Summary

This manuscript presented a framework for calibrating a hydrologic model based on taxi data. The concept is quite clever, in my opinion. There of course is a need to calibrate hydrologic models and, at the same time, a general lack of data needed to calibrate. Using taxi data for the calibration is a neat idea and I think the authors did a good job showing the reader the feasibility of this. Overall, I think the manuscript is clear, well written, and technically sound. There are a few items I think should be addressed before being accepted for publication.

Response:

We express our gratitude to the reviewer for their insightful comments and suggestions, which will substantially improve the quality of our manuscript. Following careful consideration, we have amended the manuscript in accordance with your valuable comments. Our responses to your comments are provided below.

Major comments

1 - why are you calibrating time of concentration and catchment area? These are parameters that I would not typically see calibrated. It seems like you could estimate catchment area from a DEM. Similarly, there are many methods for estimating time of concentration from catchment characteristics. Because these two parameters are relatively reasonable to calculate/estimate, I'd like to understand the author's reasoning for calibrating them.

Response:

Thanks for your question. 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, due to challenges in accurately delineating urban catchments. First, 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, subsequently affecting runoff parameters (Goodwin et al., 2009). Unlike natural catchment, it is also difficult to identify explicit urban drainage systems and road slope directly from the topographic relief of the urban region. Furthermore, high densities of residential and commercial buildings obstruct flow paths and alter flow directions of stormwater runoff, complicating rainfall-runoff and overland flow processes in urban areas (Ji & Qiuwen, 2015).

Second, accurate urban catchment delineation necessitates high-resolution Digital Elevation Model (DEM), which are not always available in many regions. Oksanen and Sarjakoski (2005) demonstrated that automatic catchment delineation is highly sensitive to DEM errors, and uncertainty in DEMs determines the lower bound for catchment size that can be computed with sufficient accuracy. In many Chinese cities, high-resolution DEMs are considered confidential data and are generally inaccessible to non-governmental organizations. Consequently, using a low-resolution DEM may introduce substantial errors.

Due to these challenges, deriving accurate catchment area and time of concentration in urban areas is difficult. This study thus aims to provide an alternative method based on taxi GPS data to calibrate these parameters.

2 - why is a curve number of 85 used for every case? This seems pretty consequential since the CN could vary between catchments. Should this be a calibrated parameter?

Response:

Thank you for your suggestion. We acknowledge that fixing the curve number as 85 is not realistic as it is influenced by various factors in urban areas, such as impervious surface percentage and soil type. Therefore, we have revised the manuscript to include curve number as one of the parameters to be calibrated. In total, we calibrate three parameters: catchment area, time of concentration, and curve number.

Figure 1 presents the probability distributions of three parameters after calibration. Each row in Fig.1 represents a different road, and each column represents the curve number. Furthermore, each subplot shows the joint probability distribution of catchment area and time of concentration given the curve number. The depth of colour in Fig.1 represents the magnitude of probability. Following two iterations of updating, the posterior probability distribution for both catchment area and time of concentration converges around the optimal parameter sets for most flood-prone roads. This demonstrates that incorporating taxi observations has substantially narrowed the uncertainty associated with catchment area and time of concentration. The probability typically attains its maximum value when the curve number is either 55 or 60. Moreover, each subplot presents a salient cluster with higher probability than other regions, suggesting that there may be multiple parameter sets which can effectively represent the acceptable ones.

Furthermore, it has been observed that the optimal catchment area under a given curve number decreases as the curve number increases, while the optimal time of concentration under a given curve number rises in relation to the curve number. This is a logical observation, as 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 observed taxi-related road conditions. Likewise, the increase in time of concentration diminishes the peak runoff produced by the additional runoff generated by a higher curve number, thus preserving the optimal runoff status.

