Flood Retention Lakes in a Rural-Urban Catchment: Climate-Dominated and Configuration-Affected Performances

Haochen Yan1,2, Mingfu Guan3, and Yong Kong1

1Department of Civil Engineering, The University of Hong Kong, Hong Kong SAR, China

Abstract Flood retention lakes (RLs) are widely employed in rural-urban catchments with low impacts on the natural environment. However, insights are lacking regarding the control of climate conditions on RLs' performances and how they are affected by different geographic configurations. This study applies a 2D hydrodynamic model to perform a catchment-scale performance assessment of RLs beyond the scope of analytical and hydrological models. We conduct extensive numerical experiments of rainstorm-induced flooding in a rural-urban catchment with a constructed RL and blueprinted ones upstream. Results demonstrate an L-shaped band of satisfactory performance of the current RL in the frequency-duration diagram, which coincides with short return periods (<5 years) and long durations (>4 hr), or short durations (<3 hr) and moderate to long return periods (5–50 years); such L-shaped pattern is also valid for additional RLs and their combinations. With the increase in event size, the first two modes of RL performance (out of four) correspond to effective flood mitigation. When working jointly, RLs with series configurations are more effective in reducing the mainstream flood peaks, while parallel connections provide a greater spatial extent of flood hazard mitigation. For both series and parallel configurations, the upstream-weighted settings tend to outperform downstream-weighted ones under more extreme events; the decentralized arrangement in the urban area yields more benefits in spatial flood hazard mitigation compared to the centralized case. The study highlights the critical role of rainstorm severity (with possible spatiotemporal variabilities) in controlling RL performances despite various configurations and hydraulic settings.

1. Introduction

Hydro-meteorological extremes are projected to increase in both frequency and intensity due to global warming (IPCC, 2021). As population and economic assets increase, particularly in rural and urban areas, flood exposure and associated damages are expected to rise consequently (Tellman et al., 2021), with both coastal and riverine cities bearing the brunt (Hallegatte et al., 2013; Lai et al., 2020; Mård et al., 2018). The lag between the pace of urbanization and the upgrade of flood protection levels especially in emerging cities makes the situation even more pressing (Merz et al., 2021; Tellman et al., 2021). Hence, timely and effective stormwater measures are essential to enhance social resilience to climate threats and improve human well-being (Maragno et al., 2018; Ruangpan et al., 2020). Nature-Based Solutions (NBSs) have been widely considered to be capable of effectively addressing environmental challenges and simultaneously providing biodiversity benefits (Cohen-Shacham et al., 2016; Qin et al., 2013) by taking actions that are inspired, supported, or copied from nature (European Commission, 2015). Flood retention lake (RL) (or retention basin), a common type of NBSs, typically intercepts part of runoff during flooding events and keeps it for subsequent release (Yang et al., 2011). Their capabilities of reducing downstream runoff and damping the flood peak generally decline with the increase of event magnitude but vary dramatically for different cases (Ayalew et al., 2015; Bellu et al., 2016; Birkinshaw et al., 2021; Chen et al., 2007). A quantitative effectiveness evaluation of RLs remains a challenging issue not only due to methodological limitations but also its complex sensitivity to climatic factors as well as catchment characteristics, including RL configurations.

Methods to evaluate the effects of RLs on catchment hydrology can be generally classified into data-driven and model-driven categories. Data-driven methods perform statistical analysis on a mass of field measurements taken at locations that RLs potentially affect, which can be compared to the sites with similar landcover and climates (Loperfido et al., 2014) or the same site in pre-construction periods (Birkinshaw et al., 2021). However, the applicability may be entangled by disparities and changes in the land cover and the climate, also limited by data quality and quantity (Habets et al., 2018). Model-driven methods can provide robust supplementary particularly for data-scarce sites/scenarios. Hydrologic models, sometimes combined with 1D hydrodynamic models, have
been extensively applied to efficiently and efficaciously simulate rural and urban stormwater systems (Ayalew et al., 2015; Iato-Espino et al., 2019; Lim & Welty, 2017; Luan et al., 2019). However, they are built upon simplifications in sub-catchment definition (preset demarcations for hydrologic units), process calculation (without or over-simplified momentum equations) and results presentation (aggregated and localized quantities). By comparison, hydrodynamic models, based on shallow water equations (SWEs), can achieve more realistic simulations when excessive surface flow occurs, as they can reproduce the physical rainfall-runoff processes both spatially and temporally (Bates et al., 2010; Guo et al., 2021; Ming et al., 2020). They are increasingly applied in rural and urban flash flooding modeling including both near-field hydraulics and large-scale rainfall-runoff processes in recent decades (Anselmo et al., 1996; Mignot et al., 2006; Ming et al., 2020; Schubert et al., 2022; Zhang et al., 2021). Besides, criteria to assess RLs’ performance await to be updated from a flood hazard mitigation perspective, where both spatially distributed inundation depth as well as water velocity get involved (Kvočka et al., 2016). This further highlights the need to probe into hydrodynamic processes on top of hydrologic levels.

Climate factors dominating floods fundamentally can be considered through intensity, duration, and the total amount of rainstorm events. It is commonly perceived and also verified that peak runoff and runoff depth increase with rainfall intensity and rainfall amount (A. J. Miller et al., 2021); longer duration storms tend to cause an increasing amount of rainfall (Lai et al., 2020), while events with time scales comparable with the catchment response time scale are likely to result in flooding. Guan et al. (2015a, 2015b) found that the ratio of runoff depth to rainfall depth almost remains constant for small events while it increases with considerable fluctuations as rainfall depth exceeds a threshold because of different intensities and duration. With the existence of RLs, the downstream peak runoff is projected to decrease in most studies with declining effectiveness as the increase of event size, but relatively less attention is paid to event durations (Ayalew et al., 2015; Birkinshaw et al., 2021; Chen et al., 2007; Vojinovic et al., 2021). While recent evidence indicates short-duration events are favored to achieve a better RL performance (Belli et al., 2016), further investigations of climate controls on RL performances (both hydrologic and hydrodynamic points of view) are needed by jointly considering intensity and duration with well-defined event processes and systematic frequency analysis.

