■Ngày nhận bài: 17/5/2023 ■Ngày sửa bài: 15/6/2023 ■Ngày chấp nhận đăng: 14/7/2023

Lựa chọn phân phối xác suất để xây dựng đường cong IDF cho tính toán hệ thống tiêu thoát nước mưa khu vực Hà Nội

Choose probability distribution to constructe IDF curve for calculation of rainwater drainage system in Hanoi area

> **HÀ XUÂN ÁNH**
Khoa Kỹ thuật hạ tầng và Môi trường đô thị, Trường Đại học Kiến trúc Hà Nội;
Email: haxuananh.hau@gmail.com

TÓM TẮT

Bài viết đề cập đến việc so sánh và lựa chọn phân phối lý thuyết để xây dựng đường cong mối quan hệ IDF (cường độ - thời đoạn - tần suất) trong tính toán hệ thống tiêu thoát nước cho TP Hà Nội. Chương trình Thủy văn quốc tế (International Hydrological Program, IHP) khu vực châu Á - Thái Bình Dương thuộc Cơ quan thường trực UNESCO có trụ sở ở Jakarta, năm 2005 đã khuyến nghị các nước thành viên, trong đó có Việt Nam xây dựng các hệ thống đường cong IDF để thúc đẩy phát triển bền vững cơ sở hạ tầng đã được xuất bản vào năm 2008. Theo đó, Việt Nam nên sử dụng phân phối Person III và Log-Person III để xây dựng đường cong IDF. Tuy nhiên, cho đến nay vẫn chưa có nhiều nghiên cứu về việc so sánh, lựa chọn phân phối lý thuyết. Bài báo tập trung vào việc tính toán việc sử dụng phân phối Person III và Log-Person III để xây dựng đường cong IDF cho TP Hà Nội. Qua đó, so sánh, phân tích và đưa ra khuyến nghị về việc sử dụng phân phối lý thuyết cho khu vực Hà Nội.

Từ khoá: Cường độ mưa; IDF; phân phối xác suất; thống kê khí tượng; phân phối thực nghiệm.

ABSTRACT

The article addresses the comparison and selection of theoretical distributions to build the IDF (intensity - duration - frequency) relationship curve in calculating the drainage system for Hanoi city. The International Hydrological Program (IHP) for the Asia-Pacific region of the UNESCO Permanent Agency headquartered in Jakarta, in 2005 recommended that member countries, including Vietnam, build IDF curve systems to promote sustainable infrastructure development were published in 2008. Accordingly, Vietnam should use Person III and Log-Person III distributions to construct IDF curves. However, up to now there have not been many studies on the comparison and selection of theoretical distributions. The article focuses on calculating the use of Person III and Log-Person III distributions to build the IDF curve for Hanoi city. Thereby, compare, analyze and make recommendations on the use of theoretical distribution for Hanoi area.

Keyword: Rainfall intensity; IDF; probability distribution; meteorological statistics; empirical distribution.

1. MỞ ĐẦU

Trong nghiên cứu khí tượng thuỷ văn, các phương pháp thống kê thường được áp dụng. Các phương pháp khảo sát, phân tích số liệu dựa trên các đặc trưng thống kê cho phép chỉ ra những thuộc tính của các đặc trưng yếu tố khí tượng, khí hậu căn cứ vào những tập số liệu cụ thể có được từ quan trắc thực tế. Tuy nhiên, do dung lượng mẫu trong nghiên cứu khí tượng thường không quá lớn, trong nhiều trường hợp những kết quả nhận được có thể phản ánh không chính xác bản chất của quá trình khí tượng được xem xét. Để khắc phục tình trạng đó, bên cạnh việc nghiên cứu dựa trên dữ liệu mẫu, chúng ta sẽ sử dụng các phân phối lý thuyết và tiến hành xấp xỉ số liệu thực nghiệm bởi những phân phối lý thuyết phù hợp. Việc sử dụng các phân phối lý thuyết để xấp xỉ số liệu thực nghiệm thực chất là lý tưởng hoá dữ liệu, coi kết quả thực nghiệm là kết quả của một số công thức toán học liên quan đến các đại lượng ngẫu nhiên. Mặc dù trong nhiều trường hợp sự biểu diễn này có độ chính xác cao, nhưng về cơ bản chúng chỉ cho chúng ta các giá trị gần đúng so với số liệu thực nghiệm. Việc xấp xỉ số liệu thực nghiệm bởi các phân phối lý thuyết mang lại nhiều ưu điểm (xem [1, 2, 6].

- Trong nhiều trường hợp, chúng ta phải lặp đi lặp lại những tính toán thống kê các đặc trưng mẫu cho một địa điểm hoặc một vùng không gian nhất định. Quá trình tính toán đó thường khá cồng kềnh, phức tạp và khó tránh khỏi những sai sót. Nếu có một phân phối lý thuyết phù hợp với số liệu thì thay cho việc phải khảo sát đầy đủ ta chỉ cần một vài tham số của phân phối này. Điều này khiến cho việc tính toán trở nên đơn giản hơn và tiết kiệm thời gian.

