Journal of Science and Technology in Civil Engineering, HUCE, 2025, 19 (1): 47–58

DEVELOPING A DESIGN RAINFALL INTENSITY EQUATION USING MEASUREMENT DATA FROM THE HA DONG METEOROLOGICAL STATION IN HANOI, VIETNAM

Ha Xuan Anh[a,*], Tran Thi Viet Nga[b], Nguyen Van Nam[a]

[a] *Faculty of Urban Environment and Infrastructure Engineering, Hanoi Architectural University, Km 10, Nguyen Trai street, Thanh Xuan district, Hanoi, Vietnam*
[b] *Faculty of Environmental Engineering, Hanoi University of Civil Engineering, 55 Giai Phong road, Hai Ba Trung district, Hanoi, Vietnam*

Article history:
Received 10/12/2024, Revised 03/01/2025, Accepted 19/3/2025

Abstract

In the design and calculation of stormwater drainage systems, the design rainfall intensity plays a crucial role and is widely used in runoff calculations. Due to the influence of local climatic factors, the selection of probability distributions for rainfall frequency analysis and the methods used to determine parameters for empirical equations result in a variety of equations compared to traditional Rainfall Intensity-Duration-Frequency curves. This paper focuses on developing a design rainfall intensity equation based on observed rainfall data from the Ha Dong meteorological station in Hanoi. The Generalized reduced gradient nonlinear method was applied to derive the equation. The results indicate that the proposed equation aligns more closely with Rainfall Intensity-Duration-Frequency curves. The methods introduced in this paper can be effectively applied to develop design rainfall intensity equations for other urban areas, providing valuable support for the design of urban stormwater drainage systems.

Keywords: design rainfall intensity; intensity-duration-frequency curves; probability distribution functions; rainwater drainage systems.

https://doi.org/10.31814/stce.huce2025-19(1)-05

1. Introduction

In the design of rainwater drainage systems, design rainfall intensity plays a crucial role as the foundation for calculating the required drainage flow. In recent years, under the influence of climate change, rainfall patterns in Hanoi have shown increasing variability in intensity, frequency, and duration. Rainfall in Hanoi is primarily influenced by atmospheric circulation systems. The main causes of rainfall include storms, tropical depressions, the activity of the Southwest or Southeast monsoons, storms combined with cold air, and upper-level cyclones. The rainy season in Hanoi typically begins in May and ends in October. During this period, total rainfall can account for 80–85% of the annual total, even though the number of rainy days only makes up 50–55% of the yearly count. Monthly rainfall can reach up to 700 mm, and in some locations, such as Ha Dong in 2008, it has exceeded 800 mm. Such high rainfall events often occur in the later months of the rainy season, contributing approximately 40% of the annual total. The changes in extreme rainfall in Hanoi [1] were evaluated using two indicators: standard deviation and rainfall variability, both of which indicate an increasing trend in extreme rainfall. This trend is evident from the series of consecutive historical rainfall events recorded in recent years in Hanoi. Over the past 50 years, the largest daily rainfall has shown a rising tendency, increasing at a rate of approximately 0.6% to 0.9% per decade.

*Corresponding author. *E-mail address:* haxuananh.hau@gmail.com (Anh, H. X.)

Typically, the design of heavy rainfall events considers the relationship between rainfall intensity, duration, and frequency, commonly referred to as the IDF relationship. The concept of developing an IDF relationship was first proposed by Bernard (1932) [2] and then has since been widely adopted and refined globally. The IDF relationship is generally represented in one of three forms: (1) rainfall intensity distribution maps, (2) IDF curve relationship charts, or (3) empirical equations used to calculate design rainfall intensity. In general, when rainfall data are available, IDF curves can be developed using statistical methods and frequency analysis. In Vietnam, design rainfall intensity equations are widely used due to limitations in the rain gauge network, including insufficient coverage and lack of short-duration measurement data records [3]. These constraints make it challenging to apply other calculation methods effectively. As a result, empirical equations remain the most practical and suitable selection for addressing design challenges in rainwater drainage systems. Additionally, many studies focus on maximizing the utilization of existing regional rainfall data to develop calculation equations, enhancing accuracy by tailoring them to the specific characteristics of local rainfall and climate. Empirical equations representing IDF relationships for areas without observation data are often derived from IDF curves established at stations or locations with available rainfall measurements. The reliability of these newly proposed equations is validated by comparing their results with those of existing equations or other traditional calculation methods.