In addition to climate factors, configurations of RLs also affect the overall catchment response to rainstorms, including the number and capacity of RLs and spatial placement. The combined effects of a group of RLs, and/or their interplay with other gray/green infrastructures can significantly differ from the simple summation of individuals, owing to the nonlinearity of the catchment system in terms of flood routing, as well as the nonlinear relation between the effectiveness of individual infrastructures and size of implementation (Luan et al., 2019). It is found that dispersed or distributed layouts yield better effectiveness than centralized or aggregated layouts assisted by a cellular model (Zellner et al., 2016), consistent with a data-driven comparative study carried out by Loperfido et al. (2014). Alongside this, RLs placed in parallel or upstream sections of the catchment have better flood control than those placed in series or near the downstream outlet demonstrated by a hydrologic model-based study (Ayalew et al., 2015). Analytical models developed by Del Giudice, Gargano, et al. (2014) and Del Giudice, Rasulo, et al. (2014) based on linear reservoir assumptions indicate that the overall effectiveness of parallel-connected RLs is the sum of each one while that of series connection is a complex function of RLs and upstream watershed characteristics and may lead to negative effects with additional RLs. Besides, RLs placed closer upstream contribute to a higher overall efficacy due to a larger ratio of RL size to an effective drainage area. Nevertheless, little is known about climate controls (e.g., event size) on various spatial configurations; evidence and insights furnished by hydrodynamic models are still lacking.

Given the points identified above, this study seeks to provide improved insights into how RLs (individuals and combined configurations) perform under a wide range of extreme climate conditions. Specifically, we employ both hydrologic (flood peak delay and reduction) and hydrodynamic (spatial flood hazard mitigation) perspectives to explore the following questions:

1. How is the best performance of an RL controlled by rainstorm frequency and duration?
2. How do different geographic configurations of multiple RLs influence their overall performance?

2. Study Site and Data

This study is primarily enlightened by the recently constructed flood RL located at the upstream reach of the Shenzhen River (ERM Hong Kong, 2010), which is the boundary river between the Hong Kong Special Administration Region (HKSAR) and the Shenzhen Special Economic Zone, China. The entire Shenzhen River flows...
from east to west into Shenzhen Bay, with a length of 37.6 km and a basin area of 312.5 km$^2$. The basin is governed by a subtropical climate and is susceptible to typhoons and low-pressure troughs that frequently induce extreme rainfalls and storm surges (Peking University, 2016). The upper reach of the Shenzhen River is generally narrow and winding with uneven width and collapsed embankment. Similar to the entire Shenzhen River Basin, the Shenzhen side of the basin is highly urbanized, while the Hong Kong side is mainly rural as shown in Figure 1. To upgrade the flood protection standard, the Phase-4 River Regulation Project was conducted during 2012–2017, starting from the confluence point, passing through the Liantang/Heung Yuen Wai Boundary Control Point, ending at about 4.5 km upstream near Pak Fu Shan in Hong Kong (ERM Hong Kong, 2010). The project consisted of (a) river channel modification using a trapezoid compound channel that follows the original watercourse and maintains the naturalness of the river and riparian habitats; (b) a flood RL covering and area of 22,000 m$^2$ with a storage capacity of about 80,000 m$^3$ (ERM Hong Kong, 2010). While the project also has comprehensive ecological considerations, this study mainly focuses on the effect of the RL on flood hazard mitigation within the upper Shenzhen River basin (about 19 km$^2$ as outlined in Figure 1) and is not limited to the current configurations.

The data used in this study (Table 1) mainly include hydrological and geographic aspects. Long-term (1986–2020) hourly rainfall at Ta Ku Ling station was obtained from Hong Kong Observatory, based on which the frequency analysis is carried out. The 4-year local hydrologic data from June 2018 to June 2022 measured at Liantang station (S2) with varying temporal resolution from minutes to hours, including runoff and local rainfall used for event-based model calibration and validation, were obtained from the Shenzhen River Regulation Office of Water Authority of Shenzhen Municipality. In addition, radar-derived rainfall data with a 6-min interval for a few
days when rainstorms occurred were obtained from the Meteorological Bureau of Shenzhen Municipality. The as-built river bathymetry of the regulation project in the form of survey points was also provided by the same department, which was then interpolated and combined with satellite-derived Digital Elevation Model (DEM) of the entire studied basin, with a 1-m spatial resolution. The land use data of the basin, which are necessary for the assignment of surface roughness and infiltration properties, were manually delineated from the Google satellite image and processed to have the same resolution as the DEM (Figure S1 in Supporting Information S1). To evaluate the effectiveness of the current RL, the DEM without the RL is produced by filling the RL with the bank elevation. Potential RLs sites (marked in Figure 1) are selected according to the DEM as well as the land use conditions, which will be interpreted in more detail in the next section.

### 3. Methods

#### 3.1. Hydrodynamic Modeling

In this study, the unsteady free surface flow of rainfall-runoff process is modeled by solving the full 2D SWEs (Brunner, 2021):

\[
\frac{\partial h}{\partial t} + \nabla \cdot (hV) = q
\]

\[
\frac{\partial V}{\partial t} + (V \cdot \nabla) V = -g \nabla z_s + \frac{1}{h} \nabla \cdot (\nu_h h \nabla V) - \frac{T_s}{\rho R}
\]

where $h = \text{water depth (L)}$, $V = \text{horizontal velocity vector (L/T)}$, $z_s = \text{water surface elevation (WSE)}$ (L), $g = \text{gravitational acceleration (L/T^2)}$, $q = I_0 + I_i = \text{source/sink flux term (L/T)}$, in which $I_0$ and $I_i$ are rainfall intensity and infiltration rate respectively, $\rho = \text{density of water (ML^{-3})}$, $R = \text{hydraulic radius (L)}$, $T_s = \frac{\rho}{\rho_n^2} |V|V = \text{bottom shear stress (ML^{-2}T^{-1})}$ and $n$ is Manning’s roughness coefficient ($\text{s/m}^{1/3}$). Infiltration is modeled by the Green-Ampt model assuming a saturated condition is already established before each rainfall event for a conservative consideration, although it is proved by preliminary runs to be secondary compared to rainfall for events with time scale of a few hours.

The numerical simulation is driven by HEC-RAS 6.2 (Brunner, 2021), a widely used and free software for 1D, 2D, or combined natural or constructed river systems modeling. Unstructured polygon meshes are generated for the simulation domain, where full 2D SWEs are solved by the Eulerian-Lagrangian scheme. In this study, 10-m uniform meshes are first established for the entire basin; refinement areas including river channels, urban streets and RLs are then delineated to have a mesh size of about 2–5 m (Figure 2a) while the meshes of mountain areas are coarsened to be 25 m. The total number of meshes is about 18k, with an average cell size of 105 m². Sensitivity analysis of mesh size shows that the current mesh layout leads to little RSME (1.3% of the flood peak) compared to a much finer mesh layout (43k) while keeping an acceptable efficiency (Figure S2 in Supporting Information S1). Besides the entire 2D domain, two culverts with a 1 m diameter are added at the upstream and downstream banks of the RL to connect the RL with the main channel at a lower elevation compared to the long broad weir (Figure 2a). The inlet/outlet elevations of culvert 1 and culvert 2 are about 1 and 0.3 m higher than the bottom of the corresponding river cross-section or the RL bottom, respectively. Therefore, water can freely enter and exit the RL under the low-flow conditions to maintain an aquifer environment and release the impounded water after floods (Figures 2b1 and 2b3). During flood conditions (Figure 2b2), a great majority of water flows into the RL through the long broad weir, while that flows across the culverts are nominal due to the small water elevation difference as well as the small diameters. Sensitivity analysis indicates little influence of culvert diameter on peak floods at S4 (will be discussed in Section 5). While this setting is only an assumption that mimics the real situation of human control, the effect is very close to the designed anticipation as it does not influence the functioning during floods and can empty excessive water within about 1 day to leave space for consecutive events.

