/10.1016/j.jhydrol.2018.01.044, 2018.

Safaei-Moghadam, A., Tarboton, D., and Minsker, B.: Estimating the likelihood of roadway pluvial flood based on crowdsourced traffic data and depression-based DEM analysis, Nat.~~ural~~ Haz.~~ards and~~ Earth Sys.~~tem~~ Sci.~~ences~~, ~~(~~preprint~~),~~. https://doi.org/10.5194/nhess-2022-77, 2022.

Shah, S. M. H., Mustaffa, Z., and Yusof, K. W.: Experimental Studies on the threshold of vehicle instability in floodwaters, J~~urnal~~. Teknol~~ogi~~., 80,~~,~~ https://doi.org/10.11113/jt.v80.11198, 2018.

Shand, T. D., Cox, R. J., Blacka, M. J., and Smith, G. P.: Australian rainfall and runoff project 10: ~~a~~Appropriate safety criteria for vehicles, Australian Rainfall & Runoff, 2016.

She, Z, Fang, Z, Zhou~~She, Zhong, Fang, Zheng, and Zhou: Extracting Flooded Roads by Fusing GPS Trajectories and Road Network, ISPRS~~: Extracting Flooded Roads by Fusing GPS Trajectories and Road Network~~.~~, Int. Geo-Inf., 8, 407,~~,~~ https://doi.org/10.3390/ijgi8090407, 2019.

Versini, P.-A., Gaume, E., and Andrieu, H.: Application of a distributed hydrological model to the design of a road inundation warning system for flash flood prone areas, Nat. Hazards~~.~~ Earth~~.~~ Syst. Sci., 10, 805–817,~~;~~ https://doi.org/10.5194/nhess-10-805-2010, 2010.

Water Authority of Shenzhen Municipality: List of 2015 ~~F~~flood-prone ~~R~~roads in Shenzhen~~,~~. http://swj.sz.gov.cn/ztzl/ndmsss/yldzl/zrrxxb/content/post_2918436.html, 2015.

Water Authority of Shenzhen Municipality: List of 2019 ~~F~~flood-prone ~~R~~roads in Shenzhen~~,~~. https://opendata.sz.gov.cn/data/dataSet/toDataDetails/29200_01403146, 2019.

Xia, J., Falconer, R. A., Xiao, X., and Wang, Y.: Criterion of vehicle stability in floodwaters based on theoretical and experimental studies, Nat. Hazards, 70, 1619–1630,~~;~~ https://doi.org/10.1007/s11069-013-0889-2, 2014.

Yabe, T., Tsubouchi, K., and Sekimoto, Y.: Fusion of ~~T~~terrain ~~I~~information and ~~M~~mobile ~~P~~phone ~~L~~location ~~D~~data for ~~F~~flood ~~A~~area ~~D~~detection in ~~R~~rural ~~A~~areas, in:~~ 2018~~ IEEE International Conference on Big Data (Big Data)~~, 2018 IEEE International Conference on Big Data (Big Data)~~, 2018, 881–890, 2018~~,~~. https://doi.org/10.1109/BigData.2018.8622156~~, 2018~~.

Yao, Y., Wu, D., Hong, Y., Chen, D., Liang, Z., Guan, Q., Liang, X., and Dai, L.: Analyzing the effects of rainfall on urban traffic-congestion bottlenecks~~, IEEE J. Sel.~~, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 13, 504–512,~~;~~ https://doi.org/10.1109/JSTARS.2020.2966591, 2020.

Yin, J., Yu, D., Yin, Z., Liu, M., and He, Q.: Evaluating the impact and risk of pluvial flash flood on intra-urban road network: A case study in the city center of Shanghai, China, J. Hydrol., 537, 138–145,~~;~~ https://doi.org/10.1016/j.jhydrol.2016.03.037, 2016.

Zahura, F. T., Goodall, J. L., Sadler, J. M., Shen, Y., Morsy, M. M., and Behl, M.: Training ~~M~~machine ~~L~~learning ~~S~~surrogate ~~M~~models ~~F~~from a ~~H~~high-~~F~~fidelity ~~P~~physics-~~B~~based ~~M~~model: Application for ~~R~~real-~~T~~time ~~S~~street-~~S~~scale ~~F~~flood ~~P~~prediction in an ~~U~~urban ~~C~~coastal ~~C~~community, Water Resour~~ces~~. Res~~earch~~., 56~~, e2019~~, WR027038, e2019,~~;~~ https://doi.org/10.1029/2019WR027038, 2020.

Zhang, T. and Huang, X.: Monitoring of ~~U~~urban ~~I~~impervious ~~S~~surfaces ~~U~~using ~~T~~time ~~S~~series of ~~H~~high-~~R~~resolution ~~R~~remote ~~S~~sensing ~~I~~images in ~~R~~rapidly ~~U~~urbanized ~~A~~areas: A ~~C~~case ~~S~~study of Shenzhen, IEEE J. Sel. Top. Appl. Earth Obs~~ervations~~. Remote Sens.~~ing~~, 11, 2692–2708,~~;~~ https://doi.org/10.1109/JSTARS.2018.2804440, 2018.

Zhang, W., Li, R., Shang, P., and Liu, H.: Impact ~~A~~analysis of ~~R~~rainfall on ~~T~~traffic ~~F~~flow ~~C~~characteristics in Beijing, Int. J. ITS Res., 17, 150–160,~~;~~ https://doi.org/10.1007/s13177-018-0162-x, 2019.