Over the past two decades, several researchers have studied the use of empirical equations to calculate rainfall intensity for designing urban stormwater drainage systems. In 1996, Dung [4] presented his doctoral thesis on refining methods for determining design rainfall runoff for urban drainage systems in Vietnam. In 2018, Hong [5] focused her doctoral research on improving methods for determining rainfall patterns and design drainage flows for systems in the Northern Delta region, Vietnam. In 2023, Giang et al. [6] conducted a study on developing a combined rainfall-water level curve for surface drainage system design in Ho Chi Minh City. Similarly, in 2018, Trang et al. [7] analyzed rainfall patterns to enhance the design of urban stormwater drainage systems in monsoon-affected areas of Vietnam. An analysis of the commonly used rainfall intensity equations in Vietnam reveals that these equations include parameters tailored to the unique characteristics of each locality, such as climate, geographical location, and rainfall patterns. Most existing equations are accompanied by appendices containing lookup tables for parameter values specific to each region. However, these parameters require periodic updates to account for changes in the influencing factors, as discussed earlier. To address these limitations, a key objective of this paper is to develop a design rainfall intensity equation for Ha Dong area using the updated rainfall data observed at the Ha Dong meteorological station and its rainfall Intensity – Duration – Frequency curves.

In this paper, the IDF relationship curves constructed for a meteorological station with long rainfall measurements will be used as an important basis for developing a new and suitable equation for design rainfall intensity in accordance with the current conditions of Hanoi city.

2. Methodology

Some basic knowledge about statistics and probability applied in meteorology is referred to in [8–10]. The flowchart below represents the approach and methods employed in this study.

Frequency is the number of times a certain value appears in the total number of tests or observations. Frequency is the ratio of the number of times a certain value appears to the total number of tests. In hydrometeorology, frequency is the ratio of the number of times a certain value appears compared to the total number of observations. In hydrometeorology, when studying the value x, people often observe how many times the value of the studied quantity is greater than or equal to x. Therefore,

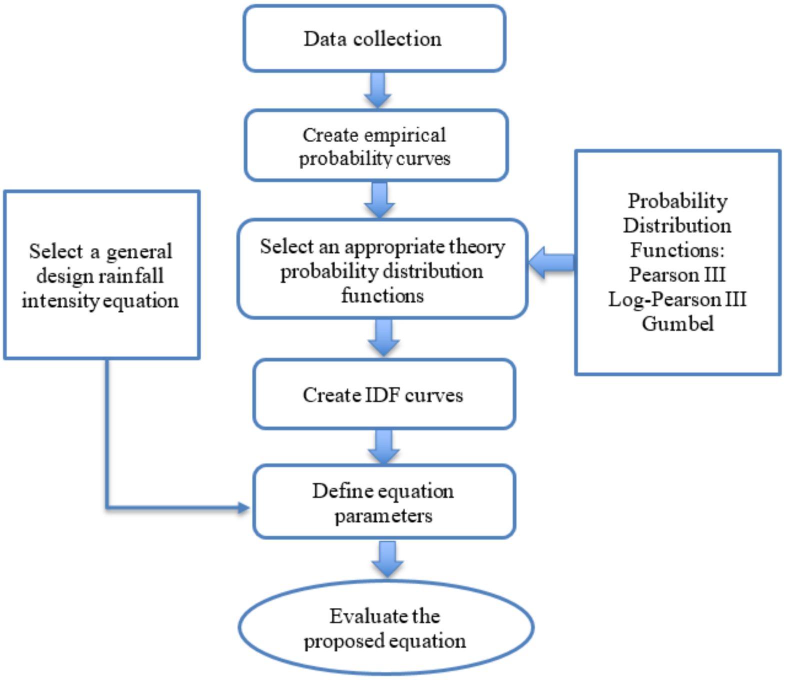

Figure 1. Flowchart of the study

people are often interested in the cumulative frequency $\frac{m}{n}$, meaning that in n years of observation, there are m times a value appears greater than or equal to a value that we are considering.

The mean $\overline{x}$ is the arithmetic mean of the data series. The mean is one of the most basic and important characteristics of a data series. If the data series consists of elements $x_1, x_2, \ldots, x_n$, the mean is calculated according to the formula

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} \tag{1}$$

The sample variance is the average of the squared deviations of the values around the mean:

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n} \tag{2}$$

Sample deviation is a descriptive statistic that measures the dispersion of a set of data that has been tabulated into a frequency table. Sample deviation is the square root of the average of the squares of the deviations of the values around the mean.