#### 3.2. Selection of Blueprinted RLs and Design for Numerical Experiments

Several groups of additional RLs upstream are blueprinted as marked in Figure 1 (chose theoretically potential sites for future development and perform initial terrain modification). Considerations for potential and suitable RL sites are based on terrain slopes, susceptibility to flooding, land use conditions, and
RL size (area) (Rezazadeh Helmi et al., 2019); specific criteria include (a) enough free space for constructions (>2,500 m²); (b) the slope is generally mild; (c) obvious inundation occurs according to preliminary runs, and the site is adjacent to major flood pathways. Besides the limitations of specific terrain geometry of individual sites, all other parameters (i.e., the weir length, the culvert diameter and friction factors, and the culvert height from the RL bottom) are kept the same; the weir elevations are set to be the maximum WSE of 1-hr 5-year flood events locally while the lengths are set to be comparable with local channel/streets (50–70 m).

In addition to the case without any RLs (Case 1) and the current case (Case 2, with an existing RL1 downstream), we develop another two major cases with blueprinted RLs in series (Case 3) and parallel (Case 4) connections (Table 2). Three potential sites for series connection (adjacent to the mainstream of the Shenzhen River, RL2–RL5 in Figure 1) and other potential sites (RL6–RL15) are determined for each sub-watershed for parallel connection. The capacity of each RL is defined by the integration of the volume from the RL bottom to the weir elevation. In total, twice the volume of RL1 (160,000 m³) is allocated to each blueprinted RL following different rules: (a) evenly weighted—directly based on the ratio of contributing watershed area; (b) downstream weighted—heavier weights are assigned to the contributing area ratio of downstream RLs (e.g., for series connection, factors of 3, 2, and 1 are multiplied by the original contributing watershed area of RL2, RL3 and RL4 to calculate the allocation); (c) upstream weighted—heavier weights are assigned to the contributing area ratio of upstream RLs. For parallel cases, each sub-watershed is allocated one RL in principle. However, due to the dispersive flood pathways in the flat urban area, we particularly add a decentralized case (Case 4d) by adding RL5, RL8—RL10 in addition to RL6 and RL7 while keeping the total capacity in the same sub-watershed to be the same. For each blueprinted case, the DEM is then modified by replacing the corresponding elevation with a lower value to produce simplified RLs. Extensive sensitivity tests regarding the influence of weir and culverts settings as well as the RL storage capacity have been performed besides the current blueprints (Text S2 and Figures S6–S25 in Supporting Information S1).

3.3. Frequency Analysis and Design for Extreme Rainstorm Events

Frequency analysis is carried out to design extreme rainfall events in the studied basin based on the 35-year hourly rainfall record at Ta Ku Ling station in Hong Kong. A global design of idealized rainstorm events is then performed covering different event sizes (intensity and duration), including return levels of 1-year, 2-year, 5-year, 10-year, 25-year, 50-year, 100-year, 200-year, and durations of 0.5-hr, 1-hr, 2-hr, 3-hr, 4-hr, 6-hr, 9-hr, 12-hr. The rainfall processes are designed by the Chicago hyetograph and are then discretized into 5-min bins when applied...
### Table 2: Designed Rainstorm Scenarios With Blueprinted Retention Lake Configurations for Numerical Experiments

<table>
<thead>
<tr>
<th>Cases</th>
<th>Rainfall duration (hour)</th>
<th>Return period (year)</th>
<th>Description</th>
<th>Sketch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>0.5, 1, 2, 3, 4, 6, 9, 12</td>
<td>1, 2, 5, 10, 25, 50, 100, 200</td>
<td>Current configuration</td>
<td><img src="image1" alt="Sketch" /></td>
</tr>
<tr>
<td>Case 2</td>
<td>0.5, 1, 2, 3, 4, 6, 9, 12</td>
<td>1, 2, 5, 10, 25, 50, 100, 200</td>
<td>Current configuration</td>
<td><img src="image2" alt="Sketch" /></td>
</tr>
<tr>
<td>Case 3a</td>
<td>1, 2, 4, 6</td>
<td>10, 25, 50, 100</td>
<td>Series, evenly weighted</td>
<td><img src="image3" alt="Sketch" /></td>
</tr>
<tr>
<td>Case 3b</td>
<td>1, 2, 4, 6</td>
<td>10, 25, 50, 100</td>
<td>Series, downstream weighted</td>
<td><img src="image4" alt="Sketch" /></td>
</tr>
<tr>
<td>Case 3c</td>
<td>1, 2, 4, 6</td>
<td>10, 25, 50, 100</td>
<td>Series, upstream weighted</td>
<td><img src="image5" alt="Sketch" /></td>
</tr>
<tr>
<td>Case 4a</td>
<td>1, 2, 4, 6</td>
<td>10, 25, 50, 100</td>
<td>Parallel, evenly weighted</td>
<td><img src="image6" alt="Sketch" /></td>
</tr>
<tr>
<td>Case 4b</td>
<td>1, 2, 4, 6</td>
<td>10, 25, 50, 100</td>
<td>Parallel, downstream weighted</td>
<td><img src="image7" alt="Sketch" /></td>
</tr>
<tr>
<td>Case 4c</td>
<td>1, 2, 4, 6</td>
<td>10, 25, 50, 100</td>
<td>Parallel, upstream weighted</td>
<td><img src="image8" alt="Sketch" /></td>
</tr>
<tr>
<td>Case 4d</td>
<td>1, 2, 4, 6</td>
<td>10, 25, 50, 100</td>
<td>Parallel, evenly weighted and decentralized in urban area</td>
<td><img src="image9" alt="Sketch" /></td>
</tr>
</tbody>
</table>

The different configurations refer to combinations of flood retention lakes (RL) that simultaneously exist at different locations in Figure 1. For example, for Case 3a, RL1 + RL2–4 means RL1, RL2, RL3, and RL4 are operating jointly. Noted that the sketches are not to scale, and the sizes distinguish different weights assigned to the capacity.