- Phân phối lý thuyết cho phép nội suy số liệu bị mất (hoặc không có số liệu), từ đó lấp đầy các khoảng trống số liệu. Điều này rất phù hợp trong nghiên cứu khí tượng thuỷ văn do trong nhiều trường hợp, số liệu của một năm nào đó bị mất.

- Do hạn chế của dung lượng mẫu, đặc biệt số liệu khí tượng thuỷ văn thường có dung lượng nhỏ, số liệu thực nghiệm chỉ phản ảnh được sự biến đổi của đặc trưng yếu tố trong phạm vị biến đổi của tập mẫu. Việc sử dụng phân phối lý thuyết cho phép chúng ta ước lượng xác suất cho những sự kiện nằm ngoài phạm vi của tập mẫu, đặc biệt là ước lượng xác suất xảy ra các tình huống cực đoan.

Thông thường, sau khi xây dựng hàm phân phối thực nghiệm, ta cần nghiên cứu, đánh giá, xem xét, lựa chọn phân phối lý thuyết sao cho nó phù hợp nhất với phân phối thực nghiệm.

Chương trình Thủy văn quốc tế (International Hydrological Program, IHP) khu vực châu Á - Thái Bình Dương thuộc Cơ quan thường trực UNESCO có trụ sở ở Jakarta, năm 2005 đã khuyến nghị các nước thành viên, trong đó có Việt Nam xây dựng các hệ thống đường cong IDF để thúc đẩy phát triển bền vững cơ sở hạ tầng đã được xuất bản vào năm 2008 (xem [4]). Trong đó, thời đoạn đo mưa, một số quy trình phân tích tần suất đã được đề xuất và một số công thức tính IDF được sử dụng trong các nước thành viên đã được nghiên cứu dựa vào số liệu tại một số trạm quan trắc mưa do các nước thành viên cung cấp. Thời đoạn đo mưa lựa chọn để tính toán phổ biến trong các nước ở khu vực là 15 thời đoạn, bao gồm: 5p, 10p, 15p, 30p, 45p, 60p, 90p, 120p, 150p, 180p, 360p, 720p và 1440p. Theo đó, các hàm phân bố tần suất được khuyến nghị sử dụng để tính tần suất là: Pearson III (Trung Quốc, Hàn Quốc, Việt Nam), Log-Pearson-III (Indonexia, Hàn Quốc, Việt Nam), Gumbel (EV1, Australia, New Zealand, Japan, Phillipin, Malaixia). Các công thức tính cường độ mưa - thời đoạn - tần suất được sử dụng phổ biến trong khu vực như Talbot, Bernard, Kimijima, Wenzel, Sherman.

Như thế, Chương trình Thủy văn quốc tế (International Hydrological Program, IHP) khu vực châu Á - Thái Bình Dương đã khuyến nghị nên sử dụng phân phối Pearson III và Log-Person III để xây dựng đường cong IDF. Đối với khu vực Hà Nội, theo tìm hiểu của chúng tôi, chưa có so sánh giữa hai phân phối trên. Bài báo này của chúng tôi nhằm so sánh thông qua một ví dụ minh hoạ việc sử dụng hai phân phối để tính toán cường độ mưa trong thời đoạn 20 phút với số đo tại trạm Láng (Hà Nội).

2. MỘT SỐ KIẾN THỨC CƠ SỞ VỀ LÝ THUYẾT XÁC SUẤT VÀ THỐNG KÊ TOÁN HỌC

2.1 Hàm mật độ xác suất và hàm phân phối tích luỹ

Trong toán học, *hàm mật độ xác suất* (tiếng Anh là *Probability density function* hay PDF) dùng để biểu diễn một phân bố xác suất theo tích phân. Một cách trực quan, hàm mật độ xác suất có thể được coi là phiên bản được *làm mịn* của một biểu đồ tần số: nếu ai đó liên tiếp đo đạc bằng thực nghiệm các giá trị của một biến ngẫu nhiên liên tục và tạo một biểu đồ tần số mô tả tần suất tương đối của các miền biến thiên của kết quả, thì biểu đồ tần số đó sẽ trông giống với mật độ xác suất của biến ngẫu nhiên đó (giả sử rằng biến được lấy mẫu đủ thường xuyên và các miền biến thiên của kết quả là đủ nhỏ).

Trong lý thuyết xác suất, *hàm phân phối tích lũy* (tiếng Anh: *Cumulative distribution function* hay viết tắt *CDF*) mô tả đầy đủ phân phối xác suất của một biến ngẫu nhiên giá trị thực *X*. Với mỗi số thực *x*, hàm phân phối tích lũy được định nghĩa như sau $F(x) = P(X \leq x)$

trong đó vế phải biểu diễn xác suất mà đại lượng ngẫu nhiên *X* lấy giá trị nhỏ hơn hay bằng *x*. Do đó, $P(a \leq X \leq b) = F(b) - F(a)$.

Theo quy ước, chữ *F* hoa được dùng cho hàm phân phối tích lũy, còn chữ *f* thường được dùng cho hàm mật độ xác suất và hàm khối xác suất. Lưu ý rằng trong định nghĩa trên, dấu "nhỏ hơn hay bằng" ('≤') có thể được thay bằng dấu "nhỏ hơn" ('<'). Điều này không làm thay đổi giá trị của $F(x)$ đối với các đại lượng ngẫu nhiên liên tục.

