Water Resources Management
https://doi.org/10.1007/s11269-025-04196-5

Probabilistic Models for Optimal Rainwater Harvesting Tank Sizing: a Comparison with Traditional Approaches

Maria Gloria Di Chiano[1] · **Mariana Marchioni**[1,2] · **Gianfranco Becciu**[1]

Received: 30 October 2024 / Accepted: 20 March 2025

Abstract

With growing water scarcity and increasing variability in rainfall patterns due to climate change, along with the pressures of urbanization and population growth, the adoption of non-conventional water resources such as rainwater harvesting and water reuse has become a crucial sustainable strategy. Traditional approaches to the design of rainwater tanks, based on the demand volume in a predefined dry period, don't take into account properly the risk of failure. Two probabilistic methods are then proposed, to address this issue. The first method is a modification of the conventional "demand-side" approach, by the use of a probabilistic estimation of the inter-event time. The second one, more reliable and more complex, takes into account in a parametric way the full stochastic rainfall process, allowing to consider also the pre-filling possibility due to consecutive storm events. Complexity in this second method is managed in order to develop direct relationships for practical applications. Although enough simple, these methods improve the design of Rainwater Harvesting Systems (RWHs), allowing to apply cost-benefit analysis procedures. Methods are compared and evaluated through the application to a case study in Milan, Italy.

Keywords Rainwater tanks design · Alternative water resources · Water sustainability · Probabilistic models · Stochastic rainfall process

1 Introduction

The United Nations' 2030 Agenda (United Nations 2024) underscores the essential role of water in achieving the Sustainable Development Goals (SDGs), particularly the aim of universal access to clean water and sanitation by 2030 (SDG 6). Attaining this goal requires sustainable water management practices that balance current needs with future demands

✉ Maria Gloria Di Chiano
mariagloria.dichiano@polimi.it

[1] D.I.C.A. Politecnico di Milano, Milano 20133, Italy

[2] Montana S.p.A, Milano 20143, Italy

(Mishra et al. 2021). Rainwater harvesting (RWH) presents a viable solution, reducing dependence on conventional water sources and enhancing urban water resilience (De Sá Silva et al. 2022; Nandi and Gonela 2022; Khan 2023).

In urban areas, harvested rainwater can be used for various non-potable purposes, including irrigation, toilet flushing, cooling, fire suppression, laundry, and street cleaning (Hviid et al. 2020). When combined with advanced treatments like membrane filtration, reverse osmosis, and UV disinfection, it can also meet potable water requirements, making it an essential resource in regions facing severe water scarcity (Adler et al. 2011; Teixeira and Ghisi 2019; Uddin Sikder 2020). Widely adopted in countries like Germany, Japan, and Australia, where policies and incentives promote their use, RWHs have expanded beyond private households to public buildings, schools, offices, and airports, highlighting their effectiveness and versatility (Wartalska et al. 2024). RWH also mitigates urban flooding by capturing and retaining runoff, thereby alleviating pressure on drainage systems and reducing flood risks (Jamali et al. 2020; Hdeib and Aouad 2023). In RWHs, the choice of tank design significantly influences the system's ability to meet water demand efficiently while minimizing waste.

Among the existing methodologies for tank design, traditional empirical models and probabilistic models can be identified.

Traditional models, including demand-side and supply-side approaches, are widely used for preliminary tank sizing due to their simplicity and minimal data requirements (Allen and Haarhoff 2015) The demand-side approach estimates tank size based on water demand during the longest dry period, whereas the supply-side approach focuses on capturing all available runoff during the wet season. However, these methods rely on simplified assumptions that overlook the stochastic nature of rainfall, often leading to conservative designs that overestimate storage needs (Becciu et al. 2018; Di Chiano et al. 2023). A key limitation of non-probabilistic methods is their inability to explicitly calculate the risk of insufficiency. Since they lack probabilistic analysis, these models depend on parameters derived from past observational data, making their reliability highly sensitive to the quantity and quality of available records. While the accuracy of such estimates improves with larger datasets, the residual risk of insufficiency remains unknown. Moreover, empirical models based on demand satisfaction assume that wet periods always provide sufficient precipitation to meet future dry-period demands, an assumption that holds true only in climates with a distinct alternation between wet and dry seasons. Despite these constraints, traditional models remain applicable in contexts where specific hydrological data are unavailable, making them a practical choice for preliminary assessments.

Probabilistic approaches in hydrological modeling define the probability distribution of output variables by leveraging the distributions of hydrological inputs. These methods have been applied to various Sustainable Drainage Systems (SUDS), including RWH systems, permeable pavements, green roofs, and Highway Filter Drains (HFDs) (Raimondi et al. 2022, 2023; Marchioni et al. 2023; Essien et al. 2024). Regarding probabilistic approaches specific to RWHs, Kim et al. (2012) introduced a probabilistic-analytical approach for estimating runoff reduction in RWHs by incorporating catchment characteristics, tank storage, and infiltration. Notably, existing probabilistic models typically consider only two consecutive rainfall events, which may limit their applicability.

This study proposes two probabilistic approaches for designing tanks in RWHs. The first method enhances the traditional demand-side approach by integrating the statistical distribution of inter-event times into the tank design process. Unlike conventional models that concentrate solely on dry periods, the second model considers the complete stochastic process of rainfall events, capturing cumulative deficits during periods of insufficient rainfall. This advancement addresses also the limitations of existing probabilistic models, which typically consider only a couple of consecutive rainfall events (Raimondi and Becciu 2014). This allows for the design of tanks that accommodate maximum variations in stored water volumes during these deficit periods, offering a more comprehensive and reliable assessment compared to current methods.