---

The red and orange dots represent the existing and additional RLs; the size distinguishes different weights assigned to the capacity.
as the global boundary condition (Alfieri et al., 2008) starting from the dry initial condition. The design of simulation cases is summarized in Table 2. Details of related information are given in Appendix A.

3.4. Hydrologic and Hydrodynamic Perspectives for Performance Assessment of RLs

To evaluate the performance of different river-RL systems, simulation results are analyzed from both hydrologic and hydrodynamic perspectives. From a hydrologic perspective, we evaluate the rate of reduction of peak runoff in the hydrograph of key locations (e.g., S1–S4 in Figure 1) relative to the case without the RL and the delay of peak time, which is traditionally and widely used for RL effectiveness evaluation. On the other hand, hydrodynamic model results provide crucial information on the spatial distribution of inundation depth as well as water velocity as a function of time, which are key factors for flood hazard assessment. A commonly used empirical flood hazard index (FHI) is applied in this study for flood hazard mapping (Wallingford, 2006):

\[ \text{FHI} = D(V + 0.5) + DF \] (3)

where \( D (\text{m}) \) is inundation depth, \( V (\text{m/s}) \) is water velocity, \( DF \) is a constant measuring the risk caused by debris. In this study, we neglect the debris effect and separately consider the contribution of maximum inundation \( D_{\text{max}} \) and maximum flux per unit width \( DV_{\text{max}} \) for each flood event, such that:

\[ \text{FHI} = 0.5D_{\text{max}} + DV_{\text{max}} \] (4)

To reduce noises in the outputs generated by sub-grid technology based on the 1 m resolution terrain data, we upscale the results to 30 m grids and further classify FHI into 4 categories, namely negligible hazard for 0–0.05 (no caution needs to take), low hazard for 0.05–0.3 (e.g., 20 cm water moving at a speed of 20 cm/s, dangerous for some people), medium hazard for 0.3–1 (e.g., 50 cm water moving at a speed of 50 cm/s, dangerous for most people), and high hazard for FHI > 1 (dangerous for all the people).

4. Results

4.1. Model Calibration and Validation

Model calibration is carried out based on events of three different peak runoff magnitudes (Figure 3). Manual iteration steps are carried out in trial targeting at Manning’s coefficients for different land use types with preset imperviousness listed in Table 3. To be specific, based on the reasonable range suggested in the reference manual (Brunner, 2021), large, moderate and small values of the parameters are grouped to be initial candidates; groups of single variable tests for important land types such as grass, forest, and urban areas are also conducted. The candidate range of the parameters is gradually narrowed by keeping the groups with more minor discrepancies between the simulated and observed runoff time series, accompanied by slight adjustment of the parameters for further improvement. As extreme events are focused on in this study, we first conduct calibration based on the large event (Figure 3a). The goodness of the simulated time series increases nominally after about 20 trials, then calibration is extended to the moderate and the small event for further adjustment. In total, it takes more than 50 runs to obtain a single set of parameters (Table 3) that fit reasonably well for the three events with different magnitudes shown in Figures 3a–3c (events 1–3). It is noticeable that the falling limb (between 37 and 50 hr) of event 1 is significantly underestimated, and this is the common behavior of all the calibration runs (Figure S3 in Supporting Information S1). As no underground storage systems or external water transfer contribute to the channel, the discrepancies could be ascribed to (a) neglect of subsurface flows during the long-duration flood event or (b) the clogging of organic debris such as logs and brushes at the gaging station, which exaggerates the actual runoff. Nevertheless, the model can capture the peak magnitudes, and satisfactorily reproduces the processes for flasher events. Another 11 events within the time span of data availability are also identified for model validation. Both the simulated values of flood duration as well as peak runoff agree with the measurements (Figures 3d and 3e). Generally, better prediction is achieved for flood peak than duration, although slight overestimation is observed. The overall performance of the model indicates a reliable performance for further application under current data availability.

4.2. Hydrological Response of the Existing River-RL System

4.2.1. Typical Responses of an Individual RL

The current RL (RL1) interacts with the upstream flood waves modifies the pattern of hydrological response for its downstream reaches. The influence of the RL can mainly be attributed to the weir flow, while the flow through

---

YAN ET AL.
the culverts is much smaller and soon decays during floods due to the rapid decrease of water level difference inside and outside the RL. The function of the RL can generally be classified into 4 Modes with increasing flood magnitude (duration and intensity) as shown in Figure 5.

- **Mode-I (M1):** The RL is not filled up during the flood event. It effectively reduces the peak runoff, but the peak time of the two cases is almost the same (Figure 4a). Only a small volume of water enters the RL, and the maximum WSE inside the RL (at $t_1$) is lower than the weir elevation (8.4 m). At this stage, the free weir flow is only dependent on the WSE in the outside channel, so the superposition of the weir flow hydrograph with the original hydrograph of Case 1 changes in the same phase. The RL WSE gradually decreases as water is released to the channel through the culverts.

- **Mode-II (M2):** The total capacity of the RL is used in the flood event while the reversing flow amount is much smaller compared with its capacity. In this Mode, the RL obviously reduces the peak runoff and delays the peak time (Figure 4b). Initially, a milder rising limb of the hydrograph can be observed. Once the maximum WSE inside the RL exceeds the weir elevation, the positive inward flow rate rapidly decreases due to the transfer from free weir flow to submerged weir flow so that the downstream rising limb becomes steeper again. Additional capacity is excited due to the surrounding bank until outside WSE reaches its peak at $t_2$.

- **Mode-III (M3):** The total capacity of the RL below the weir elevation is used in the flood event while the reversing flow amount is comparable with its capacity. As shown in Figure 4c, the RL only has a small but detectable effect for both peak time delay and peak reduction. The maximum positive inward weir flow occurred much earlier than the arrival of the flood peak so that the rising limb near the flood peak is as steep as Case 1. The period of the reversing flow becomes much longer than Mode-II (from $t_1$ to $t_2$), while the maximum flow rate (about 5 m$^3$/s) does not see an obvious change. The WSE in the RL rapidly falls below the weir elevation through the reversed weir flow and then gradually decreases through culvert flow like the first 2 Modes.