Khi đại lượng ngẫu nhiên X có hàm mật độ xác suất $f(x)$ thì hàm phân phối tích luỹ $F(x)$ được xác định $F(x) = \int_{-\infty}^{x} f(t)dt$.

2.2 Một số đặc trưng quan trọng của đại lượng ngẫu nhiên

Giá trị kỳ vọng (*expected value*), hoặc trung bình (*mean*) của một biến ngẫu nhiên là trung bình có trọng số của tất cả các giá trị cụ thể của biến đó, hay là được tính bằng tổng các tích giữa xác suất xảy ra của mỗi giá trị có thể của biến với giá trị đó. Giá trị kỳ vọng này được chứng minh xấp xỉ giá trị trung bình thực nghiệm khi số lượng được quan sát đủ lớn. Lưu ý rằng bản thân giá trị đó có thể không được mong đợi theo nghĩa thông thường; nó có thể ít có khả năng xảy ra hoặc không thể xảy ra.

Khi đại lượng ngẫu nhiên liên tục X có hàm mật độ xác suất $f(x)$ thì kỳ vọng của X là $\mu = E(X) = \int_{-\infty}^{+\infty} x.f(x)dx$.

Phương sai (*variance*) của một đại lượng ngẫu nhiên là một độ đo sự phân tán thống kê của đại lượng đó, nó hàm ý các giá trị của đại lượng đó thường ở cách giá trị kỳ vọng bao xa. Khi đại lượng ngẫu nhiên X có hàm mật độ xác suất $f(x)$ thì phương sai của X là $Var(X) = \int_{-\infty}^{+\infty} x^2.f(x)dx - \left(E(X)\right)^2$.

Độ lệch chuẩn, hay độ lệch tiêu chuẩn (tiếng Anh: *standard deviation*) là một đại lượng ngẫu nhiên dùng để đo mức độ phân tán của của các giá trị. Có thể tính ra độ lệch chuẩn bằng cách lấy căn bậc hai của phương sai, tức là $\sigma = \sqrt{Var\,(X)}$.

Độ xiên (skewness) là một đại lượng đo lường mức độ bất đối xứng của phân phối xác suất của một biến ngẫu nhiên. Nó còn tên gọi nữa là *hệ số bất đối xứng*. Để tính ra giá trị độ xiên, chúng ta sử dụng công thức: $\gamma = \frac{\mu_3}{\sigma^3}$ ở đó σ là độ lệch chuẩn và μ_3 là moment trung tâm bậc 3 của X, tức là $\mu_3 = \int_{-\infty}^{+\infty} x^3 f(x)dx$.

2.3 Tần suất và tần suất lũy tích

Tần số là số lần xuất hiện một giá trị nào đó trên tổng số lần thử nghiệm hay quan sát. *Tần suất* là tỷ số giữa số lần xuất hiện một giá trị nào đó trên tổng số lần thử nghiệm. Trong khí tượng thủy văn, tần suất là tỷ số giữa số lần xuất hiện một trị số nào đó so với tổng số lần quan trắc. *Tần suất luỹ tích* là tỷ số giữa tổng số lần xuất hiện một trị số lớn hơn hoặc bằng một trị số nào đó so với tổng số lần thử nghiệm. Tần suất lũy tích được tính bằng cách cộng dồn các tần suất từ trên xuống cho đến vị trí tương ứng với biểu hiện mà ta đang muốn tính tần suất luỹ tích. Trong khí tượng thủy văn, khi nghiên cứu tới trị số x, người ta thường quan sát có bao nhiêu lần giá trị của đại lượng được nghiên cứu lớn hơn hay bằng x. Do đó, người ta thường quan tâm đến tần suất luỹ tích $\frac{m}{n}$, nghĩa là trong n năm quan trắc có m lần xuất hiện trị số lớn hơn hoặc bằng một trị số mà ta đang xét.

2.4 Các tham số thống kê thường sử dụng trong nghiên cứu khí tượng thuỷ văn

Số trung bình $\bar{x}$ là giá trị trung bình số học của chuỗi số liệu. Số trung bình là một trong những số đặc trưng cơ bản và quan trọng nhất của chuỗi số liệu. Khi số lần quan sát đủ lớn, số trung bình có xu hướng tiến sát đến kỳ vọng toán học của đại lượng ngẫu nhiên được quan tâm. Tuy nhiên, số trung bình lại bị ảnh hưởng bởi các giá trị cực đoan (giá trị “quá lớn” hoặc “quá bé” một cách đặc biệt), nhất là khi chuỗi số liệu ngắn. Do chuỗi số liệu trong khí tượng thuỷ văn thường không quá dài nên số trung bình có nguy cơ ảnh hưởng rất mạnh bởi số liệu mang tính cực đoan.

Phương sai mẫu là trung bình của bình phương độ lệch của các giá trị xung quanh giá trị trung bình $s^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}$.

Độ lệch mẫu (Tiếng Anh: standard deviation) là một đại lượng thống kê mô tả dùng để đo mức độ phân tán của một tập dữ liệu đã được lập thành bảng tần số. Độ lêch mẫu là căn bậc hai của trung bình của bình phương độ lệch của các giá trị xung quanh giá trị trung bình $s = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}}$.