2 Methodology

2.1 Traditional Demand-Side Approaches

Traditional methods often employ simplified deterministic approaches that rely on historical rainfall records, assuming a constant daily water demand. These methods estimate storage requirements based on observed dry periods, typically following one of two approaches: the Maximum Dry Period Approach or the Average Dry Period Approach. The Maximum Dry Period Approach determines tank volume as:

$$W_{tank} = D \cdot T_{Max,Dry\ Period} \tag{1}$$

where D is the average demand for non-potable use (mm/day) and $T_{Max,\ Dry\ Period}$ (days) represents the duration of the longest dry period observed in the historical precipitation record. This method assumes the tank is full at the onset of a drought, but it does not account for the variability of rainfall leading up to the dry period.

A less conservative alternative is the Average Dry Period Approach, which determines the required tank size based on the historical mean duration of dry periods. In this case, the tank volume is given by:

$$W_{tank} = D \cdot T_{Average,Dry\ Period} \tag{2}$$

where D is the average demand for non-potable use (mm/day) and $T_{Average\ Dry\ Period}$ (days) represents the duration of the average dry period observed in the historical precipitation record. This method optimizes storage capacity for average conditions rather than extremes, generally resulting in lower storage requirements. However, reliance on average values may underestimate storage needs for longer-than-average dry spells and fails to account for climate variability or the stochastic nature of rainfall, limiting its reliability in dynamic climatic conditions.

While both approaches provide practical and accessible solutions, their deterministic nature disregards rainfall variability and probability distributions, necessitating probabilistic modelling for more accurate tank design.

2.2 Probabilistic Demand-Side Approach

The first probabilistic methodology proposed in this work builds on an adaptation of the traditional demand-side approach, incorporating probabilistic elements to account for the inherent variability in dry periods. Instead of relying on fixed values for the maximum or average dry periods, this method uses the statistical distribution of inter-event times derived from historical data to introduce a risk-based approach. In particular, the tank volume is calculated as:

$$W_{tank} = D \cdot T_q \tag{3}$$

where D (mm/day) is the constant daily water demand, and T_q (days) is the duration of the dry period corresponding to the chosen quantile. This approach allows for a more flexible and tailored design, enabling the tank size to align with the desired level of non-exceedance probability for a given dry period duration under varying climatic conditions. Formally, the determination of the dry period quantile T_q is made deriving the empirical cumulative distribution function of the inter-event time from the historical data and then identifying the theoretical distribution that best approximates the empirical behavior. Probabilistic models such as the exponential, Weibull, lognormal, or others that well describe the behaviour of the historical series can be used. The selection of the best-fitting distribution can be validated through statistical goodness-of-fit tests, which compare observed data with the theoretical model. Once the best-fitting distribution has been identified, the designer can select the dry period quantile corresponding to the desired non-exceedance probability level.

2.3 Stochastic Approach Based on the Cumulative Deficit Distribution Over *n* Consecutive Independent Rainfall Events

The probabilistic approach based on the cumulative deficit distribution over *n* consecutive independent rainfall events takes into account both rainy and dry periods in succession, capturing the variability of the process over time and enabling the evaluation of the cumulative distribution function of the active storage volume of a generic rainwater tank, which represents the minimum storage capacity required to accommodate the maximum variations in stored water volumes over a given period T. The approach proposed in this work refers to a tank in conditions of overall surplus, meaning that the design is carried out under the assumption that, over the reference period, the total volume of water available exceeds the water demands. In these conditions of global surplus, the required storage volume is given by the maximum difference between cumulative outflows and inflows during deficit sub-periods.

The starting point of this methodology is the transformation of continuous historical rainfall data into independent discrete events using the Inter-Event Time Definition (IETD), a threshold parameter that separates consecutive rainfall occurrences based on a minimum period of no precipitation (Lee and Kim 2018; Tu et al. 2023). If the interval between two rainfall instances exceeds the IETD, they are classified as distinct events; otherwise, they are considered part of the same event. In this way, each rainfall event is defined by its duration, the dry period separating it from the next event, and the recorded rainfall depth. Among the methods for determining the IETD, autocorrelation analysis defines it as the time lag at which the autocorrelation coefficient of rainfall pulses reaches zero, indicating statistical independence (Joo et al. 2013).

After identifying independent rainfall events, the random variable Δ_i is introduced to represent the difference between water demand and the cumulative rainfall depth during an event and is calculated as follows:

$$\Delta_i = [D \cdot (\theta_i + d_i) - h_i] \tag{4}$$

where D is the average water demand (mm/day), θ_i is the ith rainfall event duration (days), d_i is the ith inter-event time (days) and h_i is the ith rainfall event depth (mm).

This variable can take on both positive and negative values, depending on the rainfall conditions and water demand. A positive value of Δ_i indicates a deficit, meaning that the amount of water collected is insufficient to meet the water demand. Conversely, a negative value of Δ_i represents a surplus, indicating an excess amount of water beyond what is required. Once the variable Δ_i is defined, the next step is to calculate its mean and variance. To calculate the mean of the variable Δ_i, the inter-event time duration, the rainfall event duration, and the rainfall event depth are considered independent random variables. This assumption implies that the occurrence of one variable does not influence the probability of the others, allowing each component to be analyzed separately when determining the mean. In these conditions the expected value of the random variable Δ_i, i.e. $E[\Delta_i]$, can be evaluated as follows:

$$E[\Delta_i] = D \cdot (\mu_\theta + \mu_d) - \mu_h \tag{5}$$

Where D is the average non-potable water demand (mm/day), μ_θ, μ_d, μ_h are sample mean rainfall event duration (days), sample mean inter-event time (days) and sample mean rainfall event depth (mm), respectively. The linear expectation property plays a crucial role in this context. It states that the expectation (mean) of a sum of random variables is equal to the sum of their individual expectations. This property allows the expected values of the dry period duration and the rain event duration to be directly combined to determine their overall influence on the mean deficit or surplus. Additionally, the assumption of stationarity is essential, as it implies that the statistical characteristics of these variables, such as their mean and variance, remain consistent over time. This consistency ensures that the historical data used to calculate these expectations accurately represents future conditions, thereby making the estimates more reliable for design purposes. If stationarity were not present, the calculated mean might not accurately capture the average deficit or surplus over the long term, reducing the effectiveness of this approach.

