

The role of hyetograph shape and designer subjectivity in the design of an urban drainage system

Spyros Pritsis ^{a,*}, Vincent Pons ^{a,b}, Marius Møller Rokstad ^a, Francois H. L. R. Clemens-Meyer^{a,c}, Manfred Kleidorfer ^d and Franz Tscheikner-Gratl ^a

^a Department of Civil and Environmental Engineering, Norwegian University of Science and Technology (NTNU), S. P. Andersens veg 5, 7031 Trondheim, Norway

^b Department of Civil, Environmental, and Natural Resources Engineering, Luleå University of Technology, 97187 Luleå, Sweden

^c SkillsInMotion B.V., Esdoornlaan 11, 3454 HH, Utrecht, The Netherlands

^d Unit of Environmental Engineering, University of Innsbruck, Technikerstrasse 13, 6020 Innsbruck, Austria

*Corresponding author. E-mail: spyridon.pritsis@ntnu.no

 SP, 0009-0008-7857-0958

ABSTRACT

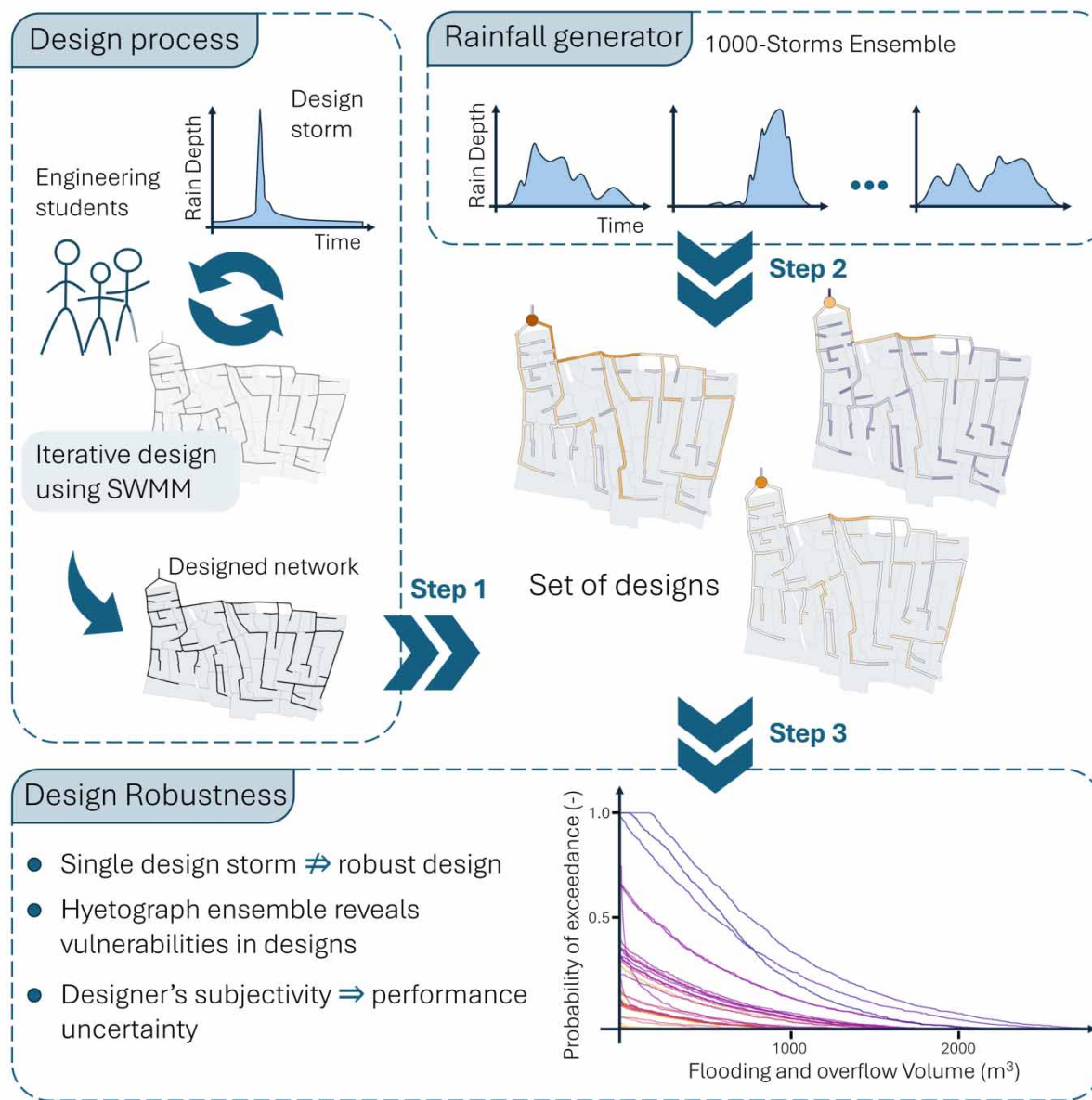
Even though it has been established that a hyetograph's shape affects the results of hydrological simulations, common engineering practice does not always account for this fact. Instead, a single design storm is often considered sufficient for designing a urban drainage system. This study examines the impact that this design paradigm, combined with the uncertainty introduced by subjective choices made during the design process, has on the robustness of a designed system. To do so, we evaluated a set of individual designs created by engineering students using the same Chicago hyetograph as a design storm. We then created ensembles of hyetographs with the same precipitation volume and duration as the Chicago hyetograph and evaluated the designs' hydrological responses. The results showed that designs, which performed equally well for the initial design storm, triggered varying responses for the storms in the ensembles and, consequently, showed different levels of robustness, hinting at a need to adapt the current design approach.

Key words: robustness, structural uncertainty, urban drainage modelling, urban hydrology

HIGHLIGHTS

- Designing an urban drainage system (UDS) using a single design storm does not produce a robust result.
- Storms with the same total precipitation volume but different temporal patterns can expose different vulnerabilities in an UDS.
- Designer subjectivity introduces uncertainty to the performance of a UDS.

GRAPHICAL ABSTRACT



INTRODUCTION

The application of (computer) models for simulating rainfall–runoff and the hydraulic response of urban drainage systems (UDSs) is routine practice for (re)designing such systems. They serve as tools for estimating how different designs respond to various scenarios (e.g., climate change or urban development). It is understood that the uncertainty in the model results affects the final design. Many studies have focused on identifying the sources of this uncertainty and quantifying its influence on the model results (Dotto *et al.* 2014; Tscheikner-Gratl *et al.* 2019). The goal should not be eliminating uncertainty completely (which is impossible), but rather acknowledging and, where possible, reducing it (Deletic *et al.* 2012). Since uncertainty is bound to influence the outcome of the design process, it should also affect the choices made to structure this process. In other words, we should acknowledge that any design approach is also a source of uncertainty in itself (Refsgaard *et al.* 2006). In this work, we describe this as a structural uncertainty – stemming from assumptions and decisions that are implicit when following an approach (e.g., choice of design event(s) and performance metrics) – insofar as it is

connected to the form of the model itself and its inadequacy to represent the processes present in the modelled system (Deletic *et al.* 2012).

The recognition of this type of uncertainty should prompt us to a reflexive look into our design approaches and the framing of the task of UDS design (Craye *et al.* 2005). Ignoring that can lead to a false level of trust in the performance of our designed system when it comes to meeting its set goals.

A common design approach is to set a goal (or a set of goals) to be met at the system scale (e.g., no flooding) and run simulations to arrive at an acceptable system design using a design hyetograph to drive the model simulation. This method results in a UDS that is heavily dependent on the design hyetograph used (Alfieri *et al.* 2008). Many guidelines prescribe the goals to be set as well as a rudimentary approach to selecting a design hyetograph. In Europe, the European Standard EN 752 'Drain and sewer systems outside buildings – Sewer system management' defines return periods for flood protection. It serves as the basis for various national guidelines, such as the ÖWAV Guideline 11 in Austria, 'DWA-A 118E' (2006) in Germany, and national versions of the European Standard like the SS-EN 752 (SIS – Bygg och anläggning, 2017) in Sweden, the BS EN752 in the United Kingdom (British Standards Institution (BSI), 2017), and the SN EN 752 in Switzerland. In the USA, the design of UDSs is described in ASCE (2013) and 'Urban Drainage Design' (Kilgore *et al.* 2024), in Queensland (Australia) in the Queensland Urban Drainage Manual (Institute of Public Works Engineering Australasia (IPWEA), 2016).