Độ thiên lệch (còn gọi là *độ xiên, hệ số bất đối xứng,* Tiếng Anh: skewness coefficient) $C_s = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^3}{s^3}$ là số dùng để đo độ bất đối xứng của chuỗi số liệu. Nếu đại lượng ngẫu nhiên có phân phối đối xứng thì $C_s = 0$. Nếu $C_s > 0$ thì mật độ phân bố có dạng đuôi lệch phải, còn nếu $C_s < 0$ thì mật độ phân bố có dạng đuôi lệch trái. .

Các công thức tính s, C_s được tính toán phía trên chỉ phù hợp với mẫu có dung lượng n lớn. Đối với hiện tượng thuỷ văn, do thường có chuỗi số liệu không dài, người ta đề nghị điều chỉnh bằng cách thay số hạng n bằng các số hạng $n-1$ (đối với s) và $n-2$ (đối với tham số C_s), tức là

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}}, C_s = \frac{\frac{1}{n-2}\sum_{i=1}^{n}(x_i-\bar{x})^3}{s^3}$$

2.5 Khoảng thời gian lặp lại như định nghĩa ở bên dưới

Khoảng thời gian lặp lại T (Tiếng Anh: return period), là khoảng thời gian hiện tượng có thể xuất hiện trở lại sau khi đã xuất hiện. Trong khí tượng thuỷ văn, khoảng thời gian lặp lại T thường được tính dựa trên xác suất vượt quá (TNTA: exceedence probability), ký hiệu là P theo công thức $T = \frac{1}{P}$. Chúng ta tính xác suất vượt quá (Tiếng Anh: exceedence probability), ký hiệu là P theo công thức $P = \frac{k}{n+1}$ trong đó k là số thứ tự của số liệu (được sắp xếp theo thứ tự giảm dần) và n là số phần tử trong chuỗi số liệu. Do các thành phần kế cận của chuỗi số liệu thường cách nhau 1 năm nên đơn vị đo của T là năm. Cần chú ý rằng giá trị của T được xét trên phương diện thống kê nên không thể hiểu rằng cứ trong T năm thì hiện tượng xảy ra 1 lần.

2.6 Phân tích tần suất

Khi cường độ của một sự kiện thủy khí hậu khác biệt đáng kể so với mức trung bình hoặc phạm vi cường độ thông thường, thì các sự kiện như vậy được gọi là các sự kiện cực đoan. Cái này có thể diễn ra trong một ngày hoặc một khoảng thời gian, ví dụ: bão lớn, lũ quét, hạn hán. Các loại sự kiện khí hậu cực đoan này ảnh hưởng đến hệ thống mức độ lớn. Phân tích tần suất được thực hiện để xác định tần suất xảy ra (hoặc xác suất xảy ra) của các sự kiện cực đoan như vậy.

Phân tích tần số thường đề cập đến phân tích tần số cố định giả định dữ liệu phải cố định. Hầu hết các chức năng phân phối tần số trong thủy khí hậu các nghiên cứu có thể được thể hiện dưới dạng phương trình sau, được gọi là tổng quát phương trình phân tích tần số, được cho bởi $x_T = \bar{x} + K_T.s$ trong đó x_T là giá trị đại lượng quan sát tương ứng với thời gian lặp lại T năm; $\bar{x}$ là giá trị trung bình của đại lượng ngẫu nhiên; s là độ lệch chuẩn của đại lượng ngẫu nhiên.

2.7 Đường tần suất kinh nghiệm

Đường tần suất kinh nghiệm là đường cong biểu thị quan hệ giữa tần suất P với giá trị x_i tương ứng, trong đó

$$P = P(X \geq x_i).$$

Với mỗi số liệu x_i từ mẫu (chuỗi số liệu thuỷ văn) xác định được tương ứng là P. Chấm quan hệ $P \sim x_i$ tạo thành một băng điểm có xu thế một đường cong của hàm phân bố tần suất. Vẽ một đường cong trơn đi qua trung tâm các điểm kinh nghiệm, ta gọi đường đó là đường tần suất kinh nghiệm. Để xây dựng đường tần suất trước hết cần phải lập bảng tính tần suất, sau đó chấm các điểm quan hệ giữa P và x_i lên giấy, từ đó xác định đường tần suất kinh nghiệm. Cách làm được tiến hành theo trình tự như sau:

- Bước 1: Sắp xếp chuỗi số liệu theo thứ tự từ lớn đến nhỏ và đánh số thứ tự kèm theo.

- Bước 2: Tính tần suất P theo công thức của Weibull và Kritsky-Menken: $P = \frac{m}{n+1} \times 100\%$.

Trong đó m là số thứ tự của x_i (đã sắp xếp từ lớn đến nhỏ) còn n là số phần tử của chuỗi số liệu thống kê.

- Bước 3: Chấm các điểm quan hệ $P \sim x_i$ (gọi là điểm kinh nghiệm) lên giấy tần suất.

- Bước 4: Vẽ đường cong đi qua các điểm kinh nghiệm, đó chính là đường tần suất kinh nghiệm (xem Hình 1 phía dưới).