To further analyze the behavior of the random variable Δ_i, it's crucial to determine its variance, which measures the degree of dispersion or variability of the deficit or surplus around its mean. The variance of Δ_i, denoted as $Var[\Delta_i]$, can be calculated as:

$$Var[\Delta_i] = D^2 \cdot \left(\sigma_\theta^2 + \sigma_d^2\right) + \sigma_h^2 \tag{6}$$

Where σ_θ^2, σ_d^2, σ_h^2 are sample variance of the same meteorological random variables. With the mean and variance of Δ_i established, the analysis is extended to consider sequences of multiple consecutive rainfall events. In this context, the variable $\varDelta_n$ is introduced to represent the cumulative deficit over *n* consecutive independent events:

$$\varDelta_n = \sum_{i=1}^{n} [D \cdot (\theta_i + d_i) - h_i] \tag{7}$$

This variable is used to capture the long-term behavior of the deficit, considering not just isolated events but sequences that can collectively impact the water balance in the tank. The mean and variance of $\varDelta_n$ are calculated using the following formulas:

$$E[\varDelta] = n \cdot E[\varDelta_i] \tag{8}$$

$$Var\ [\varDelta_n] = n \cdot Var\ [\varDelta_i] \tag{9}$$

These expressions rely on the principle that, for a sum of *n* independent and identically distributed (i.i.d.) random variables, the total mean equals *n* times the mean of each individual variable, and the total variance equals *n* times the variance of each individual variable. This principle facilitates scaling the results from single events to sequences of events, enabling an evaluation of the cumulative effect of multiple deficits over time.

The random variable $\varDelta_n$ is assumed to be normally distributed. This assumption is grounded in the Central Limit Theorem (CLT), which indicates that the sum of a large number of independent and identically distributed random variables will approximate a normal distribution, regardless of the individual distributions of these variables, as the number of variables increases (Kottegoda 2008). The variable $\varDelta_n$ can exhibit both positive and negative values, depending on the relationship between rainfall and water demand. For tank design purposes in global surplus conditions, the primary interest lies in periods of deficit, where the cumulative deficit is greater than zero. To specifically address these conditions, a new variable, denoted as $\varDelta_+$ is introduced to represent only the positive values of the cumulative deficit $\varDelta_n$. To focus solely on these positive values, a one-sided truncation of the normal distribution is applied. This truncation ensures that the analysis is concentrated exclusively on instances where $\varDelta_n > 0$, enabling a precise evaluation of deficit periods. The expressions for the mean and variance of a truncated normal distribution, when considering only values greater than 0, are formulated as follows:

$$E[\varDelta_n | \varDelta_n > 0] = E[\varDelta_+] = E[\varDelta_n] + \sqrt{Var\ [\varDelta_n]} \cdot \frac{\phi(\alpha)}{Z} \tag{10}$$

Where $E[\varDelta_n]$ and $\sqrt{Var\ [\varDelta_n]}$ are the expected value and the Standard Deviation of the original variable and φ(α) is the probability density function of the standard normal distribution. The parameter α and the variable Z can be evaluated as:

$$\alpha = \frac{-E[\triangle_n]}{\sqrt{Var\ [\triangle_n]}} \tag{11}$$

$$Z = 1 - \Phi(\alpha) \tag{12}$$

Where Φ(α) represents the cumulative distribution function of the standard normal distribution.

The variance can be obtained as follows:

$$Var\left[\Delta_n|\Delta_n > 0\right] = Var\left[\Delta_+\right] = Var\ \left[\Delta_n\right] \cdot \left[1 + \alpha \cdot \frac{\phi\ (\alpha\ )}{Z} - \left(\frac{\phi\ (\alpha\ )}{Z}\right)^2\right] \quad (13)$$

After obtaining the main statistics of the variable Δ_+, the next step involves modeling the random variable Δ_+ using an exponential distribution. So, its Cumulative Distribution function (CDF) can be expressed as:

$$F_{W_{tank}}\ (x) = F_{\Delta_+}\ (x) = 1 - exp(-\lambda\, x) \quad (14)$$

Where $x \geq 0$ and $\lambda = \frac{1}{E[\Delta_+]}$ with $E\left[\Delta_+\right]$ being the mean of the truncated normal distribution calculated previously. The parameter of this exponential distribution, denoted as λ, is derived directly from the mean of the truncated normal distribution, ensuring that it reflects the statistical characteristics of Δ_+. Since the active storage volume represents the minimum volume required to account for the maximum variations in the stored water volume during deficit periods, the cumulative distribution function (CDF) $F_{W_{tank}}\ (x) = F_{\Delta_+}\ (x)$ of the random variable Δ_+ directly corresponds to the CDF of the active storage volume. This function captures the probability that the cumulative deficit does not exceed a particular value x, effectively representing the probability that the required storage volume will not be exceeded. By utilizing this approach, the designer has the flexibility to choose the active storage volume based on the desired level of risk tolerance.