Distributions of data may not be symmetrical concerning its arithmetic mean. The coefficient of skewness C_s is used to measure the asymmetry of the data:

$$C_s = \frac{\frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^3}{s^3} \tag{3}$$

However, Eqs. (2) and (3) mentioned above are only suitable for samples with a rather large sample size. For hydrological phenomena, because there are often not long data series, we often use

$$s^2 = \frac{\sum\limits_{i=1}^{n} (x_i - \overline{x})^2}{n-1}, \quad C_s = \frac{\frac{1}{n-2}\sum\limits_{i=1}^{n} (x_i - \overline{x})^3}{s^3} \tag{4}$$

The exceedance probability, denoted by P, of a value in a data series is given by the formula

$$P = \frac{k}{n+1} \tag{5}$$

where k is the data sequence number (arranged in descending order) and n is the number of elements in the data series.

The concept of the return period of any hydrologic event plays a key role in risk and uncertainty analysis in hydroclimatic studies. The return period can be defined as the average length of time for an event of a given magnitude to be equaled or exceeded in a statistical sense. In hydrometeorology, the repetition interval T is usually calculated based on the formula

$$T = \frac{1}{P} \tag{6}$$

Frequency analysis usually refers to stationary frequency analysis which assumes that the data are stationary. Most frequency distribution functions in hydroclimatic studies can be expressed in the following equation, known as the generalized frequency analysis equation, given by

$$x_T = \overline{x} + K_T \cdot s \tag{7}$$

where x_T is the value of the observed quantity corresponding to a return period of T years; K_T is the frequency coefficient, which depends on the return period T.

The methods of data analysis and surveys based on common statistical characteristics allow us to indicate the properties of meteorological and climatic factors based on specific data sets obtained from actual observations. However, due to the limitation of sample size in meteorological research, which is often not very large, the results obtained may not accurately reflect the nature of the process under consideration in many cases. To overcome this situation, in addition to studying samples, scientists use theoretical distributions and approximate experimental data with appropriate theoretical distributions. Using theoretical distributions to approximate experimental data essentially idealizes the experimental data set, treating the experimental results as though they are derived from some mathematical formulas. Although this representation is very accurate in many instances, it is fundamentally just an approximate representation of the experimental data. However, approximating experimental data with theoretical distributions has many advantages:

+ In many cases, researchers must repeatedly calculate the statistical characteristics of samples for a certain location or space. The calculation process can be very cumbersome, complicated, and prone to unusual errors. If a theoretical distribution fits the data, we only need a few parameters of this distribution rather than conducting a complete survey.

+ Theoretical distributions allow for the interpolation of missing data (or the absence of data), thereby filling in data gaps.

+ Due to the limitation of sample size, experimental data only reflect the variation of factor characteristics within the range of variation of the sample set. Utilizing theoretical distributions allows for the estimation of the probability of events outside the sample set range, particularly for extreme situations.

Normally, after constructing the empirical distribution function, we need to study, evaluate, consider, and select the theoretical distribution that best fits the empirical distribution. The analysis of rainfall frequency distribution is based on three distributions: Pearson Type III, Log-Pearson Type III, and Gumbel, which are commonly used in Vietnam for this type of analysis [11].

A random variable X has a Pearson III distribution if its probability density function is

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} (x - x_0)^{\alpha - 1} e^{\beta(x - x_0)} \tag{8}$$

where $\Gamma(\alpha)$ is the Gamma function

$$\Gamma(\alpha) = \int_{0}^{+\infty} t^{\alpha-1} \cdot e^{t-1} dt. \tag{9}$$

A random variable X follows the Log-Pearson III distribution if the random variable $Y = \log(X)$ follows the Person III distribution. A random variable X is said to have a Gumbel distribution if its probability density function has the form

$$f(x) = \frac{\exp\left(-\dfrac{x-\beta}{\alpha} - \exp\left(-\dfrac{x-\beta}{\alpha}\right)\right)}{\alpha}. \tag{10}$$

3. Deriving Intensity-Duration-Frequency (IDF) curves using observed rainfall data from the Ha Dong rain gauge

3.1. Study area and data collection

Ha Dong District, one of the main districts of Hanoi, [12], covers a total drainage basin area of 4,995 hectares and comprises four main sub-basins, including the sub-basin north of National Highway 6, the sub-basin south of National Highway 6, the sub-basin east of the Nhue River, and the sub-basin within the Day River dike area. The primary drainage direction of Ha Dong District is toward the Nhue River, with some portions in the south draining to the Day River. The rainwater drainage system in the district is integrated with the irrigation drainage system. During periods of high river water levels, pumping stations are required to discharge water into the Nhue River.