3. TỔNG QUAN VỀ CƯỜNG ĐỘ MƯA TÍNH TOÁN

3.1 Khái niệm về cường độ mưa

Trong thiết kế thoát nước mưa, một trong những yếu tố đầu tiên là cần phải xác định được cường độ mưa q của khu vực thiết kế. Thông thường người ta sử dụng một trận mưa rào thiết kế hoặc biến cố trong đó có xét tới quan hệ giữa cường độ mưa, thời gian mưa và tần suất.

Cường độ mưa là chiều cao lớp nước mưa trên một đơn vị thời gian. Nó có thể là cường độ mưa tức thời hoặc cường độ mưa trung bình trong một thời gian mưa. Trong tính toán, ta thường dùng cường độ mưa trung bình, cường độ mưa được biểu thị bởi: $i = \frac{h}{t}$ trong đó: i - cường độ mưa còn được gọi là cường độ mưa theo lớp nước; h - độ sâu mưa, đo bằng mm; t - thời gian mưa, thường tính bằng giờ.

Cường độ mưa theo thể tích được tính theo cường độ mưa theo lớp nước $q = 166{,}7 * i$; trong đó: q - Cường độ mưa theo thể tích, l/s.ha; 166,7 là hệ số chuyển đổi; i - Cường độ mưa theo lớp nước, mm/phút.

3.2. Quan hệ cường độ mưa - thời gian - tần suất (IDF)

Khi thiết kế hệ thống thoát nước mưa, trước hết cần xác định lượng mưa dùng trong tính toán. Thông thường một trận mưa rào thiết kế được sử dụng trong đó có xét tới quan hệ cường độ mưa, thời gian và tần suất hay thời kỳ lặp lại, gọi tắt là quan hệ IDF. Thông thường quan hệ IDF được biểu thị dưới dạng đồ thị, trong đó thời gian mưa được đặt trên trục hoành, cường độ mưa trên trục tung và một họ đường cong IDF, mỗi đường tương ứng với một chu kỳ lặp lại. Khi có đủ các số liệu về mưa, các đường cong IDF có thể được xây dựng bằng phân tích tần suất. Ứng với mỗi thời đoạn mưa lựa chọn, lượng mưa lớn nhất năm được thống kê từ tập số liệu mưa thực đo, sau đó tiến hành phân tích tần suất đối với chuỗi số liệu năm.

4. VÍ DỤ LỰA CHỌN PHÂN BỐ XÁC SUẤT

4.1. Số liệu đo mưa 30 phút tại trạm Láng

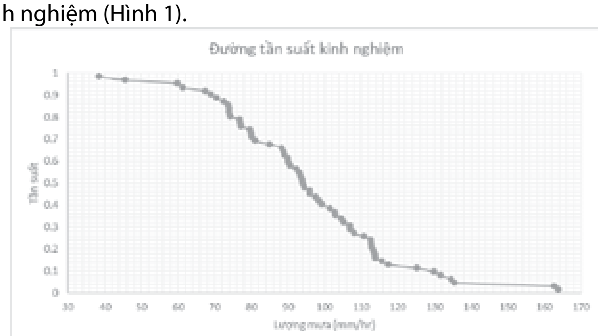
Ở đây, chúng tôi minh hoạ việc lựa chọn xây dựng đường cong IDF từ số liệu đo mưa tại Trạm Láng, Tổng cục khí tượng Thuỷ văn, Hà Nội bằng cách sử dụng các phân phối Person-III, Log-Person-III và so sánh.

Dưới đây là số liệu đo mưa 30 phút tại Trạm Láng (Hà Nội) từ 1961 đến 2021 (xem Bảng 1). Để đưa ra bộ số liệu này, trong mỗi năm, trạm Láng tiến hành đo chiều cao của lớp nước mưa trong các thời đoạn 30 phút. Từ đó, chọn ra chiều cao của lớp nước mưa lớn nhất trong thời gian 30 phút của một năm.

Bảng 1. *Số liệu đo mưa tại trạm Láng (1961-2021) thời đoạn 30 phút*

Năm	Lượng mưa	Ngày	Năm	Lượng mưa	Ngày	Năm	Lượng mưa	Ngày	Năm	Lượng mưa	Ngày
1961	48.0	30-VII	1977	52.6	15-VII	1993	51.4	22-VI	2009	50.7	20-VII
1962	39.9	9-V	1978	44.5	17-V	1994	67.2	7-VII	2010	58.7	13-VII
1963	47.0	2-VI	1979	35.3	16-IX	1995	30.6	9-IX	2011	39.8	2-VIII
1964	53.5	1-VI	1980	48.0	18-X	1996	22.8	23-VII	2012	38.6	22-VII
1965	37.1	12-VI	1981	36.9	6-IX	1997	56.6	9-V	2013	49.6	16-VII
1966	44.1	3-VI	1982	46.5	20-VII	1998	53.4	9-VI	2014	36.2	7-VI
1967	54.0	9-VII	1983	48.8	21-VI	1999	64.9	15-VII	2015	56.9	26-VIII
1968	56.8	6-VI	1984	56.4	18-IX	2000	40.0	8-VII	2016	45.3	18-VIII
1969	40.5	8-VI	1985	42.5	27-VII	2001	57.8	22-VII	2017	45.0	13-X
1970	38.4	30-VIII	1986	62.6	18-VI	2002	56.3	7-VI	2018	46.9	12-V
1971	47.2	16-IX	1987	36.9	25-IX	2003	55.4	24-V	2019	46.7	29-IV
1972	38.5	22-VIII	1988	29.9	6-VIII	2004	36.8	17-VIII	2020	45.1	24-I
1973	67.7	27-IV	1989	34.5	11-VI	2005	81.9	3-V	2021	51.5	11-V
1974	46.1	4-VII	1990	33.7	20-VII	2006	81.3	20-XI			
1975	44.4	15-VII	1991	49.2	7-V	2007	56.2	26-VIII			
1976	52.3	30-VI	1992	19.2	10-V	2008	65.8	15-VII			