All design approaches have in common that they are based on a risk/cost assessment that balances investment and damage. As a result, design return periods of rainfall events are proposed (sometimes depending on the land use, i.e. the expected damage) for which the runoff has to be calculated and for which the hydraulic design has to be carried out. They often use a frequency-based approach, which means that a return period is given to the practitioner, and steps to arrive at a hyetograph are recommended as follows: (1) obtain an intensity–duration–frequency (IDF) curve, (2) calculate the critical rainfall duration and (3) create a design hyetograph using a temporal pattern from literature (e.g., Keifer & Chu 1957; Chow *et al.* 1988; DWA-A 118E 2006; Veneziano & Villani 1999). There has been a lot of focus on addressing how to construct and use IDF curves (first two steps), with many researchers criticizing the assumption of climate stationarity (Underwood *et al.* 2020), which is part of the standard IDF curve usage. Meanwhile, others have proposed a shift towards stochastic models to replace the IDF curves (Koutsoyiannis 2021).

While in almost all other fields of engineering, safety coefficients on either load, the system's capacity, or both are applied to compensate for uncertainties in the design, this tactic is, to the authors' knowledge, not explicitly adopted in the urban drainage community. Lately, there have been climate change safety coefficients introduced in some countries (Arnbjerg-Nielsen 2012). These coefficients are used to account for increased rainfall loads in climate projections, but they do not address the uncertainties in the designed system's performance. Instead, they attempt to ensure that it can handle anticipated future conditions regarding increased volumes while maintaining existing hyetograph shapes.

Conversely, the effect of the shape of the hyetograph on the design has not been discussed as frequently. Several studies have focused on the effects of rainfall's temporal pattern (Alfieri *et al.* 2008; Willems 2012; Funke *et al.* 2021) and resolution (Ochoa-Rodriguez *et al.* 2015) on simulated runoff for both grey infrastructure and green infrastructure (GI). Even if a perfect IDF curve is constructed (and the climate stationarity assumption is validated), the issue of choosing the temporal distribution of the hyetograph remains. Commonly used hyetograph shapes are block rainfall events (Mugume & Butler 2017), alternating blocks (Chow *et al.* 1988), BLUE (Veneziano & Villani 1999), Chicago (Keifer & Chu 1957), Euler type II (DWA-A 118E 2006) and more; all created using different methodologies and different sets of assumptions. However, they are all based on a similar postulation, namely that the design hyetograph pattern aims to transform the return period of a rainfall event into a flood event with the same return period (Maidment 1993). It has been shown, however, that for both natural and urban catchments, the assumed 1:1 relationship between the two return periods does not hold (Wright *et al.* 2014).

Acknowledging the possible mismatch between the nature of the task of designing a modern UDS and the methods commonly applied, researchers have been developing alternative approaches. Some take advantage of the increased availability of computational power to run long-term simulations (Dotto *et al.* 2011), bypassing the problem of design storm definition and selection but creating the need for long observed precipitation time series. This introduces a new set of difficulties, including the risk of accumulating calculational uncertainties tied to the model structure and the numerical methods used (Beven 2011; Deletic *et al.* 2012). Others (Guo & Adams 1998; Guo & Zhuge 2008) use an analytical stochastic conceptualization of the rainfall–runoff model, which treats the rainfall and runoff time series as realizations of random variables allowing them to better represent them regarding the dynamics of their temporal distributions. This also bypasses the design storm problem but is limited when it comes to describing the processes present in a catchment (Guo & Zhuge 2008). This approach can

also seem arcane and/or cumbersome to practitioners, trammelling wider uptake. A move towards risk-based, instead of frequency-based, design has also been proposed as a different approach to effectively tackle (or attempt to circumvent) the same problem (Markolf *et al.* 2021). Many of these proposed alternatives have shown potential in research case studies, but they have not been consistently codified into guidelines. Most guidelines suggest a design methodology driven by a single generic design storm. Some guidelines use regional hyetograph shapes (characteristic storm events) (Cordery *et al.* 1984; Ball *et al.* 2019). But still, the possibility that an ensemble of plausible storms may be necessary to adequately (robustly) design a UDS is not being conceded.

In the design of UDSs, different system stability measures such as reliability, robustness, resistance, resilience or vulnerability linked to hazards and risks are used (Casal-Campos *et al.* 2018). Reliability refers to the consistent performance of a system over time without failure (Billinton & Allan 1992). Resilience is seen as the ability to return to normal operating conditions after disturbances (Mugume *et al.* 2015); together with resistance (describing that a system is not reacting to disturbances), they are seen as the main characteristics of an infrastructure system (Tahmasebi Birgani & Yazdandoost 2014). The concept of vulnerability describes a system-dependent property that links a specific hazard to its effect in the risk assessment framework (Hauger *et al.* 2006). In systems engineering, robustness is defined as the insensitivity to variations normally encountered in production and operational use (Blanchard & Blyler 2016).

For our study, we define a robustly designed UDS as a design that ‘yields outcomes that are deemed to be satisfactory according to some selected assessment criteria across a wide range of future plausible states of the world’ using the definition for a robust plan provided by Walker *et al.* (2013). We address two influencing factors of the design process: (i) the influence of the hyetograph shape, rather than the volume, on the performance of the designed networks and (ii) the structural uncertainty inherent in the subjectivity of the designers’ choices when defining the network characteristics. In our analysis, the effects of the second factor are considered in view of the information revealed by investigating the first factor (i.e., the influence of the designers’ choices is gauged after each design is tested with multiple hyetographs).

METHODS

In this study, we investigated the response of different UDS designs of the same catchment to an ensemble of hyetographs with the same duration and total precipitation volume as the original design storm, but with different temporal patterns. By doing so, we can examine the influence of varying hyetograph shapes on different UDSs that were created by different designers using the same design storm.

We can segment our process into three distinct parts: (1) creating the UDS designs, (2) creating the event ensembles, and (3) running simulations and evaluating the outcomes. The first step was to obtain a set of UDS designs for the same catchment whose variations are the product of the designer’s subjectivity. To do this, over the course of three semesters, 52 MSc students were given the same exercise in which they were asked to design a UDS given a common design guideline. For the second part, we used Multiplicative Random Cascade (MRC) models to create ensembles of hyetographs. The set of UDS designs was then tested using the hyetograph ensembles in SWMM 5.2 (Gironás *et al.* 2010) with swmm-api (Pichler 2022) to interface with Python. The modelling results were analysed to assess robustness. The methods used in these three parts are described in the following subsections.

Creating the UDS designs

The study site is a relatively small and steep urban catchment (22 ha of total area, with 54% of imperviousness and an average slope of 12%). The geographical layout of a stormwater system is provided for this catchment. It is designed as a separate drainage system for which only the stormwater system is considered in the design process. There are no inflows from residential or industrial sources. The given pipe network (consisting of 235 pipe segments) routes the runoff to a retention basin (RB), which has two outflows: OF, a conduit conveying a maximum of 250 L/s to a downstream system, and W, a weir that handles surplus water volume by overflowing to a river (see Figure 1).