3 Case Study

The proposed models were tested and compared by the application to a case study in Milan, Italy. An hypothetical rainwater harvesting system was considered, assuming a catchment area, that is a rooftop, with a surface of 250 m², and a runoff coefficient (ϕ) equal to 1. Historical rainfall data were obtained from the Monviso rain gauge station, which provided a 47-year time series (1971–2017) with a depth resolution of 0.2 mm and a time resolution of 1 min. The time series underwent preprocessing to correct for missing data from 2005 to 2017, using the Inverse Distance Weighted Method (IDWM) with data from neighboring gauging stations, including Gattamelata, Sacco, Marino, and Sondrio. Additionally, small rainfall events producing less than 0.2 mm of rainfall were censored from the analysis, as they were considered negligible in terms of water collection. This historical dataset of rainfall records was used to identify independent rainfall events and their basic characteristics, namely rainfall event depth, rainfall event duration, and inter-event time, a crucial step for applying the probabilistic methods proposed in this study, as will be discussed in the following sections. The study simulates non-potable water use in residential settings, specifically focusing on toilet flushing demands. Demand scenarios were defined with daily consumption rates (D) ranging from 2.0 to 4.0 mm/day, reflecting various levels of water use. This corresponds to an average consumption of 40 L per person per day, leading to scenarios with 13, 16, 19, 22, and 25 users, respectively.

4 Results

4.1 Statistical Analysis and Identification of Independent Rainfall Events

In order to implement the proposed probabilistic models, a comprehensive statistical analysis was required to accurately determine the parameters essential for the models' application. A crucial step for the probabilistic applications, as previously specified, is the identification of independent rainfall events from the continuous and discrete series of rainfall data. In this study, a 4-hour minimum inter-event time (IETD) was determined based on autocorrelation analysis, ensuring statistical independence between rainfall events. The rainfall events, identified using this IETD, provided the dataset used for modeling both the empirical and theoretical distributions of inter-event times, and for further probabilistic analysis of cumulative deficit volumes.

The statistical properties of the rainfall events identified using the IETD are summarized Table 1.

Table 1 shows that while the inter-event time (d) is largely independent of both rainfall height and duration, there is a moderate correlation between height and duration ($\rho_{h,\theta} = 0.706$). This correlation likely arises from the natural tendency of longer rainfall events to accumulate more precipitation, a common feature of meteorological patterns. Despite this moderate dependence, the model was applied to evaluate its performance, accepting this limitation in order to assess its robustness under real-world conditions where such dependencies may be present.

4.2 Probabilistic Demand-Side Approach: Inter-Event Time Distribution Analysis

The series of identified rainfall events was used to compute both the empirical and theoretical cumulative distribution functions (CDFs) of inter-event times. Several theoretical distributions were tested, including the one- and two-parameter exponential, gamma, lognormal, and Weibull distributions. Among these, the Weibull distribution provided the best fit, with an NRMSE of 0.052, and is presented in Fig. 1. Additionally, quantiles of inter-event times were calculated for return periods of 5, 10, 20, and 50 years, as shown in the figure. By selecting the dry period quantile corresponding to the desired probability level and multiplying it by the water demand, the required storage volumes can be estimated accordingly.

Table 1 Average values (μ), standard deviations (σ) and Spearman's rank correlation coefficients (ρ) per event, of rainfall variables (h: rainfall depth, θ: duration, d: interevent time) in Milan

Milan (1971–2017)								
Main statistics of rainfall variables						Spearman's rank correlation coefficient		
μ_h [mm]	μ_θ [days]	μ_d [days]	σ_h [mm]	σ_θ [days]	σ_d [days]	$\rho_{h,\theta}$ [-]	$\rho_{\theta,d}$ [-]	$\rho_{h,d}$ [-]
11.63.	0.31	3.59	17.42	0.41	4.99	0.706	−0.046	0.002