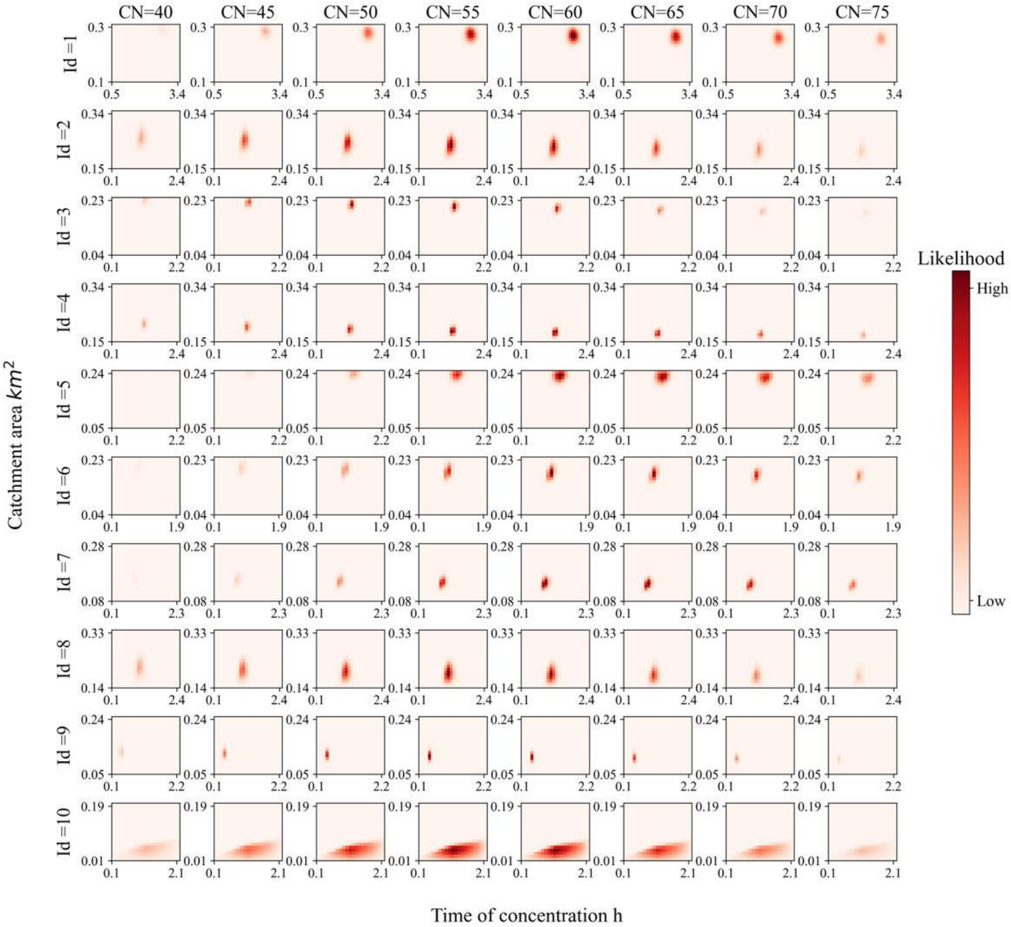

Figure 1 Posterior probability distributions of hydrological parameter sets for 10 flood-prone roads after calibration. The prior probability distributions are derived from the DEM and additional prior knowledge.

We also plotted the marginal distributions of the three parameters for ten roads before and after calibration in Fig. 2. Upon examining Fig. 2, the marginal posterior distributions of the curve number appear relatively similar to the marginal prior distributions. It seems that the method employing taxi data provides limited information about the distribution of curve numbers after calibration compared to catchment area and time of concentration. This outcome may be attributed to the range and discretization granularity of the parameter spaces. Catchment area and time of concentration encompass 20 and 30 possible values, respectively, while the curve number includes 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 catchment area and time of concentration.