This model was provided to the students as part of a course project, together with a Chicago design hyetograph as a hydraulic load for the simulations during the design process. Their task was to optimize the UDS within the limitations of the given layout. They were explicitly given the freedom to select the dimensions of the pipes, RB volume and weir (W) height, as well as the depth of the manholes and elevation of the pipes (within local regulations, e.g., regarding frost depth and minimum/maximum slopes). The objectives were to induce **no flooding** in the catchment and **no overflow** to the river while minimizing the cost of their design using a cost function taken from the literature (De Toffol *et al.* 2007; Sitzenfrey *et al.* 2013). This cost

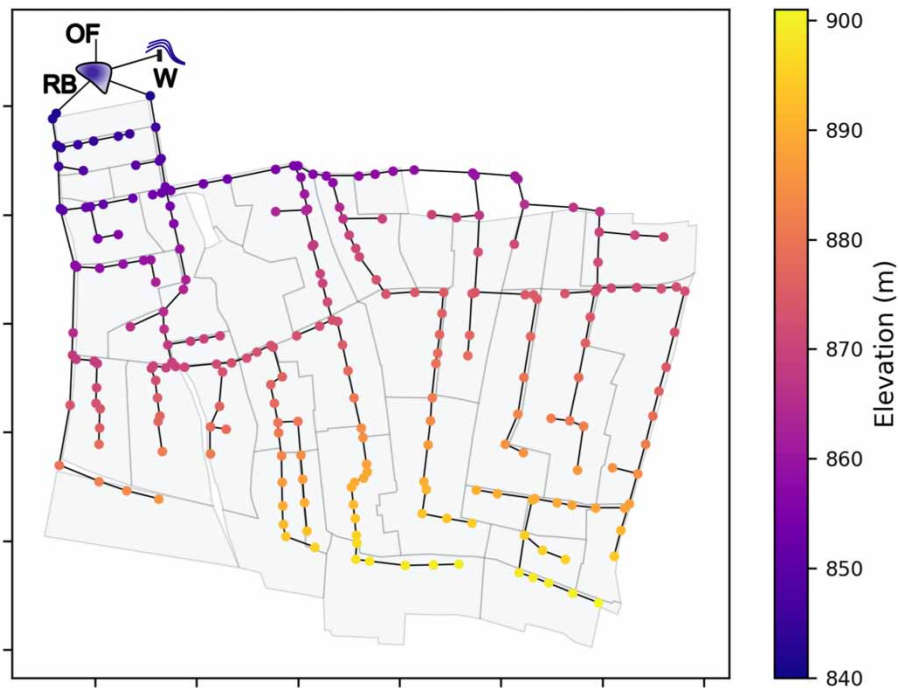


Figure 1 | Overview of the study catchment (showing elevation at the lowest invert level for each manhole). RB: retention basin; OF: outflow conduit; W: weir overflowing to the river.

function depends solely on the volumes of the pipes and the detention basin. Although the students were not restricted by the given exercise in changing Storm Water Management Model (SWMM) simulation settings (e.g., ponding), few chose to do so and to no visible effect.

Not all the designs submitted by the students met the given requirements, but we chose to initially discard only those that produced flood or overflow volumes above 10^5 m^3 (20% of the rainfall volume and approximately 30% of the runoff). This resulted in a total of 41 acceptable designs used for further evaluation. By considering designs that do not meet the design requirements as well, we wanted to assess their robustness across different temporal hyetograph patterns and compare the magnitude and mode of failure to the initial results obtained from the design storm simulation. We observe two modes of failure in our system, flooding and overflow, both of which can be categorized as functional (Butler *et al.* 2017).

The differences between designs are a result of subjective choices made by each designer (student) within the constraints of the structure of the design approach and the software applied. By analysing these designs, we aim to focus on the impact of these subjective choices on the systems' performance.

Creating the storms

Sampling of the hyetograph ensemble was performed using MRCs. The $\text{MRC}_{\text{SIT,SEP}}$ model developed by Pons *et al.* (2022a) was used. We chose the model structure that depends on the timescale, intensity, and temperature (SIT) and also utilizes a stochastic element permutation generator (SEP), as it showed the best performance in Pons *et al.* (2022a). This temporal statistical downscaling method is usually applied to time series but has been suggested to be used as a weather generator in a single framework for both time series and extreme events under climate change by Pons *et al.* (2022b). More concisely, the MRC model is trained using long time series with high temporal resolution. This allows it to downscale from time series with a coarser resolution to a finer one. To create an ensemble of n storms, we run the same (trained) model n times, setting the total precipitation depth and storm duration, resulting in n different storms (because of the stochastic nature of the model). Moreover, the current study investigates the extension of the Highly Informed Design Evaluation Strategy (HIDES) (Pons *et al.* 2022b) from the local scale to the UDS scale by examining the potential of extreme event sampling for system design purposes. Indeed, the HIDES originally used synthetic rainfall events derived from the $\text{MRC}_{\text{SIT,SEP}}$ models to analyse the robustness of designed green roofs. In Pons *et al.* (2022a) 10^6 events were sampled; however, because the

current study required the computation of a set of catchment models that increased the computational time, it was decided to only use 1,000 events per model as input. Using a smaller number of events is expected to influence the estimation of the probability of failure but still allows for assessing the behaviour of the systems under various probable hyetographs. Indeed, by simulating 10,000 events for two selected designs (see Appendix), we calculate the maximum 95% confidence interval for their probability of failure (for any threshold) as 5%. This amount of uncertainty does not influence the discussion of the results.

MRC_{SIT,SEP} models for two locations were considered in the current study: (i) the MRC_{SIT,SEP} model of Bergen, Norway, developed, calibrated and tested in Pons *et al.* (2022a) and (ii) an MRC_{SIT,SEP} model calibrated and tested in the current study using a local dataset. The local model was calibrated and tested similarly to Pons *et al.* (2022a) using precipitation and temperature time series from 1993 to 2023. The measurements were taken from a weather station near the study site with a 10-min resolution. Temperature measurements were available from two sensors located 5 and 200 cm from the ground. The precipitation measurements were performed using a tipping bucket.

Figure 2 shows the maximum depth of precipitation observed over a time window during the storm events plotted against the duration of the time window. The Chicago hyetograph produces volumes above the mean of the synthetic storms created using the MRC models for all durations. The difference is especially large for short durations, which is expected since Chicago hyetographs are designed to have one extreme peak, around which most of the precipitation volume is located. This difference decreases when longer durations are considered. The Chicago hyetograph volumes fell within the 95% range of the ensembles for time windows larger than 14 min for the Bergen model and 22 min for the local model. This means that the Bergen model generates, on average, storms with larger rain volumes concentrated around peaks than the local model. The ensembles generated by the MRC models are referred to as SYNTH_{LOC} and SYNTH_{BER}.

Simulations

Simulations of all synthetic storms were run for all different designs using SWMM 5.2 (Gironás *et al.* 2010). The same assumptions that were made for the design of the models are made here (i.e., the storage unit is empty when the simulation starts and there is no dry weather flow in the system). In the models, a single rain gauge was used to input the hyetographs and

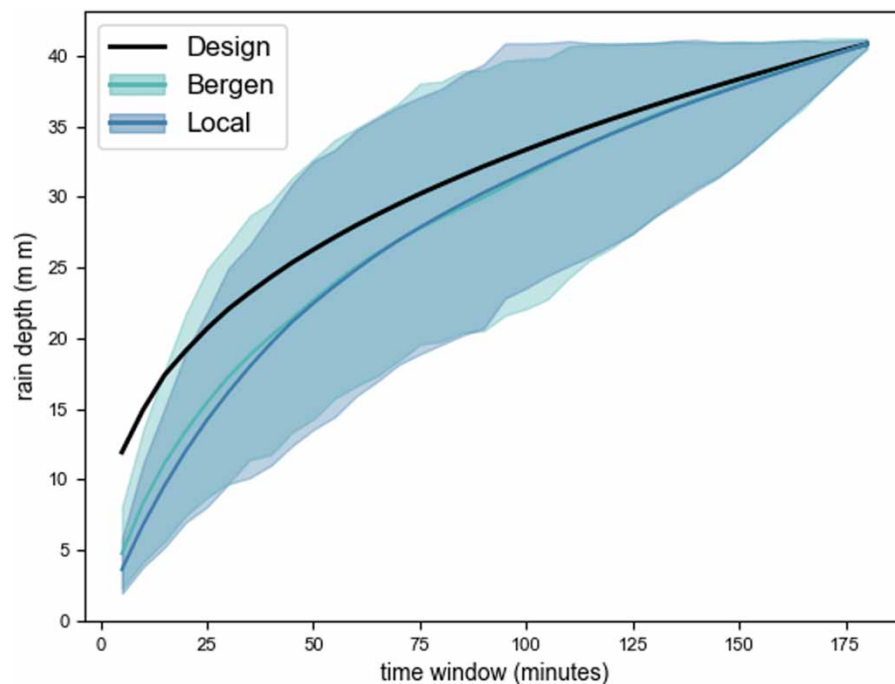


Figure 2 | Highest aggregate rain depth for different time windows. Each value shows the maximum total amount of rainfall observed within the corresponding (moving) time window. For the ensembles, the mean values are shown (the shaded areas indicate the range within which 95% of the observations fall).

the spatial distribution of the storms is considered uniform across the simulated area. Given the size of the catchment (22 ha), the effect of spatial variability in rainfall is not expected to be significant.