Similar to other urbanized areas, the current rainwater drainage system is a combined system of both rainwater and wastewater. Much of this infrastructure was constructed during the French colonial period, and has been recently upgraded. Under normal conditions with no heavy rainfall, the drainage system functions effectively for the district and surrounding areas. However, in recent years, rising water levels in the Nhue River, frequent heavy rainfall, and rapid urbanization have led to flooding and inundation in this area.

A significant challenge in developing an equation for design rainfall intensity is the limited availability of long-term, high-resolution rainfall data with adequate spatial and temporal detail for the Hanoi area. To accurately reflect rainfall variation, these datasets need to span at least 20–30 years. By collecting and analyzing rainfall data from meteorological and rain gauge stations in Hanoi, this study utilized the rainfall data series from the Ha Dong station. As a key meteorological station located in the inner city, Ha Dong station provides a detailed, long-term, and synchronous dataset that is particularly well-suited for research on calculating and determining design rainfall intensity equations. The collected dataset in this study covers 50 years, from 1973 to 2023.

3.2. Calculation for Ha Dong station with 5-minute period according to three distributions

Based on rainfall data from Vietnam Institute of Meteorology, Hydrology and Climate change (IMHEN) at Ha Dong station for the period 1973-2023, annual maximum value series for different rainfall periods, including 5 min, 10 min, 20 min, 30 min, 45 min, 1 h, 2 h, . . . , 24 h, were obtained. Frequency analysis technique is used to develop the relationship between rainfall intensity and return period from rainfall data. Table 1 is the maximum 5-minute rainfall data at Ha Dong station during the period 1973-2023.

Table 1. Rainfall data from Ha Dong station

Year	Value	Date	Year	Value	Date	Year	Value	Date
1973	7.2	17-Sep	1990	13.5	19-Feb	2007	18.6	21-Aug
1974	14.5	19-May	1991	15.3	21-Jun	2008	22.5	31-Oct
1975	14.6	7-Oct	1992	19.6	10-May	2009	14.5	2-Jun
1976	15.0	22-May	1993	22.6	23-Jun	2010	17.9	15-Aug
1977	10.6	8-Apr	1994	15.4	30-Aug	2011	8.6	11-May
1978	17.5	22-Sep	1995	13.2	22-Aug	2012	12.4	22-May
1979	11.5	7-Jul	1996	14.2	22-Jul	2013	7.4	9-Aug
1980	16.5	23-Jun	1997	19.5	16-Aug	2014	19.8	21-Sep
1981	13.2	3-Aug	1998	20.9	16-May	2015	20.0	26-Aug
1982	8.5	15-Nov	1999	13.5	15-Jul	2016	14.9	25-May
1983	12.5	29-Apr	2000	13.1	25-Jun	2017	11.0	13-Jul
1984	19.5	18-May	2001	12.9	4-Jun	2018	19.8	12-May
1985	16.5	8-Sep	2002	17.5	30-Jun	2019	12.8	19-Sep
1986	11.5	4-Sep	2003	26.5	3-May	2020	15.5	3-Mar
1987	14.5	21-Nov	2004	11.4	27- Jun	2021	10.0	17-Apr
1988	10.5	2-Aug	2005	12.8	21-Jul	2022	19.8	5-Jul
1989	11.5	4-Jun	2006	12.1	9-Sep	2023	19.8	31-Jul

Based on the rainfall data in Table 1, the sample mean $\overline{x} = 15.0$, sample deviation $s = 4.201$, and skewness $C_s = 0.405$ can be calculated.

Use the frequency coefficient and skewness coefficient table in to find the K_T values for the return periods of 2 years, 5 years, 10 years, 25 years, 50 years, and 100 years. Since the skewness coefficient $C_s = 0.405$ lies between the two coefficients given in the table, a linear approximation between the two numbers can be made to obtain the appropriate K_T value. Since $0.405 = 0.95{\cdot}0.4+(1-0.95){\cdot}0.5$, then

$$K_T\left(C_s = 0.405\right) = 0.95 \cdot K_T\left(C_s = 0.4\right) + (1 - 0.95) \cdot K_T\left(C_s = 0.5\right) \tag{11}$$

From there, the rainfall intensity corresponding to the return period T can be calculated.