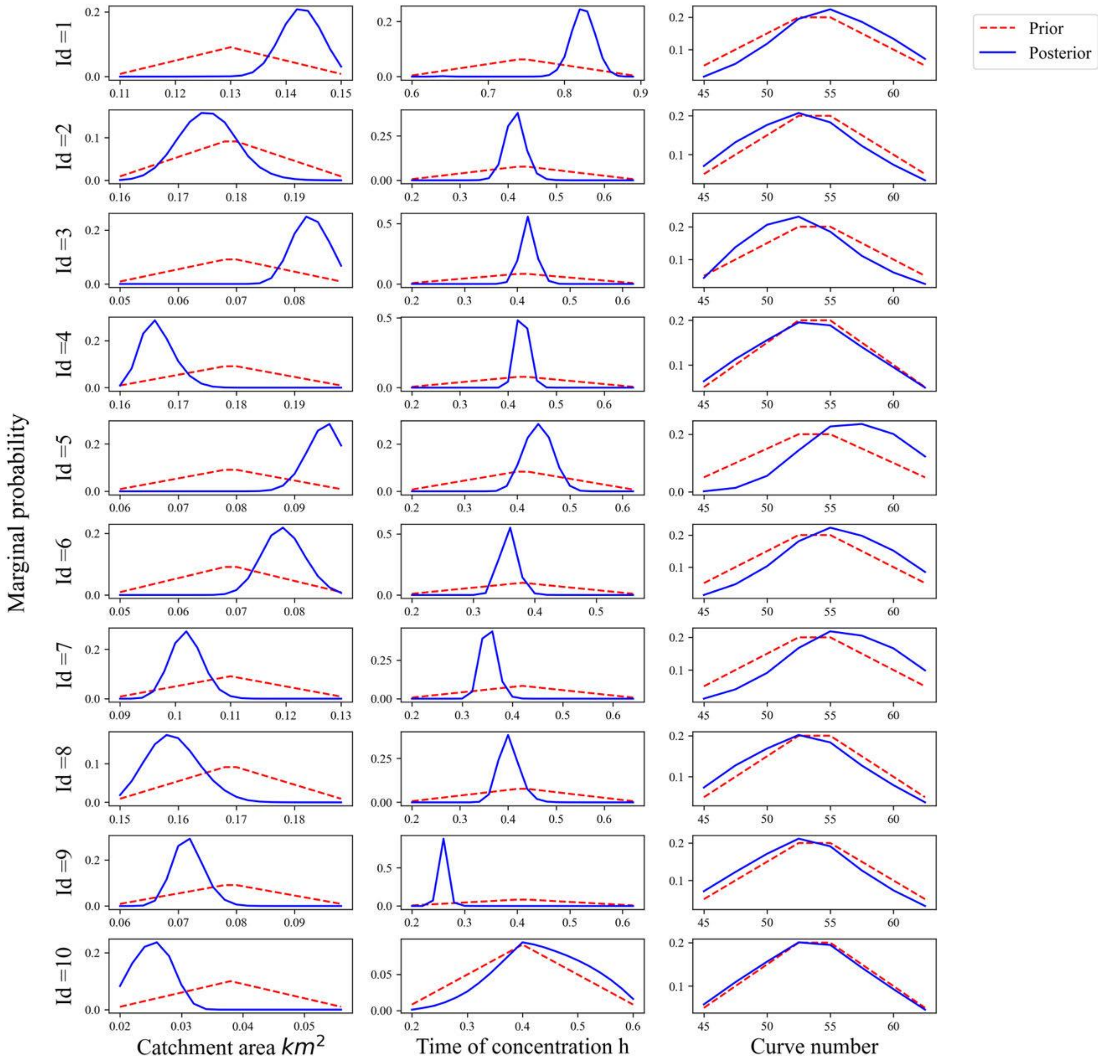

Figure 2 Marginal prior and posterior probability distribution of curve number for 10 flood-prone roads.

For instance, for road ID=6, the optimal parameter set maximizing the no-taxi-passing probability consists of a catchment area of 0.19 km², a time of concentration of 0.9 hour, and a curve number of 55. To investigate the effects of these parameters on the hydrograph and the time series of no-taxi-passing probability, we hold two parameters constant at their optimal values and observe the impact as the third parameter varied. Our findings, illustrated in Fig.3, demonstrate that when the catchment area varies from 0.04 km² to 0.23 km², the maximum no-taxi-passing probability ranges from 20% to 100%, and the duration of no-taxi-passing probability exceeding 0.5 hour extends from 0.0 to 1.3 hours. Similarly, when the time of concentration fluctuates from 0.1 to 1.9 hour, the peak time of no-taxi-passing probability spans from 0.5 to 1.8 hour. In contrast, when the curve number varies from 40 to 75, the maximum no-taxi-passing probability is fixed at 100%, the duration of no-taxi-passing probability excessing 0.5 hour extends from 1.1 to 1.3 hours, and the peak time of no-taxi-passing probability remains fixed at the 1.1 hour. The smaller fluctuations in the time series of no-taxi-passing probability interpret why the distribution of curve number remains relatively stable after calibration compared to catchment area and time of concentration.