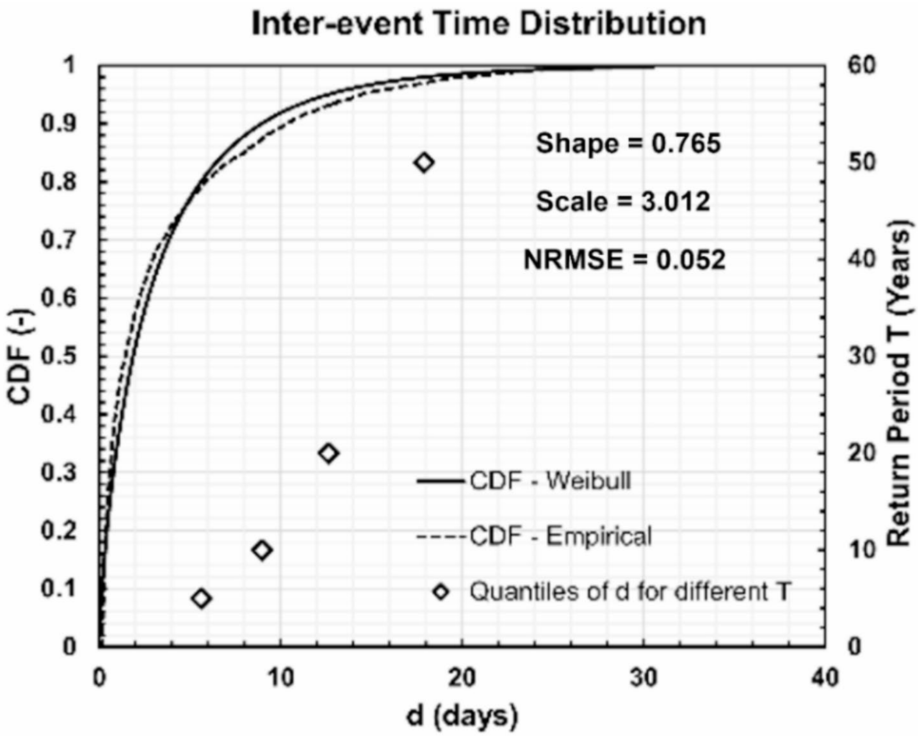

Fig. 1 Cumulative Distribution Function (CDF) of inter-event times (d) based on rainfall data from Milan, Italy, comparing the empirical CDF with a Weibull distribution fit and dry period quantiles for various return periods (T)

4.3 Comparison of Empirical and Probabilistic Cumulative Deficit Distributions for Different Demand Scenarios

After identifying the independent rainfall events, the rainfall event series was further analyzed to identify cumulative deficit sub-periods, which were defined as periods comprising events where the rainfall depth was insufficient to meet the water demand over the event duration, including the dry period that separates one event from the next. For each cumulative deficit sub-period, both the cumulative deficit volume and the number of events within the sub-period were calculated. These empirical cumulative deficit volumes were then used to construct the empirical cumulative distribution function for comparison with the theoretical curves produced by the probabilistic cumulative deficit model. The parameter n in the probabilistic model was estimated as the average number of events within each cumulative deficit sub-period, ensuring the model reflects the actual frequency and scale of deficit events.

Figure 2 shows the comparison between the empirical cumulative deficit distributions and the theoretical cumulative deficit distributions derived from the proposed probabilistic model for various water demand levels.

Figure 3 illustrates the relationship between the average number of consecutive deficit events and the mean value of the deficit variable, $E[\Delta]$.

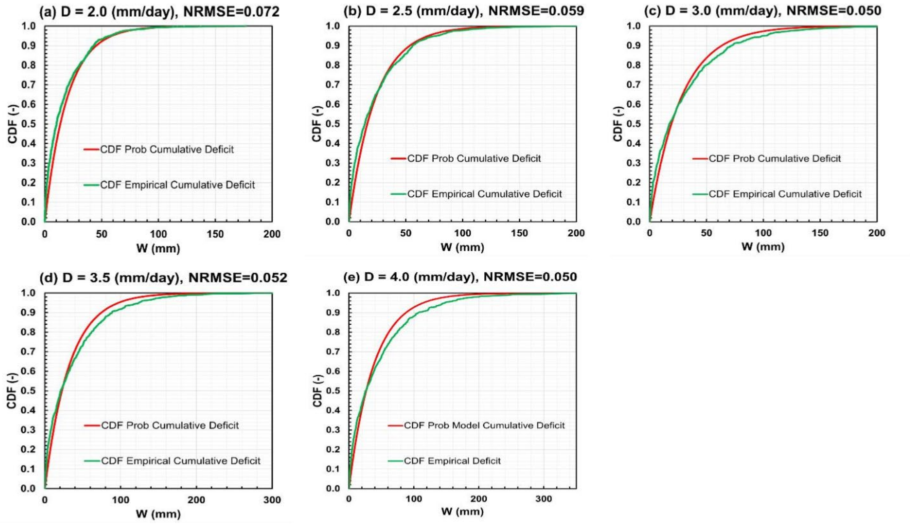

Fig. 2 Comparison of the empirical cumulative deficit curves and the theoretical cumulative distributions derived from the probabilistic model based on the cumulative deficits for different daily water demand levels.

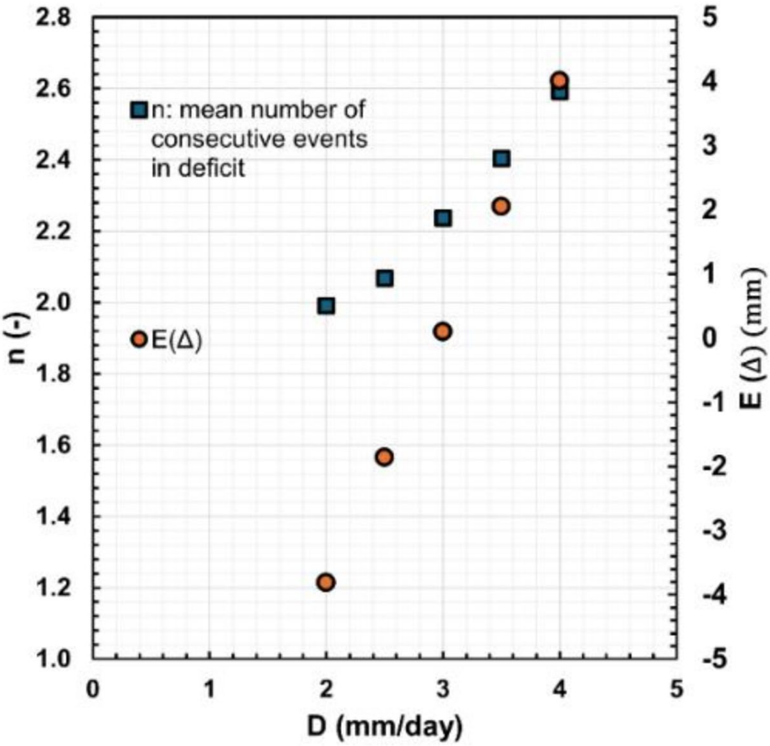

Fig. 3 Relationship between daily water demand (D), the mean number of consecutive events in deficit (n), and the mean value of the deficit variable $E[\Delta]$

4.4 Comparison of Tank Sizing Volumes Using Deterministic and Probabilistic Approaches

Figure 4 presents a comparison of tank storage volumes (W) calculated using both deterministic and probabilistic methods for varying daily water demand levels (D) and different return periods (5, 10, and 50 years).