To use the Log-Pearson III distribution, the first step is to calculate the log of the 5-minute rainfall data and calculate the characteristic numbers corresponding to this logarithmized data set as the sample $y = 1.2$, sample deviation $s = 0.125$, and skewness $C_s = -0.313$. Then, calculate the value of K_T corresponding to the repetition times and the asymmetry coefficient $C_s = -0.313$. From there, the value of $\log x$ corresponding to the repetition times T is computed. Then by exponentiating $x = 10^{\log x}$ we get the rainfall intensity.

Next, we apply the calculation formula according to the Gumbel distribution

$$x_T = \overline{x} + K_T \cdot s \tag{12}$$

where K_T is the coefficient corresponding to the return period T and the capacity of the data sample $n = 51$. From there, the rainfall intensity in the 5-minutes period can be calculated according to the Gumbel distribution corresponding to the return period T.

Table 2. Rainfall intensity converted by the Pearson III distribution

The return period (years)	K_T at $C_s = 0.5$	K_T at $C_s = 0.4$	K_T at $C_s = 0.405$	Rainfall in 5 minutes (mm)	Rainfall intensity (mm/h)
2	−0.083	−0.066	−0.068	15.875	190.5
5	0.808	0.816	0.815	20.116	241.4
10	1.323	1.317	1.318	22.530	270.4
25	1.91	1.88	1.884	25.249	303.0
50	2.311	2.261	2.268	27.091	325.1
100	2.686	2.615	2.625	28.804	345.7

Table 3. Rainfall intensity converted by the Log-Pearson III distribution

The return period (years)	K_T at $C_s = -0.3$	K_T at $C_s = -0.4$	K_T at $C_s = -0.313$	$\log x$	Rainfall in 5 minutes (mm)	Rainfall intensity (mm/h)
2	0.05	0.066	0.052	1.165	14.634	175.6
5	0.853	0.855	0.853	1.266	18.441	221.3
10	1.245	1.231	1.243	1.315	20.638	247.7
25	1.643	1.606	1.638	1.364	23.130	277.6
50	1.89	1.834	1.882	1.395	24.821	297.8
100	2.104	2.029	2.094	1.421	26.383	316.6

Table 4. Rainfall intensity converted by the Gumbel distribution

The return period (years)	K_T	Rainfall in 5 minutes (mm)	Rainfall intensity (mm/h)
2	0.160	15.670	188.0
5	0.819	18.439	221.3
10	1.464	21.149	253.8
25	2.281	24.581	295.0
50	2.885	27.119	325.4
100	3.486	29.644	355.7

From Tables 2, 3 and 4, the rainfall intensity table corresponding to the 5-minutes period is obtained according to the selected distributions including Pearson III, Log-Pearson III, Gumbel.

To test the suitability of the selected distributions to the observed data, the Chi-square distribution χ^2 is used according to the following formula:

$$\chi^2 = \sum \frac{\left(O_f - E_f\right)^2}{E_f} \tag{13}$$

where O_f is the experimental frequency (the number of observed values in the considered interval), E_f is the expected frequency (calculated as the product of the sample size and the probability that the observed value falls in the considered interval assuming the data follows the theoretical distribution).

Table 5. Rainfall intensity converted by the three distributions

The return period (years)	Exceedance probability	Rainfall intensity (mm/h)		
		Log-Pearson III	Pearson III	Gumbel
2	0.5	175.6	190.5	188.0
5	0.2	221.3	241.4	221.3
10	0.1	247.7	270.4	253.8
25	0.04	277.6	303.0	295.0
50	0.02	297.8	325.1	325.4
100	0.01	316.6	345.7	355.7

Table 6. Test results for the Gumbel distribution

Value range	Probability	E_f	O_f	χ^2
10.4	0.102	5.19	5	0.007
10.4 – 13.6	0.321	16.38	18	0.159
13.6 – 16.9	0.307	15.68	12	0.862
16.9 – 20.1	0.158	8.06	12	1.926
20.1 – 23.3	0.068	3.47	3	0.063
23.3	0.044	2.22	1	0.672
	1.000	51	51	3.688

Table 7. Test results for the Pearson III distribution

Value range	Probability	E_f	O_f	χ^2
10.4	0.129	6.58	5	0.379
10.4 – 13.6	0.256	13.04	18	1.883
13.6 – 16.9	0.295	15.05	12	0.620
16.9 – 20.1	0.186	9.49	12	0.665
20.1 – 23.3	0.081	4.12	3	0.305
23.3	0.053	2.71	1	1.083
	1.000	51	51	4.935