Table 3: Calibrated Parameters for Hydrodynamic Modeling

<table>
<thead>
<tr>
<th>Landcover classification</th>
<th>Manning's coef. (sm$^{-1/3}$)</th>
<th>Percent imperviousness (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grassland</td>
<td>0.08</td>
<td>0</td>
</tr>
<tr>
<td>Open water</td>
<td>0.06</td>
<td>100</td>
</tr>
<tr>
<td>Bare land</td>
<td>0.03</td>
<td>100</td>
</tr>
<tr>
<td>Urban low intensity</td>
<td>0.12</td>
<td>60</td>
</tr>
<tr>
<td>Urban buildings</td>
<td>0.5</td>
<td>100</td>
</tr>
<tr>
<td>Forest</td>
<td>0.2</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 3. Calibration and validation of flood event modeling of the upper Shenzhen River Basin based on the measured data at Liantang station in 2018–2022. Panels (a–c) are calibration of events with high, medium and low flood magnitudes (event 1–3), (d) and (e) are validation of peak runoff and flood duration, respectively.
4.2.2. How Climate Conditions Dominate Satisfactory Zone of RL Effectiveness

The effectiveness of the current RL (RL1) dominated by a wide range of extreme rainstorms is evaluated in terms of the rate of reduction of peak runoff compared with Case 1, and the delay of flood peak. The contour maps derived from 128 simulations, 64 points (8 durations and 8 return periods for Case 1 and Case 2 in Table 2) are presented in a frequency-duration diagram in Figure 5 without smoothing. It is seen that the RL reduces the peak runoff and retards the peak arrival time up to about 25%, 25 min, respectively. The RL generally has a decent performance within an L-shaped band limited by event duration and exceeding probability for both peak damping and peak delay. The band generally lies at moderate to large return periods (5–50 years) combined with small durations (<3 hr) toward moderate to long durations (>4 hr) combined with small return periods (<5 years). Despite the similarities, effective peak delay is more demanding compared to peak damping; almost no delay effect can be observed outside the L-shaped band as discussed in the last subsection. It is interesting to note that the optimal zones of peak damping and peak delay are different and spaced from each other to a certain degree. This is because of the different nature of the typical weir flow conditions leading to the two effects—significant peak damping expects the peak of inward weir flow to coincide with the peak of the original external hydrograph

**Mode-IV (M4):** The weir becomes an indiscriminate flood pathway downstream, and the RL has nominal effects if the event size further increases, as shown in Figure 4d. The RL terrain can hardly demarcate the inundation during the event, and the flood water straightens its pathways to downstream, overlaying the bank. The residence time of the RL (ratio between the capacity and mean positive weir flow) is significantly smaller than the time scale of the flood event so the RL runs out of storage at the very early stage.

---

**Figure 4.** Typical Modes of retention lake (RL) function as a result of river-lake interaction with an example for each Mode, viewed at the near downstream location of the RL (S4). (a) Mode-I, 5-year event with 1-hr duration; (b) Mode-II, 10-year event with 1-hr duration; (c) Mode-III, 50-year event with 1-hr duration; (d) Mode-IV, 100-year event with 3-hr duration. The orange dashed lines mark the weir elevation (8.4 m).

---
(without the RL, or Case 1), while significant peak delay requires the zero (end) point of the inward weir flow hydrograph (which implies the peak of the external hydrograph of Case 2) to be as late as possible compared to Case 1 peak time, such that the time of peak weir flow is very likely to deviate from that of Case 1 hydrograph.

If the two criteria (peak damping and delay) are taken as equally important to assess the overall effectiveness of the RL, an effectiveness score can be calculated by $E_f = 0.5P_d + 0.5T_d$, where $P_d$ and $T_d$ are normalized peak damping and time delay by the maximum values among all the scenarios. The frequency-duration diagram of such an effectiveness score is plotted in Figure 5c. A more continuous L-shaped zone is obtained, based on which (and also the individual criterion) the four modes discussed in Section 4.2.1 are outlined. It is conjectured that the L-shaped demarcations for different Modes exhibit typical characteristics of RL-catchment systems in response to a wide range of duration-intensity combinations despite differences in specific values. This conjecture is verified from the results of multiple blueprinted RLs, which have different shapes, contributing areas and local geometry (Figures S7–S12 in Supporting Information S1); additional RLs (of Case 4d) can jointly increase the overall performance by widening the band and shifting it toward larger event size, while the L-shaped pattern does not change (Figure S13 in Supporting Information S1). Therefore, to preliminarily extract the pattern of the system performance, much fewer numerical runs could be necessary by following the white dashed lines roughly perpendicular to the demarcation lines in Figure 5c.

**Figure 5.** Diagrams of performance of an individual retention lake (RL) in response to a wide range of climate conditions (64 rainstorm durations and return periods for Case 1 and Case 2). (a) By flood peak reduction; (b) By flood peak delay; (c) Effectiveness score. The black dashed lines roughly mark the demarcations of different Modes (M1–M4) of RL function, while the white dashed lines provide simplified routes for preliminary extraction of system performance pattern with much fewer number of numerical runs.
4.3. Effects of RL Configurations on Flood Mitigation Performances

4.3.1. Series V.S. Parallel Configurations

RLs connected in series show a higher rate of flood peak damping than those connected in parallel if the configurations with the same weights are compared (i.e., Case 3a with Case 4a, Case 3b with Case 4b, Case 3c with Case 4c) as shown in Figure 6. This is because, for each site, the contributing area for series configurations is larger than that of parallel configurations, such that almost all the incoming water is regulated by the RLs. At the upstream location S1, both series and parallel settings achieve relatively higher effectiveness under moderate events (1-hr 50-year and 6-hr 10-year) than the smaller (1-hr 10-year) and the more extreme (6-hr 50-year) events, as the blueprinted storage capacities are more suitable for moderate events given respective contributing areas. The best performances shift toward smaller events when viewing the results further downstream (S2 and S4) due to the decreased capacity and local flood magnitude ratio. Besides, it is expected that the series configurations cannot provide benefits for the tributary watershed (S3); instead, there are even very slightly adverse effects for larger events (6-hr events) because the decreased downstream water level alleviates the backwater effect at the confluence. In accordance with the local flood peak reduction performance, series configurations show a larger magnitude of flood hazard mitigation than their counterparts. Still, this effect is limited because the RLs can only alter the flood processes nearby and downstream. The statistics shown in Figure 8 generally sit at a higher fraction of the area of flood hazard reduction for parallel groups than series groups, which is consistent with the above observation. In addition, the series configurations have a relatively flat distribution with an increasing magnitude of flood hazard reduction, while the symbols of parallel configurations roughly follow a decreasing trend. This also implies that under the same rainstorm conditions, parallel configurations are

![Figure 6. Comparison of peak runoff reduction compared to the current system (Case 2) at the four key locations among different geographic configurations of retention lakes (RLs). Flood events of different return periods (10- and 50-year) and durations (1- and 6-hr) are extracted for comparison. Case 3a to Case 3c are series connections with evenly, downstream, and upstream weighted, respectively; Case 4a to Case 4c are parallel connections with evenly, downstream, and upstream weighted, respectively; Case 4d is the same as Case 4a except RLs are decentralized in urban areas.](image-url)
less capable of reducing flood hazard for initially high-level-hazard regions (river channel), whilst they tend to take salient effects where the initial FHI is low (riverbank and shallow urban or rural inundated regions).