Trước hết chúng ta xây dựng đường tần suất kinh nghiệm. Đầu tiên chúng ta xếp dãy số liệu trên theo thứ tự giảm dần và chuyển về lượng mưa trong 1 giờ. Sau đó, sử dụng công thức của Weibull và Kritsky-Menken: $P = \frac{m}{n+1}$, trong đó m là số thứ tự của số liệu và n là số phần tử dữ liệu (ở đây $n = 61$). Từ đó chúng ta vẽ được tần suất kinh nghiệm (Hình 1).



Hình 1. Đường tần suất kinh nghiệm

4.2. Sử dụng phân phối Person III

Đại lượng ngẫu nhiên X có phân phối Person III nếu hàm mật độ xác suất là

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

với $\Gamma(\alpha)$ là hàm Gamma của biến số α

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1}\, e^{-t}dt.$$

Ta có

$$\alpha = \frac{4}{C_s^2}, \beta = \left(\frac{C_v.C_s}{2}\right)\bar{x}, x_0 = \bar{x} - \sigma_x\sqrt{\alpha}$$

Sử dụng phân phối Person III, chúng ta tính được trung bình cộng của cột R là $avg(R) = 95.816$. Từ đó, thiết lập được các cột $(R - avg(R))^2, (R - avg(R))^3, (R - avg(R))^2 + (R - avg(R))^3$ (xem Bảng 5 phần Phụ lục)

Từ đó ta tính được phương sai của mẫu số liệu $s^2 = \frac{\sum_{i=1}^{n}(R_i - \mathrm{avg}(R_i))^2}{n-1} = 573.716$ và độ lệch mẫu $s = 23.952$.

Qua đó, chúng ta tính được *độ xiên* $C_s = \frac{n.\sum_{i=1}^{n}(R_i - \mathrm{avg}(R_i))^3}{(n-1)(n-2).s^3} = 0.432$.

Sử dụng bảng hệ số tần suất và hệ số bất đối xứng (Bảng giá trị K_T tương ứng với độ lệch C_s và thời gian lặp lại T năm của phân phối Pearson III (được đưa ra bởi Haan năm 1977) để tìm các giá trị k cho các thời gian lặp lại 2, 5, 10, 25, 50, 100 và 200 năm.

Vì hệ số bất xứng $C_s = 0.432$ nằm giữa hai hệ số $C_s = 0.4$ và $C_s = 0.5$ trong bảng nên chúng ta có thể *xấp xỉ tuyến tính giữa hai số để có được giá trị k phù hợp.*

Vì $0.432 = 0.68 \times 0.4 + 0.32 \times 0.5$ nên chúng ta có $k(T, C_s = 0.432) = 0.68 \times k(T, C_s = 0.4) + 0.32 \times k(T, C_s = 0.5)$.

Chúng ta tính được giá trị k ứng với các thời gian lặp lại T và hệ số bất đối xứng $C_s = 0.432$. Từ đó chúng ta tính được giá trị R tương ứng các thời gian lặp lại T theo công thức $R(T) = avg(R) + k(T, C_s = 0.432) \times s$.

Bảng 2. *Giá trị* $k(T, C_s = 0.432)$ *và R*

Thời gian lặp lại (năm)	$C_s = 0.4$	$C_s = 0.5$	$C_s = 0.432$	R
2	-0.066	-0.083	-0.07144	94.1
5	0.816	0.808	0.81344	115.3
10	1.317	1.323	1.31892	127.4
25	1.88	1.91	1.8896	141.1
50	2.261	2.311	2.277	150.4
100	2.615	2.686	2.63772	159.0
200	2.949	3.041	2.97844	167.2

4.3. Sử dụng phân phối Log-Person III

Đại lượng ngẫu nhiên Y có phân phối Log-Person III nếu đại lượng ngẫu nhiên $Y = \log X$ tuân theo phân phối Person III.

Sử dụng phân phối Log-Person III, tính log của lượng mưa quy đổi và trung bình mẫu của log này là $avg(log\ R) = 1.967$. Từ đó chúng ta thiết lập các cột $\left(log\ R - avg(log\ R)\right)^2, \left(log\ R - avg(log\ R)\right)^3$, và $\left(log\ R - avg(log\ R)\right)^2 + \left(log\ R - avg(log\ R)\right)^3$ (xem Bảng 6 phần Phụ lục).