The results of the two ensembles of hyetographs (SYNTH_{LOC} and SYNTH_{BER}) were analysed independently to investigate the effect of training the MRC model on local datasets.

The UDS of the catchment has two modes of failure, resulting in flood volume in the catchment (V_{FL}) and overflow to the downstream river (V_{OF}). The sum of both (V_{tot}) is also considered a variable of interest in the analysis of the results presented below, representing the total amount of water exiting the drainage system (excluding the volume carried downstream by OF, which is not of interest in this analysis). We introduce a simple robustness metric ρ to characterize each design, only depending on V_{tot} . The metric, ρ (Equation (1)), is a number between 0 and 1 quantifying robustness as we have defined it for this study (i.e., '[the frequency of] outcomes that are deemed to be satisfactory according to some selected assessment criteria across a wide range of future plausible states of the world' (Walker *et al.* 2013)). The assessment criterion we use to deem an outcome satisfactory is whether V_{tot} is below a certain threshold and the different hyetographs (or future plausible states of the world) are assumed to be equiprobable.

$$\rho = \sum_{i=1}^n \frac{T_i}{n}, \quad \text{where } T_i = \begin{cases} 1 & \text{if } V_{tot} < \tau \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where n is the number of different hyetographs and τ is the threshold.

Water that ponds above the manhole level of a node is considered in the flood volume, even if it is reintroduced into the system (this only applies to four designs where ponding was allowed in SWMM). Water that only causes node surcharge (i.e., raising the water level of a node above the crowns of all connected conduits but below the manhole level) is not considered.

After running simulations of all the selected designs for both the storm ensembles, in addition to the original design storm used in the design process, we compared the results. By using V_{FL} and V_{OF} to characterize the response of a design, we get information on the mode and magnitude of failure (if a failure occurred during a given simulation). Even though the initial design criteria mandate that both V_{FL} and V_{OF} should be zero, we used a slightly more lenient threshold of 50 m^3 to define the acceptable designs for this analysis. The reason we do not use the initial strict threshold of no flooding and no overflow is to avoid characterizing designs that produce negligible volumes as not acceptable. We chose a value of 50 m^3 as the cut-off between negligible and non-negligible volumes because it is less than 1% of the total runoff produced by the SWMM simulation using the design storm. However, it still differentiates between the cluster of designs that respond very well to the design storm and the rest of the designs. This assumption makes the results more interpretable, given the fact that there is a small number of available designs (41) and 22 of them produce $V_{tot} < 50 \text{ m}^3$ for the Chicago design storm.

RESULTS AND DISCUSSION

Figure 3(a) shows the response (V_{FL} and V_{OF}) of each design for the original design storm. More than half of the designed systems met the (updated) design criteria. The rest (19) of the designs (shown in the carousel pink-shaded area of Figure 3) gravitate towards either one or the other mode of failure, with only two having non-negligible volumes in both flooding and outflow. The main reason is that some of the designs include a large RB, whereas others use a smaller one in favour of larger pipe dimensions in the network. However, the two modes of failure are not mutually exclusive, as demonstrated by the two aforementioned designs.

Subplots (b), (c) and (d) of Figure 3 show the median values of V_{FL} and V_{OF} for the 25, 10 and 5% most critical storms of both ensembles. The criticality of a storm was calculated by summing the V_{tot} values of all (41) designs when simulated with that storm.

Looking at the designs that failed to meet the design criteria for the design storm (i.e., the designs in the shaded part of Figure 3), we see a varying response to the ensembles. Some designs seem to have a better response to the ensembles in Figure 3(b), but when we only consider the most severe storms of either ensemble (Figure 3(c) and 3(d)), all but four designs produce median V_{tot} volumes larger than the design storms. Focusing on the four exceptions, they produce a negligible V_{tot} even for the most critical storms of the ensembles. The common characteristic of these four designs is that they use very large RBs ($5,600 \text{ m}^3$). In contrast to that choice, the pipe dimensions are not chosen conservatively (maybe to keep the total cost

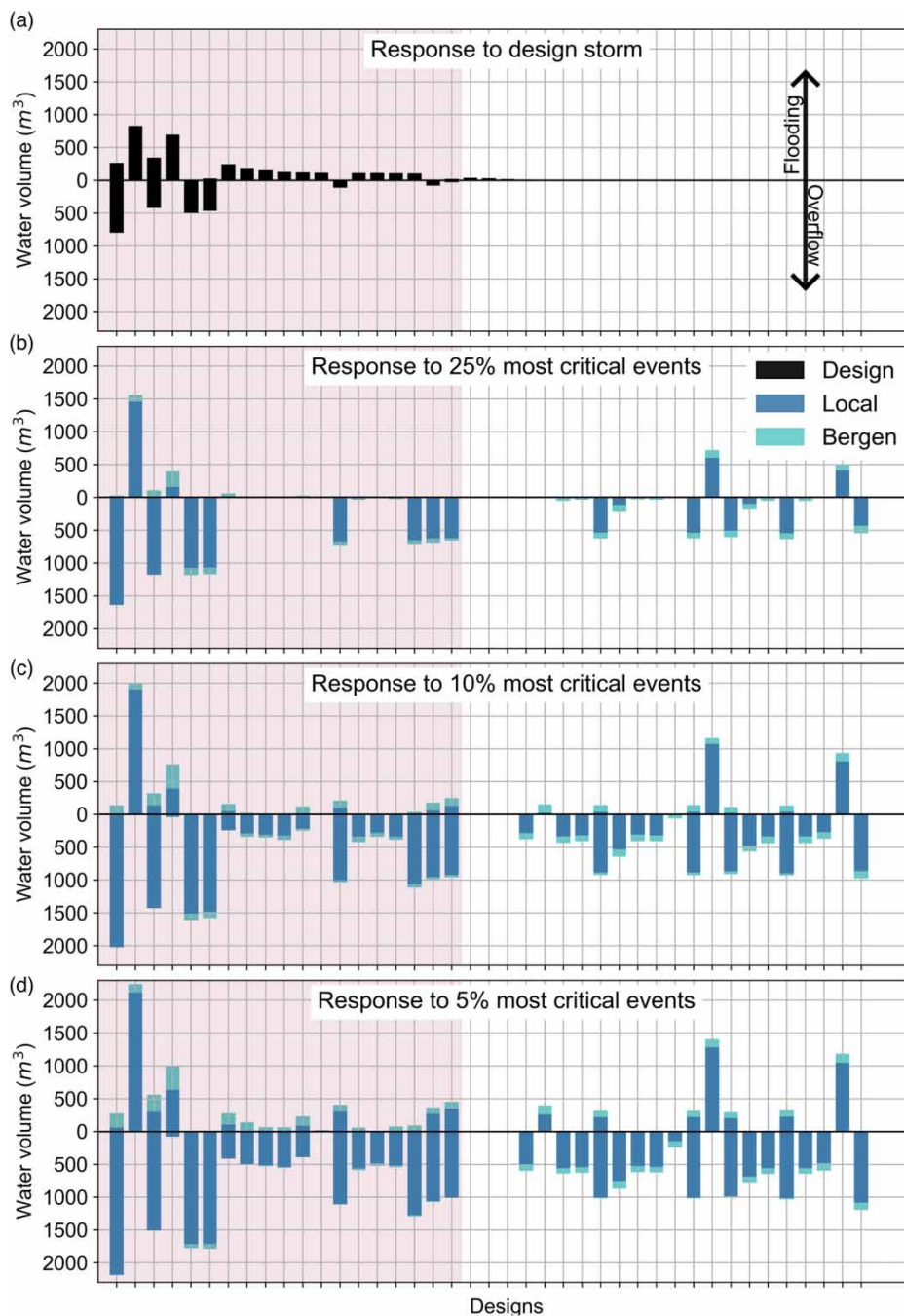


Figure 3 | Hydrological response of UDS designs for (a) the design storm. (b), (c), and (d): 25, 10, and 5% of the most critical events in the ensembles. The designs that produce $V_{tot} > 50 \text{ m}^3$ for the design storm are shown in the shaded area.

reasonable). The result is that these designs flooded (slightly, three out of four had acceptable flooding volumes) when stressed with the design storm, which had a large peak, but showed very robust behaviour when stressed with the ensembles. The mode of failure of nine designs also changes when they are stressed with the ensembles. More specifically, designs that failed by only flooding with the design storm change to either an overflow or a combined failure. The inverse is not observed.