### 4.3.2. Upstream-Weighted V.S. Downstream-Weighted Configurations

Assigning uneven weights to RLs at different locations meaning to attach preferential capacity to address the flood hazards at parts of the watershed at the cost of impairing others. For the upstream-weighted groups, it is expected that the effectiveness of flood peak reduction at S1 is the best for all the rainstorm scenarios (Figure 6), and this is true for both series and parallel connections. When viewing further downstream (S2 and S4), for series configurations, the downstream-weighted group outperforms the upstream-weighted ones for short-duration events, while the outcome is reversed for larger events. However, parallel configurations show better performance for the upstream-weighted group for all events. This is due to the RLs in series configurations being also impacted by the upstream accumulation of excessive flood water. When event size increases, the accumulation of upstream excessive flood water overwhelms that of the available capacity downstream and leads to the rapid decay of downstream RL effectiveness. In contrast, as there are almost no interactions occurring in the flooding processes among the sub-watersheds, the decay of assembling effectiveness of the parallel-connected RLs is linearly linked with that of individuals. In addition, the performance of evenly weighted groups lies in the middle of upstream and downstream-weighted ones for all cases. Whilst a greater proportion of storage capacity in the upstream reach is supposed to provide larger flood hazard mitigation along the upstream channel, there

Figure 7. Spatial distribution of (a) maximum inundation $D_{\text{max}}$, (b) maximum flux per unit width $DV_{\text{max}}$, (c) flood hazard index (FHI) FHI for 1-hr 50-year event for Case 2, and (d–j) the change of FHI between Case 3a to Case 4d and Case 2 for 1-hr 50-year event. Case 3a to Case 3c are series connections with evenly, downstream, and upstream weighted, respectively; Case 4a to Case 4c are parallel connections with evenly, downstream, and upstream weighted, respectively; Case 4d is the same as Case 4a except retention lakes are decentralized in urban areas.
are no apparent differences for the 1-hr 50-year event (Figures 7d–7i), while a more apparent difference can be observed in more extreme events like the 6-hr 50-year event shown in Figure S4 in Supporting Information S1. The detailed comparison of the four scenarios (10- and 50-year, 1- and 6-hr events) in Figure 8 indeed shows that larger area benefits from more significant hazard mitigation (ΔFHI > 0.1) in the upstream-weighted group while less significant hazard mitigation (ΔFHI < 0.1) takes up a larger area in the downstream-weighted group for all tested rainstorm scenarios.

4.3.3. Centralized V.S. Decentralized Layouts

The inundation patterns in the urban region are distinct from rural regions, as shown in Figures 7a–7c. In the urban area, flooding is more dispersed across the flat and low-lying terrain and mainly occurs along the streets, while in rural areas, flood waters shape more confined pathways due to steeper slopes as well as more homogeneous roughness. The dispersed distribution of flood waters in the urban area hinders effective stormwater management if centralized RLs are placed downstream; instead, decentralized RLs with the same total capacity distributed across a wider domain can effectively increase the benefit, especially in terms of the area of flood hazard mitigation (Figure 7j). It is also shown in Figure 8 that Case 4d always has a more significant fraction of the area of hazard mitigation of all the ranges compared to Case 4a. However, the decentralized arrangements cannot provide visible improvements for flood peak reduction downstream (Figure 6). On the other hand, viewing
from the entire catchment, the series configurations can also arguably be regarded as centralized arrangements since only three additional RLs are employed, in contrast to the parallel configurations with 7–11 RLs. The obvious benefits of parallel groups that greater area achieve flood hazard mitigation and the apparent merits of series groups in reducing mainstream flood peaks resemble the above comparison between the decentralized and centralized settings.

5. Discussions

5.1. Control of Event Size on Individual RL Performances

As shown in Section 4, rainstorm conditions dominate the performance of an existing RL. With the increase in event severity (duration and intensity), four modes of RL functions have been characterized and demarcated. Alternatively, the event size can be indicated by the total upstream inflow to the RL $V_{\text{inlet}}$ (the integration of the flow volume at the immediately upstream cross section over the time period of an event when the RL weir flow occurs, normalized by the RL capacity). It is seen that the net inward weir flow volume $V_{\text{ret}}$ named retention volume here (also normalized by the capacity), positively correlates with $V_{\text{inlet}}$ (Figure 9a). Thus, $V_{\text{ret}}$ can serve as a local and convenient indicator of the flood magnitude, so of the event severity. On the other hand, we can observe the direct relationship between $V_{\text{ret}}$ and different modes of RL function (Figure 9b). With the increase of flood magnitude and thus $V_{\text{ret}}$, the RL first effectively damps the flood peak (Mode-I). However, as $V_{\text{ret}}$ further increases to around 1 so that the RL runs out of its capacity, reversed weir flow during the falling limb of the external flooding process starts to occur (Mode-II). Around this range of $V_{\text{ret}}$, the RL can achieve the optimal rate of flood peak reduction $R_p$ (Figure 9c). As $V_{\text{ret}}$ further increases, meaning that the flood event becomes overwhelming in terms of the RL capacity (Mode-IV), both $R_p$ and $V_{\text{ret}}$ approach 0 since the potential volume of reversed flow is limited by the bank above the weir elevation, while a greater portion of the flood water rushes through the weir directly downstream. When the weir elevation is increased while the capacity remains unchanged, the hill-shaped $R_p$ curve shifts toward a larger event size while the maximum $R_p$ slightly decreases; the magnitude of $V_{\text{ret}}$ significantly drops because the reversal volume (above the weir and below the bank) is decreased. The hill-shaped curves show that weir flow volume can coherently reflect how event size controls the RL performance, which is constituted by two factors (frequency and duration) in the L-shaped band. For either perspective, the linear assumption of RL effectiveness cannot accurately reflect
reality when estimating flood peak reduction, especially for extreme events. Only when the weir flow volume only accounts for a small portion of the capacity, the linear assumption may be valid (e.g., Mode-I, when the curve of $R_p$ is flatter than other modes). If the inlet flood volume is employed to estimate the RL performance, the nonlinearity (i.e., the changing effectiveness with event magnitude) also needs to be taken into account particularly around $V_{ret} = 1$, when the RL capacity is exhausted and the incoming flood flow starts to be diverted at a different ratio, thus a tipping point marking the change of the slope is observed (Figure 9a). These findings provide critical references to improve the fidelity of simplified analytical model-based assessment of RL-assisted flood mitigation at the catchment-scale, and thus its applicability in optimization problems.