Từ đó ta tính được phương sai của mẫu số liệu log hoá

$$s^2 = \frac{\sum_{i=1}^{n}\left(\log R_i - \mathrm{avg}(\log R_i)\right)^2}{n-1} = 0{,}013$$

và độ lệch mẫu $s = 0{,}114$.

Bên cạnh đó, chúng ta tính được *độ xiên*

$$C_s = \frac{n.\sum_{i=1}^{n}\left(\log R_i - \mathrm{avg}(\log R_i)\right)^3}{(n-1)(n-2).s^3} = -0{,}676.$$

Sử dụng bảng hệ số tần suất và hệ số bất đối xứng (Bảng giá trị K_T tương ứng với độ lệch C_s và thời gian lặp lại T năm của phân phối Pearson III để tìm các giá trị k cho các thời gian lặp lại 2, 5, 10, 25, 50, 100 và 200 năm.

Vì hệ số bất xứng $C_s = -0.676$ nằm giữa hai hệ số $C_s = -0.7$ và $C_s = -0.6$ trong bảng nên chúng ta có thể *xấp xỉ tuyến tính giữa hai số để có được giá trị k phù hợp.*

Vì $-0.676 = 0.76 \times (-0.7) + 0.24 \times (-0.6)$ nên chúng ta có

$$k(T, C_s = -0.676) = 0.76 \times k(T, C_s = -0.7) + 0.24 \times k(T, C_s = -0.6).$$

Chúng ta tính được giá trị k ứng với các thời gian lặp lại T và hệ số bất đối xứng $C_s = -0.676$. Từ đó chúng ta tính được giá trị R tương ứng các thời gian lặp lại T theo công thức $log\ R(T) = avg(\log R\ ) + k(T, C_s = -0.676) \times s$.

Bảng 3. *Giá trị $k(T, C_s = -0.676)$ và R*

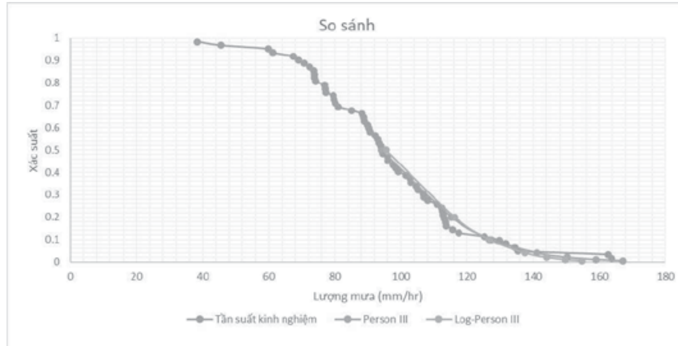
Thời gian lặp lại (năm)	$C_s = -0.7$	$C_s = -0.6$	$C_s = -0.676$	log R	R
2	0.116	0.099	-0.07144	1.980	95.5
5	0.857	0.857	0.81344	2.065	116.2
10	1.183	1.2	1.31892	2.103	126.7
25	1.488	1.528	1.8896	2.138	137.5
50	1.663	1.72	2.277	2.159	144.1
100	1.806	1.88	2.63772	2.175	149.8
200	1.926	2.016	2.97844	2.190	154.7

4.4. So sánh

Bảng 4. *Cường độ mưa tính theo hai phân phối Log-Person III và Person III*

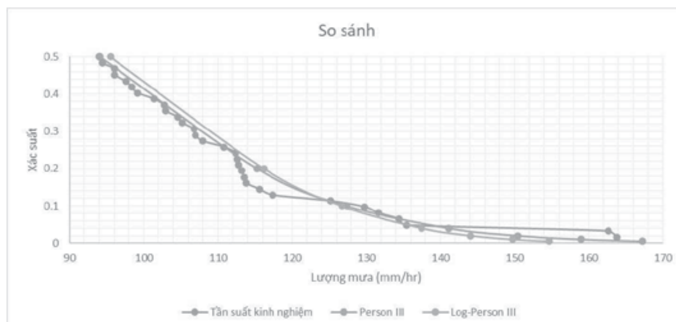
Thời gian lặp lại (năm)	Cường độ mưa R (mm/hr)	
	Log-Person III	Person III
2	95.5	94.1
5	116.2	115.3
10	126.7	127.4
25	137.5	141.1
50	144.1	150.4
100	149.8	159.0
200	154.7	167.2

Chúng ta vẽ hai phân phối cùng đường tần suất kinh nghiệm trên cùng biểu đồ.



Hình 2. Đường tần suất kinh nghiệm và số liệu tính qua hai phân phối.

Tuy nhiên, chúng ta có thể thấy rằng phần xác suất lớn hơn 0.5 không có trong đường Person III và Log-Person III. Lý do thời gian lặp lại ít nhất được tính là T=2 năm. Để nhìn rõ hơn tính phù hợp của hai phân phối với đường tần suất kinh nghiệm chúng ta chỉ quan sát phần biểu đồ ứng với xác suất từ 0.5 trở xuống.



Hình 3. Đường tần suất kinh nghiệm và số liệu tính qua hai phân phối với xác suất không quá 0.5.