5 Discussion

Analyzing the comparison between the empirical distribution of dry spell durations and the theoretical Weibull-modeled distribution (Fig. 1), it is observed that below a probability level of approximately 75%, the empirical distribution tends to overestimate the Weibull curve at the same quantile. This suggests that for dry periods with lower probabilities, meaning shorter dry spells, the empirical distribution indicates more frequent dry intervals than the theoretical Weibull model. Conversely, beyond the 75% probability threshold, the Weibull distribution overestimates the empirical one, indicating that for extreme events, the Weibull model predicts longer dry periods compared to empirical observations. This discrepancy can be attributed to the more regular nature of the Weibull model, which smooths out the inherent variability present in historical data, capturing a narrower range

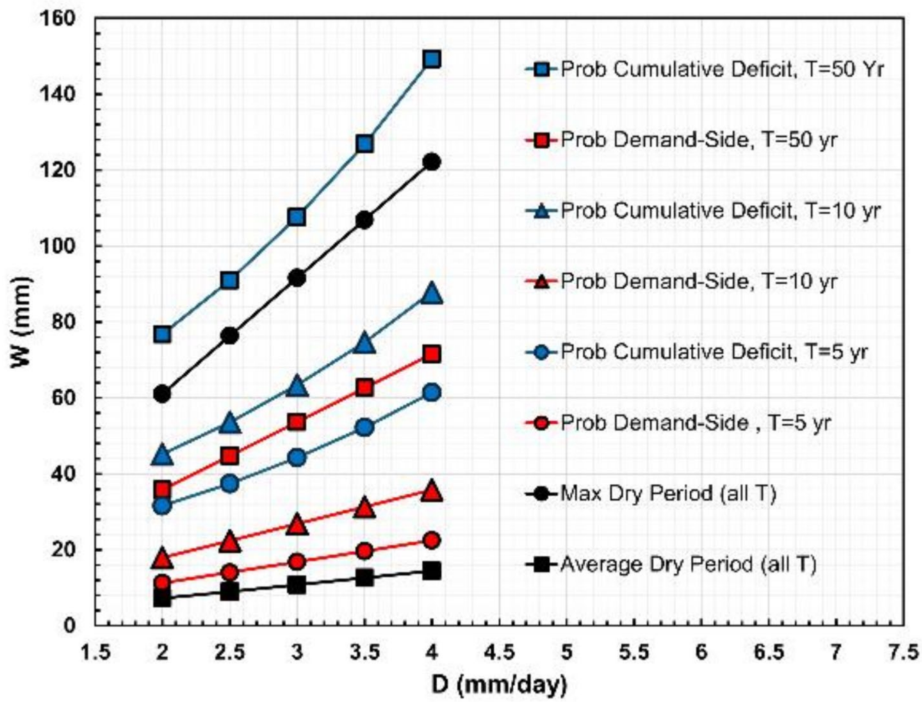

Fig. 4 Comparison of tank storage volumes (W) calculated using both deterministic and probabilistic methods for varying daily water demand levels (D) and different return periods

of fluctuations. Despite this difference, the overall fitting is considered acceptable due to the low NRMSE value obtained, confirming the suitability of the Weibull distribution for modeling inter-event times in the application of the proposed probabilistic demand-side approach. Regarding the comparison of the empirical and the probabilistic cumulative deficit distributions for different demand scenarios (Fig. 2), in all the considered scenarios, the normalized root mean square error (NRMSE) remains below 0.10, demonstrating that the probabilistic model provides a reliable approximation of cumulative deficit volumes when compared with empirical data. For lower demand levels, such as 2.0 mm/day and 2.5 mm/day, the NRMSE values are slightly higher than those observed in other cases, yet the alignment between the empirical and probabilistic curves is closer. As demand increases, a trend emerges where the probabilistic model tends to overestimate the cumulative deficit, particularly at higher probability levels corresponding to larger quantiles. This overestimation becomes more pronounced with increasing demand, indicating that while the probabilistic model performs well for moderate deficits, it may slightly over-predict more extreme deficit scenarios. The observed behavior is explained by the fact that the probabilistic cumulative deficit model is developed under the assumption of global surplus conditions, where cumulative water supply, on average, exceeds demand. This is reflected in the behavior of the parameter Δ, where a negative mean value of Δ (E[Δ]) indicates average surplus conditions, while a positive E[Δ] signals average deficit conditions. As it can be seen in Fig. 3, for lower demand levels, such as 2.0 mm/day and 2.5 mm/day, E[Δ] remains negative, indicating an average state of water surplus in which the tank is replenished more frequently, leading to smaller cumulative deficits that the model captures with greater precision. Despite slightly higher NRMSE values, the model demonstrates good visual alignment between empirical and probabilistic curves, effectively handling the smaller and more consistent variations in cumulative deficits. However, as demand exceeds 2.5 mm/day, E[Δ] becomes positive, marking a shift from surplus to deficit conditions. At these higher demand levels, the system experiences more substantial cumulative deficits, which the probabilistic model tends to overestimate, particularly at higher quantiles corresponding to larger deficit events. This behavior highlights the model's sensitivity to changing supply-demand dynamics, transitioning from surplus to deficit conditions as demand increases. While the model is expected to perform better under low-demand conditions due to its foundation on the assumption of global surplus, it also requires at least two consecutive deficit events to function accurately. Consequently, a minimum demand threshold of 2.0 mm/day was set in the study, ensuring an average of at least two consecutive deficit events ($n \approx 2$). Falling below this threshold violates a fundamental assumption of the model, which is designed to account for consecutive deficit events and thus requires an average number of events of at least two. Therefore, while the model is based on global surplus conditions, it simultaneously necessitates a series of at least two consecutive deficit events to remain applicable. This apparent contradiction actually reflects the need to balance system behavior under varying operating conditions, as even in global surplus conditions, short-term deficits can still occur, and the model aims to capture the distribution of these consecutive deficit events. If demand is too low, the system enters surplus too frequently, drastically reducing the likelihood of experiencing two or more consecutive deficit events.