More interestingly, the designs that met the criteria for the design storm show a varying response to the ensembles as well. All but three designs produce large V_{tot} values for the 10 and 5% most critical storms in the ensembles. Even though the

distributions of the median values of V_{FL} and V_{OF} (and V_{tot}) have a lower mean for these designs than for the ones that fail for the design storm (null hypothesis of a permutation test (Good 1994) is rejected at the 5% significance level for the p -value), this difference does not warrant the assumption that a design's performance for the design storm is a sufficient criterion for its overall performance. In short, this means that even when considering only designs that performed well for the design storm, we cannot expect them to perform well for other storms with the same duration and precipitation depth.

It is also clear for these designs that the main mode of failure is overflow. Three designs flooded (without overflow). They produce different flood volumes, but the flooding location for all three was the same – the RB. They have a shared design flaw that was not evident in the design storm simulations. The weir level was at the same height as the maximum depth of the basin. This makes overflow to the river impossible, leading to flooding because of the RB reaching its capacity. Understanding the failure mechanism of these specific cases helps us set them aside from the other instances where flooding occurs. This is because even though they fail due to flooding, the failure is localized at the RB because of a clear design flaw rather than spreading between nodes in the catchment. There are also four designs (out of the 22 that meet the design criteria) that have a combined mode of failure when stressed with the 5% most severe storms of each ensemble. The locations of the flooding points were different for each design, but they all tended to be at nodes near the exit of the catchment. These nodes have a bigger part of the UDS (more than 30% of the total area) contributing to the water volume they receive. The different temporal patterns of the synthetic storms are more probable to stress those locations rather than more upstream ones that are only affected by one or two sub-catchments.

Figure 3 reveals two main vulnerabilities of using a single design storm to design a UDS, the resulting designs are not guaranteed to be able to handle events similar to the design one, and the way that they respond to these events cannot be known. Two dimensions of Figure 3 provide a qualitative overview of these two vulnerabilities. In the horizontal direction, we see how different (equally acceptable) designs respond to the same rainfall event. While in the vertical direction (from (a) to (d)), we see how different rainfall temporal patterns affect these designs.

It is concluded that all the weaknesses highlighted by the ensembles are connected either to the RB or the pipes near the exit of the catchment. This is a result of designs that are tailored to the design storm, dimensioning the pipes and RB, so that they marginally do not fail for the exact conditions created by the design hyetograph. However, when stressed with different temporal patterns (e.g., a storm that has a mild start, filling up a part of the basin, and a late peak, stressing the basin above its capacity), they can fail, even if these patterns have lower peak intensities than the design storm. This highlights a weakness of branched grey infrastructure systems. During the design process, the focus is solely on system reliability and resistance, while resilience and robustness are neglected (Hesarkazzazi *et al.* 2022). We see here that when these systems are subjected to an ensemble of expected storms, encountering conditions that were not simulated in the design process, they can fail.

The storms in the ensemble created using the MRC model that was trained on the Bergen dataset consistently produce larger volumes across all designs, compared to those in the ensemble created using the local MRC model. This could be attributed to the higher peaks observed in the storms of that ensemble (Figure 2).

Another way to visualize and analyse the results is with the use of survival curves (Ng *et al.* 2020; Pons *et al.* 2022a). In this way, we examined the distribution of each design's response to the different storms. More precisely, the probability of exceedance for a given V_{tot} , irrespective of the mode of failure. Figure 4(a) shows the survival curves of all the designs for the SYNTH_{BER} ensemble (the curves for the SYNTH_{LOC} ensemble were almost identical). The V_{tot} produced by the design storm is shown with an 'x' on each curve. The probability that corresponds to that point on a design's curve gives us the percentage of events that produced a larger V_{tot} than the design storm. Reading the curves from this perspective, we see that a design's response to the design storm is not indicative of its robustness against different temporal patterns of rainfall. However, the estimate of the cost function (which essentially calculates a weighted sum of the storage volumes in the pipe network and the storage basin) seems to be correlated with robustness. We can also derive a robustness metric, ρ , from a design's survival curve by choosing a threshold and finding the probability of non-exceedance ($1 - \text{probability of exceedance}$). There are five designs with a ρ below 0.7 for a threshold of 50 m^3 , which responded perfectly to the design storm (no flooding or overflow).

We chose five designs representing different degrees of robustness to compare with each other (Figure 4(b)). The most robust one (lowest curve) is used as a reference design and the other four are compared against it. It is noteworthy that the reference design produced a flood volume of 30 m^3 for the design storm. Design 1 produced a flood volume of 153 m^3 , meaning that it failed even by our updated, more flexible, criteria. While Design 2, which has a worse overall