Table 8. Test results for the Log-Pearson III distribution

Value range	Probability	E_f	O_f	χ^2
10.4	0.126	6.45	5	0.325
10.4 – 13.6	0.251	12.78	18	2.130
13.6 – 16.9	0.289	14.75	12	0.514
16.9 – 20.1	0.182	9.30	12	0.785
20.1 – 23.3	0.108	5.49	3	1.128
23.3	0.044	2.23	1	0.680
	1.000	51	51	5.561

From Tables 6, 7, and 8, it can be seen that the Gumbel distribution test value is the smallest, proving that the Gumbel distribution is the most suitable when approximating the rainfall data.

3.3. IDF curves for Ha Dong station

Based on the method implemented in the previous subsection, the study obtained a table of rainfall intensity data according to the time periods of 5 minutes, 10 minutes, 15 minutes, 30 minutes, 45 minutes, 60 minutes, 90 minutes, 120 minutes, 180 minutes, 360 minutes, 720 minutes, 1440 minutes with a recurrence period of 2 years, 5 years, 10 years according to Gumbel distribution.

Table 9. IDF table for Ha Dong station

Duration (mins)	Rainfall intensity (mm/h)					
	2 years	5 years	10 years	25 years	50 years	100 years
5	188	221.3	253.8	295	325.4	355.7
10	144.7	170.5	195.9	227.9	251.6	275.2
15	127.6	149.5	170.9	198.1	218.1	238.1
30	97.6	114.7	131.5	152.7	168.4	184.1
45	79.3	94	108.4	126.7	140.2	153.6
60	69.2	81.9	94.3	110	121.6	133.2
90	51.1	61.3	71.2	83.8	93.1	102.4
120	42.3	51.1	59.8	70.8	78.9	87
180	32.1	40	47.6	57.4	64.6	71.7
360	20.2	25.8	31.4	38.5	43.7	48.9
720	12.5	16.6	20.5	25.5	29.1	32.8
1440	7.5	10.1	12.6	15.9	18.3	20.6

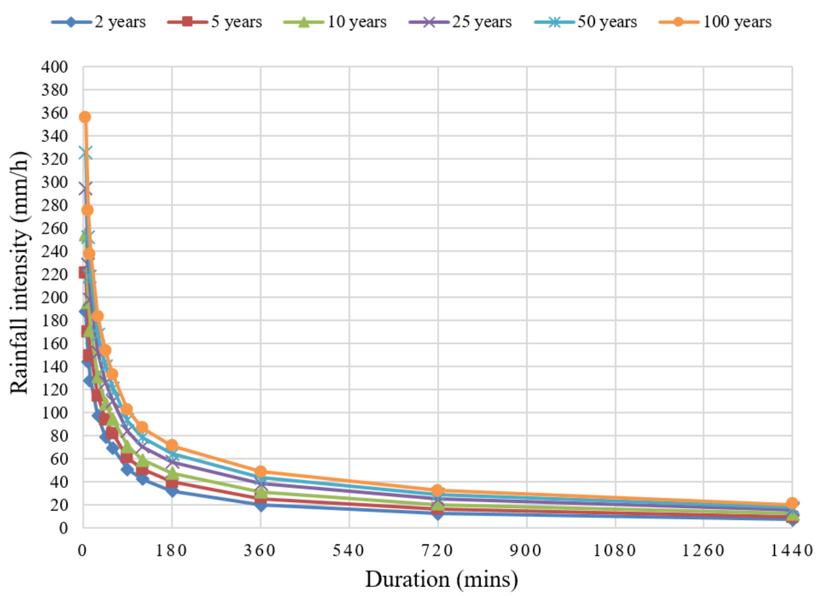

Figure 2. The new IDF curves

From there, draw IDF curves corresponding to the return periods of 2 years, 5 years, 10 years, 25 years, 50 years, and 100 years. Fig. 2 shows the IDF curves derived from observed data at the

Ha Dong meteorological station. It can be observed from the figure that longer durations of rainfall are associated with lower rain intensity values, while longer return periods correspond to higher rain intensity values. Specifically, the figure shows that for a return period of 2 years, the rain intensity value is the smallest, whereas for a return period of 100 years, the rain intensity value is the highest.