5.2. Sensitivities of RLs Performances to Weir Settings and Capacity

While the effectiveness of individual RL is a complex function of climate conditions, RL locations and hydraulic properties, multiple RLs working in different configurations exponentially increase the complexity of the performance assessment. Rather than solving a specific optimization problem, our study aims to explore qualitatively what type of geographic configurations tend to be advantageous over others under different event magnitudes. To achieve this goal, we applied the same and ordinary criteria to the blueprinted RLs when allocating the capacity and determining the weir settings (Section 3.2). To further test the sensitivity of preliminary conclusions to those factors, extensive simulation runs are conducted with the objective quantities varying at a wide range about the original settings, for example, 50% longer and shorter weirs. A more detailed description of the sensitivity tests is given in Text S2 in Supporting Information S1. The results of the sensitivity tests regarding the competition of different RL configurations are summarized in Figure 10. In brief, from the hydrologic perspective (reduction of downstream flood peak), series connections generally outperform parallel connections; downstream-weighted allocation prefers less severe events compared to the upstream-weighted allocation. For the hydrodynamic perspective (mitigation of the spatial flood hazard), the parallel configurations generally perform better than series ones with larger uncertainty than the hydrologic perspective; the effect of upstream/downstream weighting is marginal for series connection, but clearly shows the preference for upstream-weighted allocation for parallel-connected systems. The above outcomes seldom meet contradictions when the sensitivity variables are modified, and they agree with the major conclusions of the original assignments (also the conclusions in Section 4.3), thus demonstrating the robustness of the findings in a qualitative sense. The sensitivity tests also show that the change in RL storage capacity, weir elevation, and weir length generally influence the overall RLs

Figure 10. Summary of sensitivity of retention lake (RL) performance with different geographic configurations to weir settings and RL storage capacity, viewed from four levels of event size and both hydrologic and hydrodynamic perspectives. Detailed descriptions of the sensitivity test as well as data plots are presented in Text S2 in Supporting Information S1.
performance in descending order in terms of sensitivity level. Additional sensitivity tests on weir elevation and weir length are also performed for an individual RL, for example, RL1 in Case 2 and RL4 in Case 3. The results (Figure 10 and Figure S24 in Supporting Information S1) indicate the same trends along with various event sizes despite different absolute quantities. Our sensitivity test also shows that the culvert diameter has no influence on the modeled downstream flood peaks (Figure S25 in Supporting Information S1).

5.3. Effects of Spatial and Temporal Rainfall Variability

The numerical experiments conducted in this study assume a uniform rain field across the entire basin and a single rainfall temporal pattern of the Chicago hyetograph for each event. However, rainfall events may occur with spatially varying intensities at the same time with complex hyetographs (including consecutive events). As widely discussed in previous studies, the above situations are nontrivial at all in terms of catchment hydrological response across different land cover features (Cristiano et al., 2017; Zhu et al., 2018), failure to consider the spatiotemporal variability with a resolution higher than a certain threshold (also controlled by catchment scales) results in the greater uncertainties and inaccuracy (Cristiano et al., 2019; Ochoa-Rodriguez et al., 2015).

Examples of 12-hr events with various spatiotemporal distributions are drawn herein for illustration. Temporally, the intensity-duration frequency analysis views original rainfall data by various time scales rather than durations, that is, a single event may be recorded as extreme candidates under different moving windows and thus time scales (Singh, 2017). Compared to the fact that a single event corresponds to one duration, such views by different time scales result in the overestimation of the peak intensity particularly for long durations, as shown in Figure 11a, based on more than 1,400 events measured at Ta Ku Ling station. The Chicago hyetograph further overestimates the rainfall amount compared to real cases as it assumes the same exceeding probability sustains for all smaller timescales. However, various temporal processes may lead to the same upscaled intensity. For example, a 12-hr 10-year event has roughly the same rainfall amount as the combination of a 6-hr 2-year event and a 2-hr 25-year event, very close to the 12-hr process during 27–38 hr in event 1 (Figure 3a) (239, 245, 238 mm, respectively, plotted in Figure 11b). Simulation results in Figure 11d show the highest peak for the consecutive event with the 25-year event occurring after the 2-year event, and all the Chicago hyetographs yield higher peaks than event 1 (hourly rainfall data). Spatially, the uniform pattern of rain fields assumes the entire catchment reaches the same exceeding probability. An example of the spatial distribution of the total rainfall during 27–38 hr in event 1 (Figure 3a) derived from radar data normalized by the mean value is shown in Figure 11c, clearly indicating heavier rainfall in the southwest. Compared with the results simulated by uniform rain gauge data, the actual flood hydrograph (calculated by the radar data, also used for model calibration) has a peak nearly 40% smaller. All the above results demonstrate how neglect of spatiotemporal rainstorm variability may lead to serious discrepancies in flood magnitude estimation—more concentrated peaks in the hyetograph, and consecutive events may cause considerably larger flood peaks compared to the hourly uniform patterns and single events.

5.4. Limitations and Future Work

We are aware of the following major limitations along with this study. First, the urban drainage network, particularly in the urban region of the Shenzhen side, is not systematically incorporated into the model. Plenty of previous studies demonstrate that urbanization processes cause increases in runoff volume, peak flowrate, and reduction in runoff duration and time to the peak (Guan et al., 2015b; J. D. Miller et al., 2014; Rogger et al., 2017), however, it is challenging to disentangle the effect of drainage system alone from land cover change. Unfortunately, publicly accessible fully coupled surface-subsurface hydrodynamic models are still lacking; detailed information on the real drainage systems in the studied catchment is also unavailable. In this study, a parsimonious subsurface drainage network is established along major streets with manholes and culverts coupled with the 2D model domain (details are provided in Figure S5 in Supporting Information S1). Simulations are carried out using the models with/without the drainage network. Results show that the existence of the drainage network causes earlier rising limbs of downstream flood hydrograph while a slightly smaller flood peak (<2% for the 10-year event and <1% for the 50-year event), and there is almost no change in peak time. The small discrepancies may be because unmodeled drainage processes have already been partially accounted...
The contribution of total drainage flows to the flood peak decreases from 15% to 7% as flood magnitude increases from 10- to 50-year, implying a limited impact of existing drainage systems on the hydrologic response to extreme events (>10-year return level). Nevertheless, a more accurate quantification of drainage capacity assisted by detailed drainage data is necessary to avoid overestimating surface flood hazards. Second, the hydrodynamic model relies heavily on high-quality DEMs, low resolution or erroneous representation of surface elevation may cause disconnection of flow paths and significantly alter surface hydrologic patterns, which is also the common challenges faced by other studies using hydrodynamic models (Horritt & Bates, 2001; Jarihani et al., 2015; Saksena & Merwade, 2015). Lack of higher levels of details leads to an overestimation of flood extent as well as depth due to underestimated conveyance of, or even undefined drainage pathways (Muthusamy et al., 2021). Third, the numerical experiments conducted by the catchment-scale hydrodynamic model under globally designed scenarios are still computation-intensive compared to hydrologic models. For example, a PC equipped with Intel i7-11700 CPU with 32 GB RAM and 12 cores has a real-time/simulation-time ratio of 2 under the numerical settings of this study. Finally, the impacts of spatio-temporal characteristics of local climatology on RL performances with different configurations require further studies.