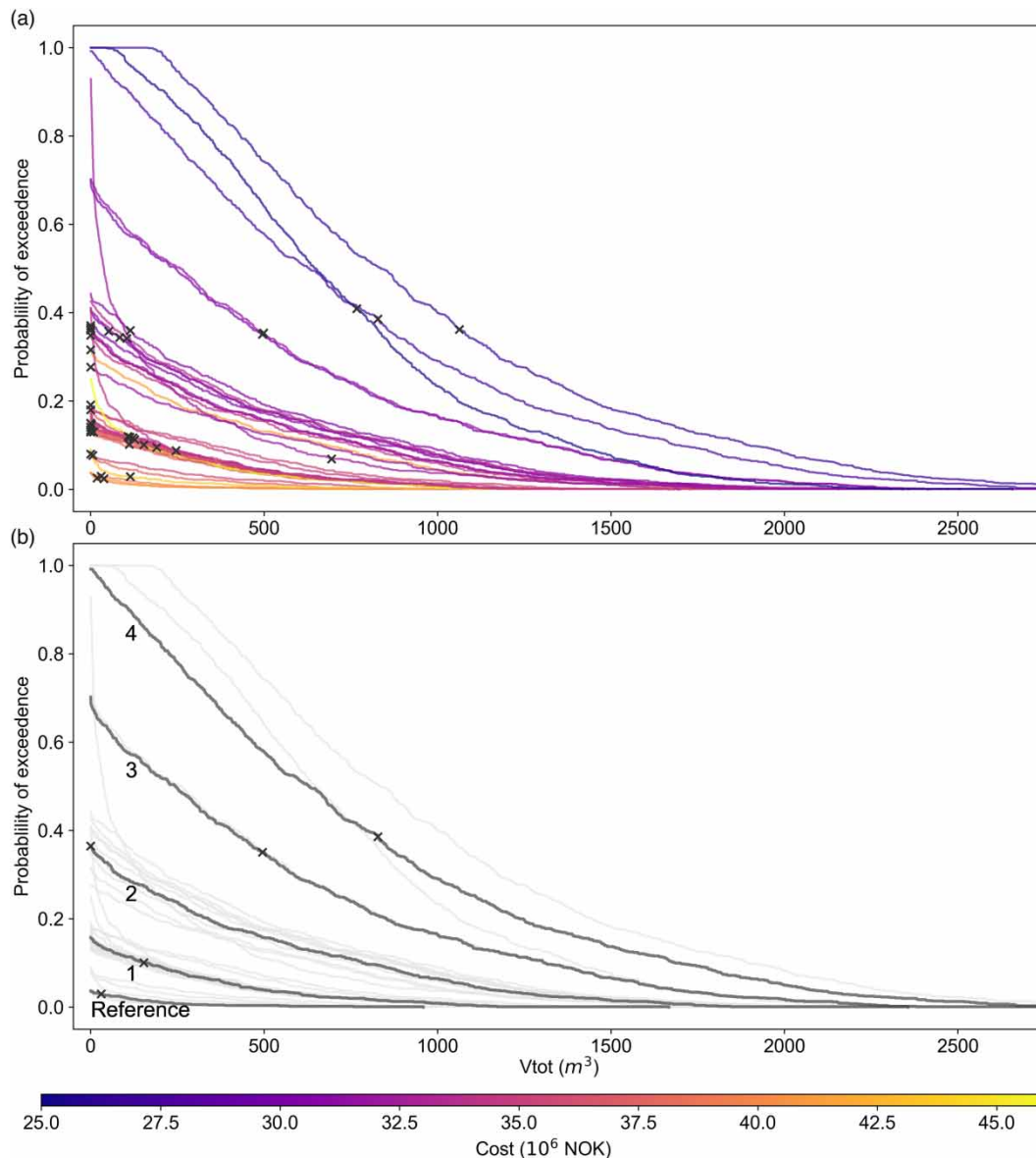


Figure 4 | Survival curves of combined flooding and overflow volume (V_{tot}) for the SYNTH_{BER} ensemble. V_{tot} produced by the design storm is marked with an «x» on each curve. In (b), five designs of interest are highlighted to be further analysed in Figure 5.

response to the ensemble, produced no V_{tot} for the design storm. Designs 3 and 4 both failed for the design storm and showed very low robustness to the ensemble. In general, these five designs represent (to an extent) the different choices made by the designers as they are part of design groups with similar behaviour.

Each subplot in Figure 5 shows the differences in cost, pipe diameters, and RB volume between the chosen designs and the reference design. If an element's dimensions (or cost) in the chosen design are larger than the reference one, it is shown with a negative value, and *vice versa*.

Designs 1, 2 and 4 use larger diameters than the reference design in major parts of the network. This does not improve their performance for any of the simulated storms, as the diameters used at those locations by the reference design are already adequate to route the runoff. The common flaw in all four of the designs is that they use a smaller RB than the reference design. Compared to the reference design, the storage capacity of the RB is decreased by 28, 42, 46, and 64% for Designs 1–4, respectively. They also use smaller pipes in some downstream locations. These are the two reasons Designs 1–4 are, in varying degrees, less robust than the reference design.

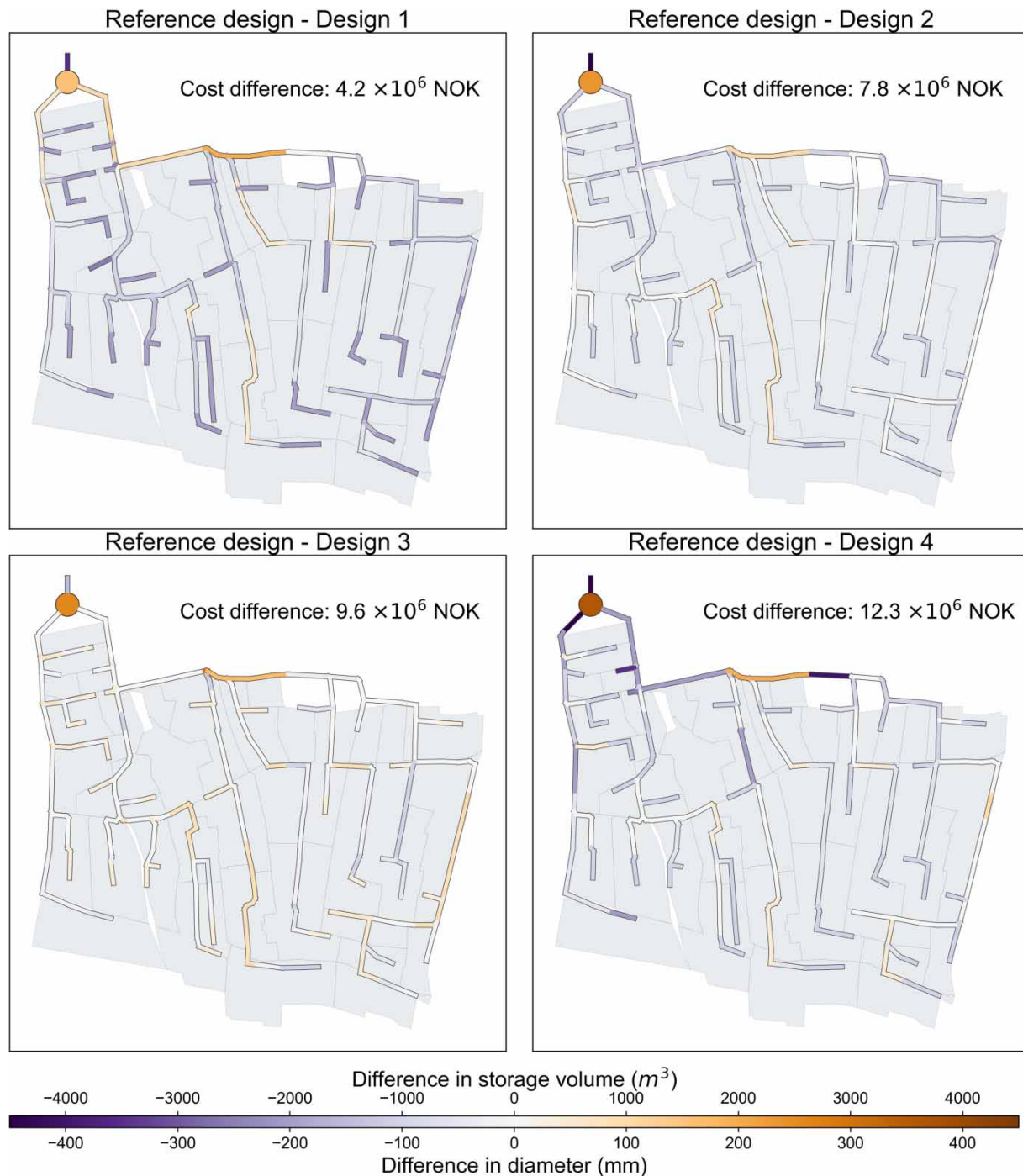


Figure 5 | Pipe diameter and basin volume comparison between designs highlighted in Figure 4(b). Negative values show that the dimensions of the selected design are larger than the reference design.

Design 4 has a combined mode of failure, producing V_{tot} above the acceptable threshold for the original design storm as well as most of the storms in the ensembles. This is because most of the downstream pipes as well as the RB are underdimensioned.

The effect of the designers' subjectivity during the dimensioning of the systems is made especially clear between the reference design and Design 2. Even though both systems performed well for the design storm, one designer (reference design)

opted for a 74% larger RB and overall slightly smaller pipe diameters compared to Design 2. This resulted in a system that is almost 8 million Norwegian kroner (approximately 800,000 euros) more expensive but also considerably more robust (see Figure 4).

Limitations and outlook

This work is a case study that uses modelling results from one catchment to illustrate vulnerabilities of the design approach used for the UDS. Consequently, the results have limited generalizability. However, they may still serve as an example of the presence and source of structural uncertainty in the designs we studied. These designs are created by MSc students and not by real practitioners, but, given an explicit design approach and a limited decision space (fixed system topology, predefined sub-catchment delineation and properties, and no GIs), they can be used as proxies for designer uncertainty.

Furthermore, no possible malfunctions of the UDS (pipe blockages, RB sedimentation, etc.) are considered in the simulations. Meaning that one dimension of robustness is overlooked, in order to more specifically focus on the factors we consider (i.e., the hyetograph shape and design choices). That is to say that the modes of failure studied are only functional, and no possible structural and operational modes of failure are considered (Butler *et al.* 2017).