3.4. Determine the rainfall intensity formula for Ha Dong station and compare

Based on the formula for determining rain intensity in Vietnam Standard 7957:2008, we research and choose the formula to calculate rain intensity as [13]:

$$q = \frac{0.36 \cdot A(1 + C \cdot \log T)}{(d + b)^n} \tag{14}$$

where q is the rainfall intensity (mm/hour); T is the return period (year); d is the rainfall duration (minutes); A, C, b, n are parameters determined based on the IDF table of the rain gauge station.

The parameters A, C, b, n are determined so that the square root error (i.e. the average of the sum of squared errors between the calculated value by the formula and the IDF data is the smallest) is the smallest. Here the root mean square error (RMSE) [3, 6] is determined by:

$$RMSE = \sqrt{\frac{1}{72} \sum_{k=1}^{72} \left(\frac{0.36 \cdot A(1 + C \cdot \log T_k)}{(d_k + b)^n} - q_k \right)^2} \tag{15}$$

where (T_k, d_k, q_k) are the triples of the return period, rainfall duration, and rainfall intensity determined from Table 9. It can be seen that there are 12 rainfall durations and 6 the return periods, so there will be $12 \cdot 6 = 72$ sets of values. The optimized equation for this study is presented as follows:

$$f(A, C, b, n) = \sum_{k=1}^{72} \left(\frac{0.36 \cdot A(1 + C \cdot \log T_k)}{(d_k + b)^n} - q_k \right)^2 \to \min \tag{16}$$

Usually, the value of b is chosen to be a non-negative integer. Therefore, with the chosen value of b, the objective function is considered as a function of three variables. With the IDF data of Ha Dong station, using the generalized reduced gradient (GRG) method (used to optimize a non-linear smooth function), we obtain the smallest square root errors corresponding to non-negative integer values of b.

Therefore, to obtain the smallest square root error, the rainfall intensity formula for Ha Dong station is determined as follows

$$q = \frac{0.36 \cdot 2320 \cdot (1 + 0.655 \cdot \log T)}{(d + 9)^{0.633}} \tag{17}$$

With the IDF data available, there is a comparison table of the newly obtained formula and some other existing formulas.

Based on the results, it can be observed that (1) Probability Distribution Function Selection: Using the Chi-square statistical test index (χ^2), the study identified the Gumbel distribution as the most suitable function for constructing the rainfall frequency curve. This distribution demonstrated the highest correlation with the observed variation in short-duration rainfall data at the Ha Dong station. (2) Design Rainfall Intensity Equation: By applying the least squares method, optimized using the Generalized Reduced Gradient Nonlinear approach, the study developed a design rainfall intensity equation for the Ha Dong station based on IDF data up to 2023. The parameter values (climatic variables) of the equation were determined as $A = 2320, C = 0.655, b = 9$, and $n = 0.633$.

Table 10. Coefficients with values of b

b	A	C	n	$RMSE$
0	886	0.664	0.435	12.389
1	995	0.664	0.461	10.228
2	1108	0.663	0.485	11.432
3	1229	0.663	0.507	9.805
4	1357	0.663	0.528	8.706
5	1494	0.663	0.549	8.089
6	1639	0.662	0.568	7.893
7	1796	0.662	0.587	8.030
8	1963	0.662	0.606	8.404
9	2320	0.655	0.633	7.736
10	2335	0.662	0.641	9.547
11	2543	0.662	0.659	10.211
12	2765	0.661	0.675	10.895
13	3005	0.661	0.692	11.582
14	3263	0.661	0.709	11.035
15	3541	0.661	0.725	11.633
16	3840	0.661	0.741	12.215
17	4162	0.661	0.757	12.780
18	4162	0.661	0.757	14.667
19	4883	0.661	0.788	13.855
20	5285	0.661	0.803	14.364

Table 11. Result comparison of the proposed equation with other equations

Name	Formula	RMSE	Correlation coefficient
Proposed equation	$q = \dfrac{0.36 \cdot 2320 \cdot (1 + 0.655 \cdot \log T)}{(d + 9)^{0.633}}$	7.736	0.9985
TCVN 7957	$q = \dfrac{0.36 \cdot 5890 \cdot (1 + 0.65 \cdot \log T)}{(d + 20)^{0.84}}$	10.365	0.9959
IMHEN equation	$q = \dfrac{0.36 \cdot 2511 \cdot (1 + 0.8938 \cdot \log T)}{(d + 14)^{0.7143}}$	10.200	0.9961

(3) Evaluation of reliability and accuracy: The proposed equation was validated against reliability and accuracy criteria. It achieved a root mean square error (RMSE) of 7.736, significantly lower than the existing equations, which have RMSE values of 10.365 and 10.200. Furthermore, the proposed equation showed an exceptionally high correlation coefficient, close to 1. These findings indicate that the proposed equation provides a more accurate and reliable representation of the IDF data compared to the existing equations.