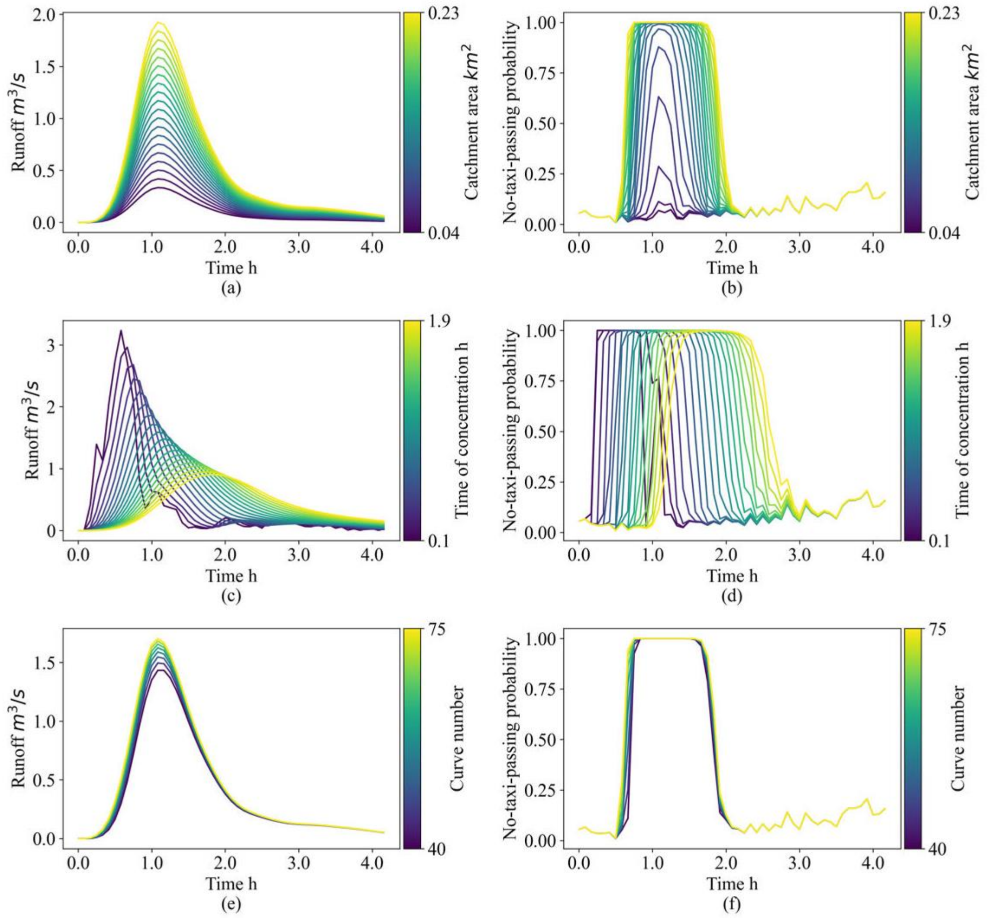

Figure 3 Impacts of three parameters on the variation of time series of runoff and no-taxi-passing probability. (a) Catchment area on the runoff. (b) Catchment area on the no-taxi-passing probability. (c) Time of concentration on the runoff. (d) Time of concentration on the no-taxi-passing probability (e) Curve number on the runoff. (f) Curve number on the no-taxi-passing probability

3 - Assuming it is reasonable to calibrate the catchment area and time of concentration, I question whether it's reasonable to have uniform priors for those parameters. Maybe you can't exactly know what the area of a catchment is going to be, but would you have enough of a guess to make a reasonable prior distribution? I'm guessing you'd know if a road segment has a relatively large or small catchment. Knowing this, it doesn't seem right to keep the prior distribution uniform.

Response:

In the original manuscript, we utilized only uniform priors for all parameters, leading to the inadequate use of prior knowledge, such as topography. In the revised manuscript, we introduce two types of prior distributions to demonstrate the effects of these distributions on calibrated parameters. The first prior distribution is determined based on the DEM of Shenzhen, obtained from ASTER GDEM V3, a product of NASA and Japan's Ministry of Economy, Trade, and Industry (METI) (Jet Propulsion Laboratory, 2019). This global digital elevation dataset covers the entire land surface of the earth with a 30-meter resolution and exhibits significant improvements in horizontal and vertical accuracy while reducing anomalies compared to previous versions. We input the DEM of Shenzhen into the hydrological software PCSWMM to delineate the catchment and calculate the catchment area. Subsequently, we compute the time of concentration using the watershed lag method (Natural Resources Conservation Service, 2010). According to Zhang and Huang (2018), the average curve number for Shenzhen in 2015 was 60, which we adopt as the estimated curve number for the 10 roads.

We construct the discretized parameter space for each road's three parameters as follows: for the catchment area, we consider 20 possible values centered on the estimated value with 0.01 km^2 intervals; for the time of concentration, we explore 30 possible values centered on the estimated value with 0.1 hour intervals; and for the curve number, we examine 8 possible values centered on 60 with intervals of 5. After constructing the parameter space for three parameters, we assign a triangular prior distribution to each, which assumes maximum probability at the estimated value and linearly reduces to zero at the parameter space boundaries, as depicted in Fig.4.