Future research could aim to introduce more designer freedom (e.g., by combining the design approach with a GI placement framework) to investigate the effect of this added uncertainty on the resulting designs' robustness. As we move towards more GI adaptation and resilient designs, there will be a need to better understand the effect of all parts of our design approaches and adapt our methods (Funke *et al.* 2021; Markolf *et al.* 2021). The effect of rainfall's spatial variability could also be investigated in the context of UDS design focusing on larger urban catchments since its effect on urban hydrodynamic modelling is already documented (Ochoa-Rodriguez *et al.* 2015). Future climate predictions can also be utilized by training the MRC model on 'future' time series to create rainfall events based on data accounting for climate change. That would, of course, introduce a new set of uncertainties to the event ensemble, which should not be taken lightly. More generally, the effect that uncertainty has on the robustness of UDS designs could be further investigated. For example, in the choice of design approach, quantifying and decomposing the structural uncertainty inherent in different guidelines could be the first step to reducing it. Furthermore, the effect of the choice of software used could be investigated in future work (similar to Dotto *et al.* (2011)). Regarding the choice of simulation input data, we have shown that a single design storm might be insufficient when attempting to describe the behaviour of a whole catchment. Using an ensemble of design storms and setting an acceptable probability of exceedance are possible solutions. To create this ensemble, especially in cases that require considerable computational power, evaluating an initial set of events to arrive at a relevant subset to be used in the design process might be necessary.

CONCLUSIONS

The results of this study show that following common engineering practices and designing a UDS using a single design storm do not consistently produce a robust system. This echoes the conclusion of studies done in the last years (Watt & Marsalek 2013; Ng *et al.* 2020; Markolf *et al.* 2021), which demonstrate, in different ways, the shortcomings of the current implementations of design storms. This points towards the need to incorporate robust decision-making (RDM) in our UDS design approaches. RDM is an approach that aims to help arrive at robust solutions, given a problem that involves deep uncertainties (Walker *et al.* 2013). Typically, an RDM process consists of first defining the scope of variables and uncertainties to consider and simulating the response of the system to an ensemble of scenarios representing a variety of conditions within the scope. Then, the results are analysed, and vulnerabilities of the initial system are identified and confronted by adapting its design framework.

In this study, we narrowed down the scope to two influencing factors: (i) the influence of the hyetograph shape on the performance of the designed networks and (ii) the structural uncertainty inherent in the subjectivity of the designers' choice when defining the network qualities. The results show a lack of robustness that is not aligned with the level of trust placed in current UDS design guidelines. By necessity, the vulnerabilities we can spot originate from the influencing factors we chose to investigate. This means that the lack of robustness we observed in our analysis is a product of two vulnerabilities: (i) the inability of one hyetograph shape to represent the entire range of events that should be anticipated and (ii) the lack of performance consistency in the designs produced by the design approach.

These issues can be confronted by adapting the design approach. For example, introducing an ensemble of events could provide a more rigorous (and more realistic) benchmark for the designs. In this direction, of applying a wider range of conditions to the designs, simulating multiple successive events can also illuminate vulnerabilities in a system (especially for

systems with high retention or detention capacities (Adams & Howard 1986)). Finally, aiming for more resilient designs (e.g., by requiring that the network has redundant connections to avoid bottlenecks) can lead to better consistency and more robustness (Casal-Campos *et al.* 2018; Hesarkazzazi *et al.* 2022).

ACKNOWLEDGEMENTS

The authors are grateful to the anonymous reviewers for their constructive feedback leading to an improved manuscript and to the anonymized students for their invaluable contribution to this work.

DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

- Adams, B. J. & Howard, C. D. D. 1986 *Design storm pathology*. *Canadian Water Resources Journal* **11**, 49–55. <https://doi.org/10.4296/cwrj1103049>.
- Alfieri, L., Laio, F. & Claps, P. 2008 *A simulation experiment for optimal design hyetograph selection*. *Hydrological Processes* **22**, 813–820. <https://doi.org/10.1002/hyp.6646>.
- American Society of Civil Engineers 2013 *Standard Guidelines for the Design, Installation, and Operation and Maintenance of Urban Subsurface Drainage*. American Society of Civil Engineers, Reston, VA, USA. <https://doi.org/10.1061/9780784413050>.
- Arnbjerg-Nielsen, K. 2012 *Quantification of climate change effects on extreme precipitation used for high resolution hydrologic design*. *Urban Water Journal* **9**, 57–65. <https://doi.org/10.1080/1573062X.2011.630091>.
- Ball, J., Babister, M., Nathan, R., Weeks, W., Weinmann, E., Retallick, M. & Testoni, I. 2019 *Australian Rainfall and Runoff: A Guide to Flood Estimation*, © Commonwealth of Australia (Geoscience Australia).
- Beven, K. J. 2011 *Rainfall-Runoff Modelling: The Primer*. John Wiley & Sons, Chichester, UK.
- Billinton, R. & Allan, R. N. 1992 *Reliability Evaluation of Engineering Systems*. Springer US, Boston, MA, USA. <https://doi.org/10.1007/978-1-4899-0685-4>.
- Blanchard, B. S. & Blyler, J. E. 2016 *System Engineering Management*, 5th edn. Wiley Series in System Engineering and Management. Wiley, Hoboken, NJ, USA.
- British Standards Institution (BSI) 2017 BS EN 752:2017. Drain and sewer systems outside buildings. Sewer system management.
- Butler, D., Ward, S., Sweetapple, C., Astaraie-Imani, M., Diao, K., Farmani, R. & Fu, G. 2017 *Reliable, resilient and sustainable water management: The Safe & SuRe approach*. *Global Challenges* **1**, 63–77. <https://doi.org/10.1002/gch2.1010>.
- Casal-Campos, A., Sadr, S. M. K., Fu, G. & Butler, D. 2018 *Reliable, resilient and sustainable urban drainage systems: An analysis of robustness under deep uncertainty*. *Environmental Science Technology* **52**, 9008–9021. <https://doi.org/10.1021/acs.est.8b01193>.
- Chow, V. T., Maidment, D. R. & Mays, L. W. 1988 *Applied Hydrology*. McGraw-Hill Series in Water Resources and Environmental Engineering. McGraw-Hill, Singapore.
- Cordery, I., Pilgrim, D. H. & Rowbottom, I. A. 1984 *Time patterns of rainfall for estimating design floods on a frequency basis*. *Water Science and Technology* **16**, 155–165. <https://doi.org/10.2166/wst.1984.0187>.
- Craye, M., Funtowicz, S. & Sluijs, J. P. V. D. 2005 *A reflexive approach to dealing with uncertainties in environmental health risk science and policy*. *International Journal of Risk Assessment and Management* **5**, 216. <https://doi.org/10.1504/IJRAM.2005.007169>.
- Deletic, A., Dotto, C. B. S., McCarthy, D. T., Kleidorfer, M., Freni, G., Mannina, G., Uhl, M., Henrichs, M., Fletcher, T. D., Rauch, W., Bertrand-Krajewski, J. L. & Tait, S. 2012 *Assessing uncertainties in urban drainage models*. *Physics and Chemistry of the Earth, Parts A/B/C, Estimating and Representing Uncertainty in Applied Hydrology, Hydraulics and Water Quality Studies* **42–44**, 3–10. <https://doi.org/10.1016/j.pce.2011.04.007>.
- De Toffol, S., Engelhard, C. & Rauch, W. 2007 *Combined sewer system versus separate system – A comparison of ecological and economical performance indicators*. *Water Science and Technology* **55**, 255–264. <https://doi.org/10.2166/wst.2007.116>.
- Dotto, C. B. S., Kleidorfer, M., Deletic, A., Rauch, W., McCarthy, D. T. & Fletcher, T. D. 2011 *Performance and sensitivity analysis of stormwater models using a Bayesian approach and long-term high resolution data*. *Environmental Modelling & Software* **26**, 1225–1239. <https://doi.org/10.1016/j.envsoft.2011.03.013>.
- Dotto, C. B. S., Kleidorfer, M., Deletic, A., Rauch, W. & McCarthy, D. T. 2014 *Impacts of measured data uncertainty on urban stormwater models*. *Journal of Hydrology* **508**, 28–42. <https://doi.org/10.1016/j.jhydrol.2013.10.025>.
- DWA-A 118E 2006 *Hydraulic Dimensioning and Verification of Drain and Sewer Systems*.
- Funke, F., Matzinger, A. & Kleidorfer, M. 2021 Sensitivity of sustainable urban drainage systems to precipitation events and malfunctions based on 60-year long-term modelling. In *Presented at the ICUD 2021*.