4. Conclusions

This paper indicates the need to update parameters in rainfall intensity equations to account for fluctuations in influencing factors, such as changes in rainfall characteristics and local climate. By analyzing and comparing commonly used probability distribution functions, the study identified the Gumbel distribution as the most suitable for modeling short-duration rainfall data in Hanoi. Furthermore, the study established a relationship between the IDF curve and the rainfall intensity equation based on rainfall observation data from the Ha Dong meteorological station in Hanoi, spanning the period from 1973 to 2023.

The proposed equation of design rainfall intensity was validated by comparing with other existing equations (Equations suggested by TCVN 7957 and IMHEN). The results demonstrate that the proposed design rainfall intensity equation aligns more closely with IDF-derived data, offering improved accuracy for calculating rainfall intensity. The results of this study also align with those of other studies when using the latest short-duration rainfall data. Additionally, the use of IDF curves in formulating design rainfall intensity demonstrates clear efficiency, as it more accurately reflects actual observation results compared to previous methods. This provides a valuable tool for addressing challenges in designing rainwater drainage systems for residential and urban areas.

Acknowledgments

The authors would like to thank the PhD Scholarship Programme of Vingroup Innovation Foundation (VINIF), code VINIF.2023.TS.007 for helping us complete this research.

References

[1] Consortium of Northern Water Resources Planning and Investigation Federation, Institute of Hydrology, Environment and Climate Change (2012). *Report of the Project "Building a flood risk map of Hanoi City considering the impact of Climate Change"*. Hanoi.

[2] Bernard, M. M. (1932). Formulas for rainfall intensities of long durations. *Transactions of the American Society of Civil Engineers*, 96:592–624.

[3] Institute of Meteorology, Hydrology and Climate Change (2019). *Project Summary Report "Building the Rainfall Intensity – Duration – Frequency (IDF) Curves to Strengthen State Management of Hydrometeorological Information for Socio-Economic Development"*. Hanoi.

[4] Dung, U. Q. (1996). Research to complete the method of determining the calculated rainwater volume when designing drainage systems for urban areas in Vietnam. PhD thesis, Hanoi University of Civil Engineering, Hanoi.

[5] Hong, N. T. V. (2018). Research on improving the method of determining rainfall pattern and design drainage flow for drainage systems in the Northern Delta region. PhD thesis, University of Water Resources, Hanoi.

[6] Giang, L. S., Binh, N. H. T., Hoa, N. T. T., Thinh, D. N. (2023). Constructing combined rainfall - water level curve for designing surface drainage system in Ho Chi Minh City. *Can Tho University Journal of Science*, 59:213–220.

[7] Trang, B. T. T., Hanh, N. T. H., Thuong, N. T. H., Thuan, N. V. (2018). Analysis of Storm Pattern for Design Urban Drainage System in the Monsoon Areas of Vietnam. *Journal of Environmental Science and Engineering*, 7:49–68.

[8] Chow, V. T. (1988). *Handbook of Applied Hydrology*. McGraw-Hill, New York.

[9] Hosking, J. R. M., Wallis, J. R. (1977). *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press.

[10] Maity, R. (2018). *Statistical Methods in Hydrology and Hydroclimatology*. Springer Transactions in Civil and Environmental Engineering, Springer Singapore.

[11] International Hydrological Programme IHP-VII (2008). *Technical Documents in Hydrology No. 2: Asian Pacific FRIEND, Rainfall Intensity Duration Frequency (IDF), Analysis for the Asia Pacific Region*. Reported by Regional Steering Committee for Southeast Asia and the Pacific, UNESCO Office, Jakarta.

[12] Hanoi People's Committee (2020). *Project on developing drainage and wastewater treatment service prices in Hanoi*. Completed according to social feedback from the Vietnam Fatherland Front Committee of Hanoi and comments from the Department of Finance No. 5835/STC-QLG dated September 10, 2020; No. 6816/STC-QLG dated October 21, 2020.

[13] Vietnam Technical Standard (TCVN 7957:2008) (2008). *Drainage and sewerage - External Networks and Facilities - Design Standard*. Ministry of Construction.