Regarding the comparison of tank volumes obtained with probabilistic and deterministic approaches (Fig. 4), it emerges that increasing the return period results in a significant rise in required storage volume, particularly for higher demand levels, highlighting the inherent trade-off between ensuring higher reliability and the costs associated with constructing larger tanks. The probabilistic approach based on cumulative deficit consistently results in higher storage volumes compared to the probabilistic demand-side approach, especially at higher demand levels. This discrepancy arises because the cumulative deficit model accounts not only for dry periods with no rainfall but also for instances where rainfall occurs yet remains insufficient to meet demand. In other words, the cumulative deficit model considers both rainfall intensity variability and the potential mismatch between rainfall and demand during non-dry periods. This is particularly critical during water scarcity periods when even minor rainfall events may not fully replenish storage due to high demand. As a result, the cumulative deficit model captures the cumulative effect of these deficits over consecutive events, leading to higher storage volume requirements and providing a more comprehensive assessment of the system's water balance. This contrasts with models based solely on inter-event times, which may overlook deficits occurring within rainy periods, ultimately underestimating the required storage volume. Deterministic approaches show distinct behavior, with the method based on the average dry period resulting in smaller tank sizes than the one based on the maximum dry period, reflecting its less conservative nature by balancing typical rainfall conditions rather than focusing on worst-case scenarios. The maximum dry period approach, in contrast, leads to storage volumes surprisingly close to those obtained from the cumulative deficit model with a 50-year return period, highlighting its extreme conservatism, particularly given that such a long return period is uncommon for typical rainwater harvesting infrastructure. Consequently, this method may significantly overestimate required storage volumes compared to more refined probabilistic models, which offer a better balance between efficiency and risk management. Probabilistic methods provide intermediate estimates that are more balanced than either deterministic method, as they consider both dry periods and the occurrence of rainfall events, leading to a more accurate estimation of tank size. Unlike deterministic approaches that focus solely on historical drought extremes or averages, probabilistic methods capture the stochastic nature of rainfall events, offering greater flexibility in accounting for variability in water supply.

6 Conclusions

This study introduces two advanced probabilistic methodologies for estimating storage tank volumes in rainwater harvesting systems, offering a substantial improvement over traditional deterministic demand-side approaches. The first methodology refines the classical demand-side model by incorporating the frequency distribution of inter-event times, providing a more flexible framework for calculating storage volumes based on desired reliability levels, such as return periods or failure probabilities. This allows for a clearer understanding of how temporal variability in rainfall events influences the storage requirements. The second methodology is based on cumulative deficits over multiple independent consecutive rainfall events, considering not only dry periods but the entire stochastic rainfall process, including rainy intervals. By evaluating the cumulative distribution function of the storage volume, this approach accounts for both the frequency and intensity of rainfall events specific to the local climate, resulting in a more tailored design that reflects the region's precipitation characteristics. A critical theoretical analysis has been conducted to highlight the strengths and limitations of each methodology, further supported by a comparison with traditional demand-side sizing approaches. Based on the analyses conducted in this study, it can be concluded that while deterministic methods tend to overestimate storage volumes when relying on extreme dry periods or underestimate them when based solely on average conditions, probabilistic approaches provide a more accurate representation by capturing the full range of variability in rainfall patterns. Additionally, deterministic methods lack the flexibility to incorporate risk management, making them less adaptable to varying reliability requirements in system design. By establishing a clearer connection between storage volume and the desired reliability level, probabilistic methodologies enable more efficient use of resources, ensuring that rainwater harvesting systems are both cost-effective and resilient to periods of water scarcity.

Author Contributions Conceptualization, G.B. and M.G.D.; methodology, G.B. and M.G.D.; validation, G.B., M.M. and M.G.D.; formal analysis, G.B., M.M. and M.G.D.; investigation, G.B. and M.G.D.; data curation, M.G.D.; writing—original draft preparation, M.G.D.; writing—review and editing, G.B., M.M. and M.G.D.; visualization, G.B.; supervision, G.B.; project administration, G.B.; funding acquisition, G.B. All authors have read and agreed to the published version of the manuscript.

Funding Open access funding provided by Politecnico di Milano within the CRUI-CARE Agreement.
The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data Availability The authors would like to express their gratitude to MM S.p.A. for providing the data used in this study.

Declarations

Ethical Approval Not applicable.

Consent to Participate Not applicable.

Consent to Publish The authors confirm that they have obtained the necessary consent for the publication of this work. All data and materials comply with ethical guidelines and privacy regulations.

Competing Interests The authors have no relevant financial or non-financial interests to disclose.