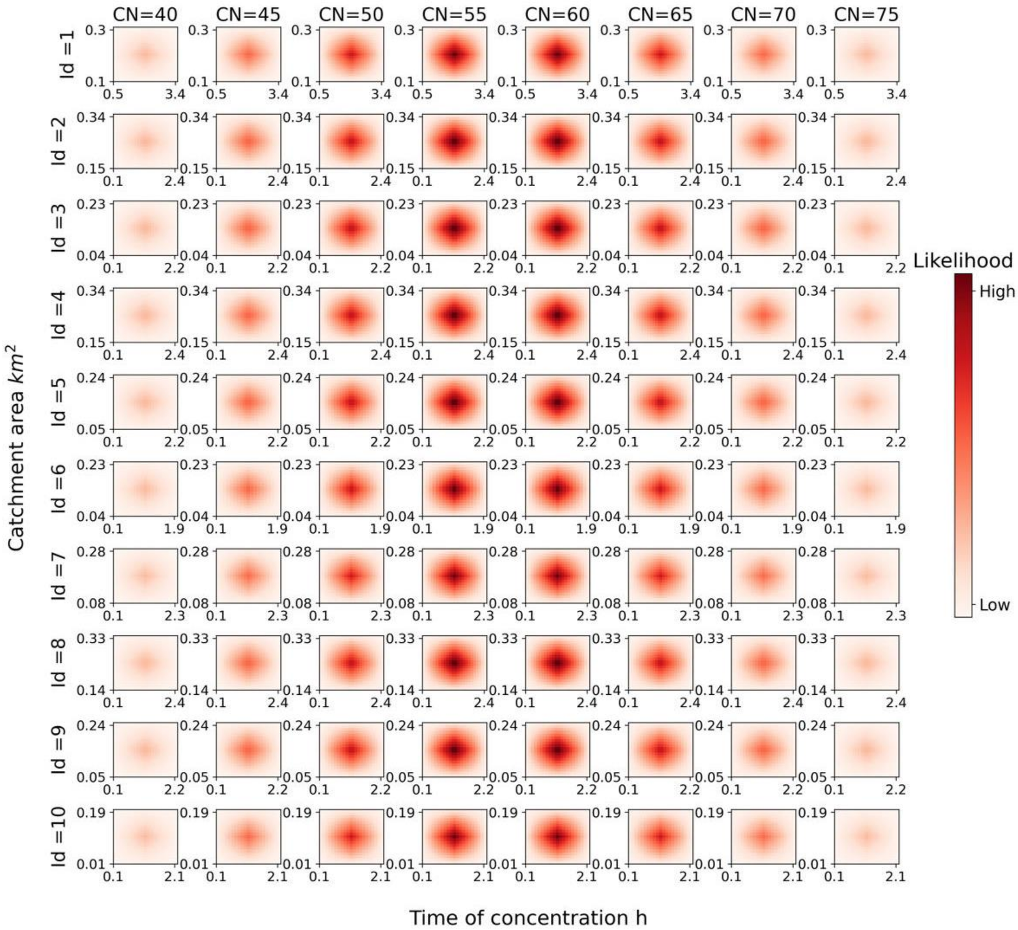

Figure 4 Prior probability distributions of hydrological parameter sets based on DEM and other prior knowledge for 10 flood-prone roads.

The second prior distribution assumes that the three parameters follow uniform distributions. The parameter space of the second prior distribution is the same as the first one. As a result, the joint probability of each parameter set equals 1/20 × 1/30 × 1/8. Figure 5 presents the posterior distributions calibrated based on the uniform prior distribution. By comparing two posterior distributions derived from two prior distributions, it is evident that the posterior distributions of catchment area and time of concentration are close to each other, 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.”