Figure 11. Influence of spatiotemporal rainstorm variability on the estimated flood magnitude. (a) Rainfall events statistics viewing from event duration and time scales of the time series at Ta Ku Ling station; (b) hyetographs with the same mean intensity under 12-hr timescale; (c) normalized rainfall amount spatial distribution around the studied catchment of the calibrated event 1; (d) simulated hydrograph at S2 by the inputs in (b); legends of (d) are the same as (b).
6. Conclusions

Systematic numerical experiments by a full 2D hydrodynamic model are conducted to explore how extreme climate conditions and geographic configurations control the performance of flood RLs. Synthesized results indicate that decent effectiveness in terms of damping and delay of flood peaks generally lies in an L-shaped band in the diagram of rainstorm duration and return period, which, for the existing RL, coincides with small return periods (<5 years) combined with long durations (>4 hr) toward short durations (<3 hr) combined with moderate to large return periods (5–50 years), while the L-shaped band also exist for other sub-watersheds. Four typical Modes of RL performance are characterized based on the performances. RL’s net weir flow volume can serve as a convenient indicator of the Mode demarcation of RL functions under different rainstorm magnitudes. Additional RLs with various rules of configurations are blueprinted upstream of the current RL. RLs configured in series are more effective in reducing the flood peaks in the mainstream, while parallel configurations benefit a greater area in terms of flood hazard mitigation; the upstream-weighted allocations perform better under more extreme conditions while less effective for smaller events compared to downstream-weighted settings; the decentralized arrangement in the urban area yields greater benefits in spatial flood hazard mitigation than the centralized case. The above assessment of geographic configurations is almost insensitive to RL storage capacity, weir elevation and length, and culvert diameter within the wide range of variation tested. However, further efforts are required to account for the potential impacts of spatiotemporal rainstorm variabilities on the RL-river system. Overall, this study shows the critical control of climate conditions (rainstorm severity and possible spatiotemporal variabilities) on RL performance, while for different levels of rainstorm severity, preferences of geographic configurations hinge on the criteria/perspectives adopted.

Appendix A: Frequency Analysis and Design for Extreme Rainstorm Events

Frequency analysis of extreme rainfall measured at Ta Ku Ling station is performed for multiple durations. The Peak-Over-Threshold (POT) method is applied for sample selection because it can make sufficient use of historical events as it selects all events above a certain threshold rather than merely picking up the annual maximum and neglecting others (Hosseinzadehtalaei et al., 2020). The moving-average technique is first applied to the original time series in order to perform the analysis for various event durations (Willems, 2000), essentially allowing replication of the same extreme events viewing from various time scales (Singh, 2017). Rainfall events are identified based on the criteria that at least 5 mm rainfall is observed; the rainfall occurring within a 12-hr consecutive period is considered as the same event to guarantee independence (Willems, 2000). Next, only the events larger than 95 percentiles in terms of maximum hourly rainfall (MHR) and at least the same number of years are guaranteed to be utilized for the calibration of probability distribution models. For reference, based on hourly rainfall data, a total of 1,409 events with a mean MHR of 15.3 mm and a mean duration of 4.4 hr are identified. Finally, there remain 72 events with MHR >35 mm, larger than the Yellow Rain Warning in Hong Kong (30 mm/hr).

To estimate the event size corresponding to different exceeding frequencies, a variety of probability distribution models that are widely used in hydrological extreme event analysis are tested, including generalized extreme value (GEV) distribution, generalized Pareto (GP) distribution, Log Pearson Type-III (Gamma) (LM3) distribution, and Weibull (WB) distribution (Johnson, 1995), respectively. The best parameter estimation is selected among all the groups of parameters, based on the coefficient of determination between sample data and the predicted values from the model. In addition, the Anderson-Darling test as well as the Chi-square test (Meylan, 2019) are conducted for the calibrated distribution models; the RSME of the probability between calibrated models and sample points are also calculated for reference. The results of frequency analysis are plotted in Figure A1, including identified events of various durations and best-fit LM3 curves. The performance of the four probability models based on the four criteria is ranked from 1 to 4 (the higher, the better) and presented in the subplot of corresponding durations. Although all the models perform reasonably well, the LM3 distribution generally performs the best, and then is the Weibull distribution, while the GP distribution has the worst overall performance (Figure A1a).

Intensity-Duration-Frequency (IDF) curves of different durations are obtained and plotted in Figure A1b. The Chicago hyetograph is adopted for the design of the rainfall process so that for each duration, the rainfall intensity is congruent with the IDF curve, yielding a conservative estimation (Alfieri et al., 2008). Example designed processes of three different durations are presented in Figure A1c.
The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

The datasets that support this study are available at http://doi.org/10.5281/zenodo.5853044. The software employed to perform the numerical simulations in this study, Version 6.2 of HEC-RAS, can be downloaded from https://www.hec.usace.army.mil/software/hec-ras/download.aspx.

Acknowledgments

This study is financially supported by the National Key Research and Development Program of China for Intergovernmental Cooperation (No. 2019YFE0110100), Environment and Conservation Fund Project (ref: 108/2019) from Hong Kong Environmental Protection Department, and General Research Fund (No. 17202020) from Hong Kong Research Grants Council. We acknowledge the Shenzhen River Regulation Office of the Water Authority of Shenzhen Municipality for providing river cross-section data and hydrological data, and the Meteorological Bureau of Shenzhen Municipality for providing radar rainfall data.

References


Figure A1. Frequency analysis and design of rainstorms. (a1–a3) show examples of identified and fitted Log-M3 distributions of events with 2-, 4-, and 6-hr durations; the subplots of radar maps show the performance of four probability models based on the four criteria namely Chi-Square test (Chi), Anderson-Darling test (AD), coefficient of determination (R2) and RSME; (b) shows the IDF curve constructed based on the frequency analysis; (c) plots examples of designed rainstorm based on Chicago model.