- Gironás, J., Roesner, L. A., Rossman, L. A. & Davis, J. 2010 A new applications manual for the Storm Water Management Model (SWMM). *Environmental Modelling & Software* **25**, 813–814. <https://doi.org/10.1016/j.envsoft.2009.11.009>.
- Good, P. 1994 *Permutation Tests*, Springer Series in Statistics. Springer, New York, NY, USA. <https://doi.org/10.1007/978-1-4757-2346-5>.
- Guo, Y. & Adams, B. J. 1998 Hydrologic analysis of urban catchments with event-based probabilistic models: 1. Runoff volume. *Water Resources Research* **34**, 3421–3431. <https://doi.org/10.1029/98WR02449>.
- Guo, Y. & Zhuge, Z. 2008 Analytical probabilistic flood routing for urban stormwater management purposes. *Canadian Journal of Civil Engineering* **35**, 487–499. <https://doi.org/10.1139/L07-131>.
- Hauger, M. B., Mouchel, J.-M. & Mikkelsen, P. S. 2006 Indicators of hazard, vulnerability and risk in urban drainage. *Water Science and Technology* **54**, 441–450. <https://doi.org/10.2166/wst.2006.622>.
- Hesarkazzazi, S., Bakhshpour, A. E., Hajibabaei, M., Dittmer, U., Haghghi, A. & Sitzenfrei, R. 2022 Battle of centralized and decentralized urban stormwater networks: From redundancy perspective. *Water Research* **222**, 118910. <https://doi.org/10.1016/j.watres.2022.118910>.
- Institute of Public Works Engineering Australasia (IPWEA), Queensland Division 2016 Queensland Urban Drainage Manual, Fourth Edition.
- Keifer, C. J. & Chu, H. H. 1957 Synthetic storm pattern for drainage design. *Journal of the Hydraulics Division* **83**(4). <https://doi.org/10.1061/JYCEAJ.0000104>.
- Kilgore, R., Atayee, A. T. & Herrmann, G. R. 2024 *Urban Drainage Design*, Fourth Edition. United States Federal Highway Administration. Office of Bridges and Structures, Denver, CO, USA.
- Koutsoyiannis, D. 2021 *Στοχαστική των Ακραίων Υδροκλιματικών Φαινομένων - Μια ψύχραμη ματιά στη διακινδύνευση (Stochastics of Hydroclimatic Extremes – A Cool Look at Risk)*. Open Academic Editions, Kallipos. <https://doi.org/10.57713/KALLIPOS-1>.
- Maidment, D. R. 1993 *Handbook of Hydrology*. McGraw-Hill, New York, NY, USA.
- Markolf, S. A., Chester, M. V., Helmrich, A. M. & Shannon, K. 2021 Re-imagining design storm criteria for the challenges of the 21st century. *Cities* **109**, 102981. <https://doi.org/10.1016/j.cities.2020.102981>.
- Mugume, S. N. & Butler, D. 2017 Evaluation of functional resilience in urban drainage and flood management systems using a global analysis approach. *Urban Water Journal* **14**, 727–736. <https://doi.org/10.1080/1573062X.2016.1253754>.
- Mugume, S. N., Diao, K., Astaraie-Imani, M., Fu, G., Farmani, R. & Butler, D. 2015 Enhancing resilience in urban water systems for future cities. *Water Supply* **15**, 1343–1352. <https://doi.org/10.2166/ws.2015.098>.
- Ng, J. Y., Fazlollahi, S. & Galelli, S. 2020 Do design storms yield robust drainage systems? How rainfall duration, intensity, and profile can affect drainage performance. *Journal of Water Resources Planning and Management* **146**, 04020003. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001167](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001167).
- Ochoa-Rodriguez, S., Wang, L.-P., Gires, A., Pina, R. D., Reinoso-Rondinel, R., Bruni, G., Ichiba, A., Gaitan, S., Cristiano, E., van Assel, J., Kroll, S., Murlà-Tuyls, D., Tisserand, B., Schertzer, D., Tchiguirinskaia, I., Onof, C., Willems, P. & ten Veldhuis, M.-C. 2015 Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: A multi-catchment investigation. *Journal of Hydrology, Hydrologic Applications of Weather Radar* **531**, 389–407. <https://doi.org/10.1016/j.jhydrol.2015.05.035>.
- Pichler, M. 2022 *swmm-api: API for Reading, Manipulating and Running SWMM-Projects with Python*. <https://doi.org/10.5281/zenodo.7054804>.
- Pons, V., Benestad, R., Sivertsen, E., Muthanna, T. M. & Bertrand-Krajewski, J.-L. 2022a Forecasting green roof detention performance by temporal downscaling of precipitation time-series projections. *Hydrology and Earth System Sciences* **26**, 2855–2874. <https://doi.org/10.5194/hess-26-2855-2022>.
- Pons, V., Muthanna, T. M., Sivertsen, E. & Bertrand-Krajewski, J.-L. 2022b Revising green roof design methods with downscaling model of rainfall time series. *Water Science and Technology* **85**, 1363–1371. <https://doi.org/10.2166/wst.2022.023>.
- Refsgaard, J. C., van der Sluijs, J. P., Brown, J. & van der Keur, P. 2006 A framework for dealing with uncertainty due to model structure error. *Advances in Water Resources* **29**, 1586–1597. <https://doi.org/10.1016/j.advwatres.2005.11.013>.
- SIS – Bygg och anläggning 2017 *Drain and Sewer Systems Outside Buildings – Sewer System Management (No. SS-EN 752:2017)*.
- Sitzenfrei, R., Urich, C., Möderl, M. & Rauch, W. 2013 Assessing the efficiency of different CSO positions based on network graph characteristics. *Water Science and Technology* **67**, 1574–1580. <https://doi.org/10.2166/wst.2013.029>.
- Tahmasebi Birgani, Y. & Yazdandoost, F. 2014 A framework for evaluating the persistence of urban drainage risk management systems. *Journal of Hydro-Environment Research* **8**, 330–342. <https://doi.org/10.1016/j.jher.2014.04.004>.
- Tscheikner-Gratl, F., Bellos, V., Schellart, A., Moreno-Rodenas, A., Muthusamy, M., Langeveld, J., Clemens, F., Benedetti, L., Rico-Ramirez, M. A., de Carvalho, R. F., Breuer, L., Shucksmith, J., Heuvelink, G. B. M. & Tait, S. 2019 Recent insights on uncertainties present in integrated catchment water quality modelling. *Water Research* **150**, 368–379. <https://doi.org/10.1016/j.watres.2018.11.079>.
- Underwood, B. S., Mascaro, G., Chester, M. V., Fraser, A., Lopez-Cantu, T. & Samaras, C. 2020 Past and present design practices and uncertainty in climate projections are challenges for designing infrastructure to future conditions. *Journal of Infrastructure Systems* **26**, 04020026. [https://doi.org/10.1061/\(ASCE\)IS.1943-555X.0000567](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000567).
- Veneziano, D. & Villani, P. 1999 Best linear unbiased design hyetograph. *Water Resources Research* **35**, 2725–2738. <https://doi.org/10.1029/1999WR900156>.
- Walker, W. E., Haasnoot, M. & Kwakkel, J. H. 2013 Adapt or perish: A review of planning approaches for adaptation under deep uncertainty. *Sustainability* **5**, 955–979. <https://doi.org/10.3390/su5030955>.
- Watt, E. & Marsalek, J. 2013 Critical review of the evolution of the design storm event concept. *Canadian Journal of Civil Engineering* **40**, 105–113. <https://doi.org/10.1139/cjce-2011-0594>.

- Willems, P. 2012 [Model uncertainty analysis by variance decomposition](#). *Physics and Chemistry of the Earth, Parts A/B/C* **42–44**, 21–30. <https://doi.org/10.1016/j.pce.2011.07.003>.
- Wright, D. B., Smith, J. A. & Baeck, M. L. 2014 [Flood frequency analysis using radar rainfall fields and stochastic storm transposition](#). *Water Resources Research* **50**, 1592–1615. <https://doi.org/10.1002/2013WR014224>.

First received 15 March 2024; accepted in revised form 8 July 2024. Available online 26 July 2024