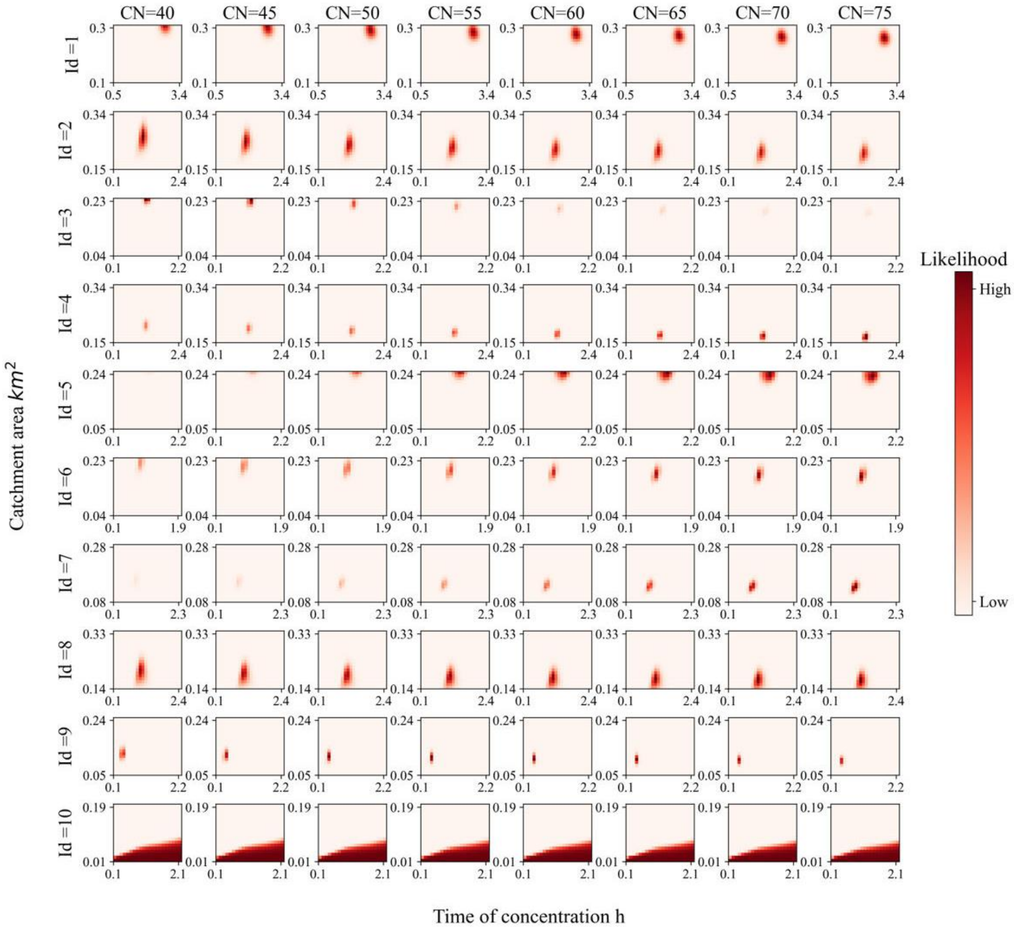

Figure 5 Posterior probability distributions of hydrological parameter sets for 10 flood-prone roads after calibration. The prior probability distributions are derived from the uniform distribution.

4 - It may be that I didn’t understand correctly, but how did you account for time of day/ day of week when considering whether or not a taxi would be passing? Or did you? For example, let’s say that at a given roadway segment, there is a day and time of the week that there are hardly any taxis. Can you take that into account in your calibration scheme so that a lack of taxis then does not suggest to the model that the roadway is flooded?

Response:

We appreciate your valuable suggestion. In the previous version of our manuscript, we did not account for variations in taxi volume concerning the time-of-day or day-of-week. We assumed that the average number of taxis arriving on the road was constant, and the no-taxi-passing probability is given by:

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

where λ is the average taxi volume per 5 min interval, calculated by averaging all 5 min taxi volumes using historical taxi GPS data for a specific road.

However, the value of λ fluctuates according to the time of day, exhibiting higher taxi volume during congested periods and lower volume during non-congested periods. In the revised version, we incorporated the time-of-day variation in taxi volume when computing the no-taxi-passing probability:

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

where λ_t is the 5 min taxi volume during the *t*th period, calculated by averaging the taxi volume of the *t*th period from May 1, 2015, to May 31, 2015. Compared with λ, λ_t has smaller deviance because it excludes more non-flooding factors.

Minor comments

- Figure 5 - does it make sense to have intermittent “have taxis” and "no taxi" times after a large rain event? I guess I’m just wondering at graph (C) in particular where it looks like there is just one taxi between 16:15-16:20. Does that mean that one taxi is just really willing to risk it and drive through the water? If it’s just one taxi, should it really be counted as "have taxi"?

Response:

This is a valid point. We also noticed that some drivers may take risks by driving through inundated roads, potentially resulting in intermittent “have-taxis” and “no-taxi” periods. We have examined the taxi volume between 16:15-16:20 and confirmed that one taxi drove through the water. Although the road appeared to be inundated and obstructed during this period, we would not categorize it as a theoretical “no-taxi” period. This is because our method determines the road’s status (“have-taxis” or “no-taxi”) based on the taxi volume, and the disruption period is inferred from the road’s status. In other words, we predict the flood period according to the road’s status, rather than vice versa. Furthermore, establishing an explicit rule to define “no-taxi” periods may cause confusion, as it implies that we have already observed the field data of flooding and constructed the rule based on it.

- Table 4 - if you had 171 flood gaging sites, why did you only pick 10 to test the model on? Why not test it on all 171?