References

Adler I, Hudson-Edwards KA, Campos LC (2011) Converting rain into drinking water: quality issues and technological advances. Water Supply 11(6):659–667. https://doi.org/10.2166/ws.2011.117

Allen J, Haarhoff J (2015) A proposal for the probabilistic sizing of rainwater tanks for constant demand. J South Afr Institution Civil Eng 57(2):22–27. https://doi.org/10.17159/2309-8775/2015/v57n2a3

Becciu G, Raimondi A, Dresti C (2018) Semi-probabilistic design of rainwater tanks: A case study in Northern Italy. Urban Water J 15(3):192–199. https://doi.org/10.1080/1573062X.2016.1148177

De Sá Silva ACR, Bimbato AM, Balestieri JAP, Vilanova MRN (2022) Exploring environmental, economic and social aspects of rainwater harvesting systems: A review. Sustainable Cities Soc 76:103475. https://doi.org/10.1016/j.scs.2021.103475

Di Chiano MG, Marchioni M, Raimondi A, Sanfilippo U, Becciu G (2023) Probabilistic approach to tank design in rainwater harvesting systems. Hydrology 10(3):59. https://doi.org/10.3390/hydrology10030059

Essien AE, Guo Y, Khafagy M, Dickson-Anderson SE (2024) Design and hydrologic performance Estimation of highway filter drains using a novel analytical probabilistic model. Sci Rep 14(1):2350. https://doi.org/10.1038/s41598-024-52760-7

Hdeib R, Aouad M (2023) Rainwater harvesting systems: an urban flood risk mitigation measure in arid areas. Water Sci Eng 16(3):219–225. https://doi.org/10.1016/j.wse.2023.04.004

Hviid CA, Zukowska-Tejsen D, Nielsen V (2020) Cooling of schools – results from a demonstration project using adiabatic evaporative cooling with harvested rainwater. E3S Web Conferences 172:02003. https://doi.org/10.1051/e3sconf/202017202003

Jamali B, Bach PM, Deletic A (2020) Rainwater harvesting for urban flood management – An integrated modelling framework. Water Res 171:115372. https://doi.org/10.1016/j.watres.2019.115372

Joo J, Lee J, Kim J, Jun H, Jo D (2013) Inter-Event time definition setting procedure for urban drainage systems. Water 6(1):45–58. https://doi.org/10.3390/w6010045

Khan AS (2023) A comparative analysis of rainwater harvesting system and conventional sources of water. Water Resour Manage 37(5):2083–2106. https://doi.org/10.1007/s11269-023-03479-z

Kim H, Han M, Lee JY (2012) The application of an analytical probabilistic model for estimating the rainfall–runoff reductions achieved using a rainwater harvesting system. Sci Total Environ 424:213–218. https://doi.org/10.1016/j.scitotenv.2012.02.021

Kottegoda NT (2008) Applied statistics for civil and environmental engineers, 2nd edn. Blackwell Pub

Lee EH, Kim JH (2018) Development of new Inter-Event time definition technique in urban areas. KSCE J Civ Eng 22(10):3764–3771. https://doi.org/10.1007/s12205-018-1120-5

Marchioni M, Raimondi A, Di Chiano MG, Becciu G (2023) Permeable pavement hydraulic optimization by using an analytical-probabilistic model. Blue-Green Syst 5(2):252–274. https://doi.org/10.2166/bgs.2023.016

Mishra B, Kumar P, Saraswat C, Chakraborty S, Gautam A (2021) Water security in a changing environment: concept, challenges and solutions. Water 13(4):490. https://doi.org/10.3390/w13040490

Nandi S, Gonela V (2022) Rainwater harvesting for domestic use: A systematic review and outlook from the utility policy and management perspectives. Utilities Policy 77:101383. https://doi.org/10.1016/j.jup.2022.101383

Raimondi A, Becciu G (2014) Probabilistic modeling of rainwater tanks. Procedia Eng 89:1493–1499

Raimondi A, Marchioni M, Sanfilippo U, Stroppa FF, Becciu G (2022) Probabilistic Estimation of runoff from green roofs. Int J Comput Methods Experimental Measurements 10(1):13–25. https://doi.org/10.2495/CMEM-V10-N1-13-25

Raimondi A, Di Chiano MG, Marchioni M, Sanfilippo U, Becciu G (2023) Probabilistic modeling of sustainable urban drainage systems. Urban Ecosyst 26(2):493–502. https://doi.org/10.1007/s11252-022-01299-4

Teixeira CA, Ghisi E (2019) Comparative analysis of granular and membrane filters for rainwater treatment. Water 11(5):1004. https://doi.org/10.3390/w11051004

Tu A, Zeng J, Liu Z, Zheng H, Xie S (2023) Effect of minimum inter-event time for rainfall event separation on rainfall properties and rainfall erosivity in a humid area of Southern China. Geoderma 431:116332. https://doi.org/10.1016/j.geoderma.2023.116332

Uddin Sikder MK (2020) Ultraviolet disinfection of potable water in rainwater harvesting systems. Int J Eng Appl Sci Technol 04(11):557–563. https://doi.org/10.33564/IJEAST.2020.v04i11.098

United Nations (2024) The sustainable development goals report 2024. https://unstats.un.org/sdgs/report/2024/The-Sustainable-Development-Goals-Report-2024.pdf

Wartalska K, Grzegorzek M, Bełcik M, Wdowikowski M, Kolanek A, Niemierka E, Jadwiszczak P, Kaźmierczak B (2024) The potential of rainwater harvesting systems in Europe – Current state of Art and future perspectives. Water Resour Manage 38(12):4657–4683. https://doi.org/10.1007/s11269-024-03882-0