Response:

The data used for parameter calibration were collected in 2015, while the data for method validation were collected in 2019, resulting in a four-year gap between the two datasets due to data availability. Furthermore, Shenzhen, as a coastal city, frequently experiences extreme storm events during summers. To mitigate flooding risks, the Shenzhen Municipal Government annually amends some flood-prone roads. As a result, the hydrological environment of certain roads may change over time, rendering parameters calibrated based on data in 2015 potentially inaccurate in 2019. To minimize validation errors caused by the time difference, we selected roads for validation that were vulnerable to flooding in both 2015 and 2019, increasing the likelihood that the hydrological parameters of these roads remained unchanged. Approximately 10 roads met this criterion. We will clarify this point in the revised manuscript.

- l355 - how did you make a rating curve for each road? How did you get the flow data to relate the stage data to?

Response:

The rating curve is usually determined by conducting field measurements and establish the relationship between the observed water level and the corresponding observed flow rate at a measuring point. In this study, however, we had no knowledge of empirical flow data for each road, thus we could not build a real rating curve. Instead, we establish a “rating curve” by plotting the water level which is field measured and the corresponding runoff which is predicted based on the proposed calibration method. If the derived “rating curve” is linearly related, indicating that the predicted runoff has the similar evolution trend of the observed water level, we thus assume that trends of runoff could be correctly predicted. To avoid confusion, we will not use the term “rating curve” to represent the relationship between the predicted runoff and the observed water level in the revised manuscript.

- Section 4.2 - I personally don’t think you need this section. While it’s interesting to see how you applied the framework, I don’t think it is needed. I think it is enough to have described (section 2), illustrated (section 3), and validated (section 4) the method.

Response:

Thank you for your insightful suggestion regarding the section in question. After careful consideration, we agree that the mentioned section may not be necessary for our paper. As you pointed out, the method has been sufficiently described in Section 2, illustrated in Section 3, and validated in Section 4. In response to your suggestion, we will remove the section to streamline the manuscript and maintain focus on the key aspects of our research. We believe this revision will enhance the overall clarity and concision of our paper.

- L54: You might consider citing the following since they are related to this topic (full disclosure: I am an author on both):

- Sadler, J. M., Goodall, J. L., Morsy, M. M., & Spencer, K. (2018). Modeling urban coastal flood severity from crowd-sourced flood reports using Poisson regression and Random Forest. Journal of hydrology, 559, 43-55.

- Zahura, F. T., Goodall, J. L., Sadler, J. M., Shen, Y., Morsy, M. M., & Behl, M. (2020). Training machine learning surrogate models from a high-fidelity physics-based model: Application for real-time street-scale flood prediction in an urban coastal community. Water Resources Research, 56, e2019WR027038. https://doi.org/10.1029/2019WR027038

Response:

Thank you for your suggestion. We will cite these articles in the introduction section to enhance our review of existing research:

Citizens voluntarily or passively acting as human sensors generate georeferenced data to improve flood monitoring. Typical studies involve the use of crowdsourcing social media data (Brouwer & 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).

- Figure 10: Could you explain why for some runoff values there is more than one level value? For empirically derived rating curves, each runoff value corresponds to only one water level.

Response:

As previously mentioned, the rating curve developed in our study relies on predicted runoff rather than observed runoff. Consequently, the temporal trends of the predicted results may not consistently align with those of the observed water levels. This discrepancy can result in one water level having two distinct runoff values. For instance, consider the road with ID=1 illustrated in Fig.6. When the water level reaches 0.27 m, the corresponding times are 1.1 hour and 1.2 hour. Due to the incongruity between the predicted runoff and observed water level, the runoff values at these two time points are 5.8 m^3/s (Point A in Fig. 6) and 12 m^3/s (Point B in Fig. 6), which accounts for the presence of two runoff values for a single water level.

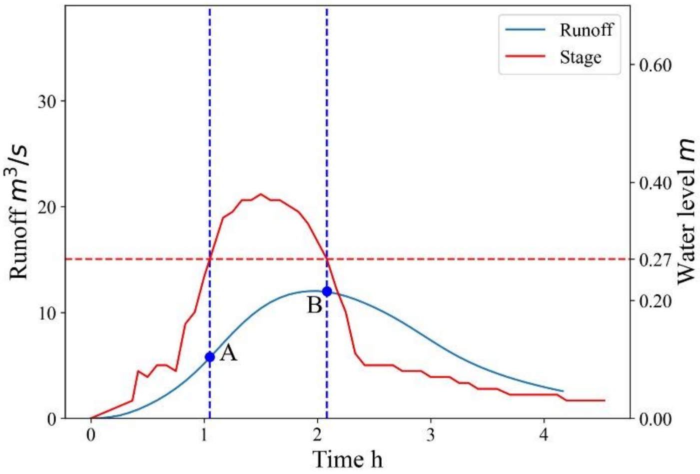

Figure 6 An example to show why some runoff values correspond to two level values.

Editorial comments

- l23 - suggest changing "metropolis" to plural "metropolises"

Response:

Modified as suggested.

- l31 - suggest changing "false" to "incomplete" or "over-simplified"

Response:

Modified to “incomplete” as suggested.

- l60 - suggest changing "critical" to "useful"

Response:

Modified as suggested.

- l87 - I do not think you need to define a hydrograph. I think you can safely assume HESS readers will know what a hydrograph is.

Response:

The definition of hydrograph is removed.

- l141 - "can absorb *a* light shower" (add "a")

Response:

Modified as suggested.

- l154 - I suggest changing "converts rainfall excess to direct runoff" to "converts rainfall excess to a temporal distribution of direct runoff" or something like that to communicate that it is a distribution of runoff over time.

Response:

Modified as suggested.

- l160 - "the paucity of runoff" instead of "the paucity of the runoff"

Response:

Modified as suggested.

- l161 - "sparkled" is probably not the right word here. Maybe "sparked" or "motivated"

Response:

Modified to “motivated” as suggested.

- l191 - "road" instead of "rood"

Response:

Modified as suggested.

- l195 - "equals the probability" instead of "equals to the probability"

Response:

Modified as suggested.

- l197 - suggest "impossible" instead of "difficult" because I think it is actually impossible to "obtain precise knowledge of all taxi-flooded intersections"

Response:

Modified as suggested.

- Table 1: Is it correct to have the "/"s for Feature in several of the rows? If so, maybe you should define that means.

Response:

Modified the “/” to “Not mention” in Table 1.

- l295 - suggest changing "a little bit" to "slightly" or something similar. "a little bit" is imprecise and colloquial

Response:

Modified to “slightly” as suggested.

- l308 - "waterlogging" is not a term I typically hear. Do you mean something like "flood-prone?"

Response:

Modified to “flood-prone” to enhance clarity.

- Figure 9 - is the x-axis "Time of Concentration?" If so, please change. I didn't know what "Time" meant.

Response:

Modified to “Time of Concentration” to enhance clarity.

- l396 - suggest replace "great" with "good"

Response:

Modified as suggested.

- l434 - suggest remove "great" to read "This study illustrates the potential ... "

Response:

Modified as suggested.

Reference

Beven, K., & Binley, A. (1992). The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes*, *6*(3), 279–298. https://doi.org/10.1002/hyp.3360060305

Brouwer, T., & Eilander, D. (2017). Probabilistic flood extent estimates from social media flood observations. *Natural Hazards and Earth System Sciences*, *17*(5), 735–747.

Goodwin, N. R., Coops, N. C., Tooke, T. R., Christen, A., & Voogt, J. A. (2009). Characterizing urban surface cover and structure with airborne lidar technology. *Canadian Journal of Remote Sensing*, *35*(3), 297–309. https://doi.org/10.5589/m09-015

Jet Propulsion Laboratory. (2019). *ASTER Global Digital Elevation Map*. https://asterweb.jpl.nasa.gov/gdem.asp

Ji, S., & Qiuwen, Z. (2015). A GIS-based Subcatchments Division Approach for SWMM. *The Open Civil Engineering Journal*, *9*(1), 515–521. https://doi.org/10.2174/1874149501509010515

Natural Resources Conservation Service. (2010). *Hydrology National Engineering Handbook Chapter 15 Time of Concentration* (pp. 5–6). United States Department of Agriculture.

Oksanen, J., & Sarjakoski, T. (2005). Error propagation analysis of DEM-based drainage basin delineation. *International Journal of Remote Sensing*, *26*(14), 3085–3102. https://doi.org/10.1080/01431160500057947

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

Zahura, F. T., Goodall, J. L., Sadler, J. M., Shen, Y., Morsy, M. M., & Behl, M. (2020). Training Machine Learning Surrogate Models From a High-Fidelity Physics-Based Model: Application for Real-Time Street-Scale Flood Prediction in an Urban Coastal Community. *Water Resources Research*, *56*(10), e2019WR027038. https://doi.org/10.1029/2019WR027038

Zhang, T., & Huang, X. (2018). Monitoring of Urban Impervious Surfaces Using Time Series of High-Resolution Remote Sensing Images in Rapidly Urbanized Areas: A Case Study of Shenzhen. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *11*(8), 2692–2708. https://doi.org/10.1109/JSTARS.2018.2804440