Hình ảnh trên cho chúng ta thấy hai phân phối Person III và Log-Person III đều phản ánh tương đối sát đường tần suất kinh nghiệm. Với khoảng thời gian lặp lại T (chu lỳ lặp lại trận mưa tính toán) từ 10 năm trở xuống thì giá trị cường độ mưa tính theo hai phân phối tương đối sát nhau. Khi thời gian lặp lại càng dài thì giá trị của cường độ mưa khi tính theo phân phối Pearson III lớn hơn. Bên cạnh đó, phân phối Perrson III phản ánh sát hơn đường tần suất kinh nghiệm.

5. KẾT LUẬN

Bài báo trình bày cơ sở lý thuyết lựa chọn phân phối để tính toán xây dựng đường cong IDF cho khu vực Hà Nội. Thông qua một ví dụ minh hoạ việc sử dụng hai phân phối Person III và Log-Person III để tính toán cường độ mưa giới hạn trong thời đoạn 30 phút với số đo tại trạm Láng (Hà Nội), bài báo so sánh việc sử dụng hai phân phối nói trên để tính toán cường độ mưa giới hạn. Chúng ta thấy hai phân phối Person III và Log-Person III đều phản ánh tương đối sát đường tần suất kinh nghiệm. Tuy nhiên, phân phối Perrson III phản ánh sát hơn đường tần suất kinh nghiệm. Chúng ta có thể sử dụng quy trình tính toán này cho các thời đoạn khác. Từ đó có thể xây dựng đường cong IDF cũng như công thức tính cường độ mưa giới hạn cho khu vực Hà Nội. Điều này là cơ sở cho việc tính toán thiết kế hệ thống thoát nước mưa cho các khu đô thị ở Hà Nội.

PHỤ LỤC

Bảng 5. *Tính toán theo phân phối Person III*

STT	Năm	Lượng mưa (mm)	R	$(R - avg(R))^2$	$(R - avg(R))^3$	$(R - avg(R))^2 + (R - avg(R))^3$
1	2005	81.9	163.8	4621.771	314204.645	318826.416
2	2006	81.3	162.6	4460.050	297858.231	302318.282
3	1973	67.7	135.4	1566.862	62022.045	63588.907
4	1994	67.2	134.4	1488.695	57439.210	58927.905
5	2008	65.8	131.6	1280.466	45819.709	47100.176
6	1999	64.9	129.8	1154.886	39247.175	40402.060
7	1986	62.6	125.2	863.396	25369.698	26233.095
8	2010	58.7	117.4	465.852	10054.768	10520.620
9	2001	57.8	115.6	391.391	7743.127	8134.518
10	2015	56.9	113.8	323.410	5816.080	6139.490
11	1968	56.8	113.6	316.257	5624.184	5940.441
12	1997	56.6	113.2	302.190	5253.148	5555.338
13	1984	56.4	112.8	288.443	4898.801	5187.243
14	2002	56.3	112.6	281.689	4727.765	5009.454
15	2007	56.2	112.4	275.016	4560.757	4835.773
16	2003	55.4	110.8	224.508	3363.947	3588.455
17	1967	54.0	108	148.440	1808.538	1956.978
18	1964	53.5	107	125.073	1398.768	1523.841
19	1998	53.4	106.8	120.640	1325.058	1445.698
20	1977	52.6	105.2	88.052	826.246	914.298
21	1976	52.3	104.6	77.152	677.671	754.822
22	2021	51.5	103	51.604	370.704	422.308
23	1993	51.4	102.8	48.771	340.596	389.367
24	2009	50.7	101.4	31.177	174.078	205.255
25	2013	49.6	99.2	11.449	38.738	50.187
26	1991	49.2	98.4	6.675	17.246	23.921
27	1983	48.8	97.6	3.181	5.674	8.855
28	1961	48.0	96	0.034	0.006	0.040
29	1980	48.0	96	0.034	0.006	0.040
30	1971	47.2	94.4	2.006	-2.842	-0.835
31	1963	47.0	94	3.299	-5.993	-2.694
32	2018	46.9	93.8	4.066	-8.198	-4.132
33	2019	46.7	93.4	5.839	-14.109	-8.270
34	1982	46.5	93	7.932	-22.340	-14.408
35	1974	46.1	92.2	13.078	-47.296	-34.218
36	2016	45.3	90.6	27.211	-141.942	-114.731
37	2020	45.1	90.2	31.544	-177.163	-145.619
38	2017	45.0	90	33.830	-196.771	-162.941
39	1978	44.5	89	46.463	-316.712	-270.248
40	1975	44.4	88.8	49.230	-345.415	-296.186
41	1966	44.1	88.2	58.009	-441.823	-383.813
42	1985	42.5	85	116.994	-1265.457	-1148.463
43	1969	40.5	81	219.526	-3252.576	-3033.051
44	2000	40.0	80	250.158	-3956.602	-3706.444
45	1962	39.9	79.8	256.525	-4108.603	-3852.078
46	2011	39.8	79.6	262.971	-4264.448	-4001.477
47	2012	38.6	77.2	346.570	-6451.885	-6105.315
48	1972	38.5	77	354.057	-6662.069	-6308.013

49	1970	38.4	76.8	361.623	-6876.769	-6515.146
50	1965	37.1	74.2	467.268	-10100.659	-9633.391
51	1981	36.9	73.8	484.722	-10671.821	-10187.099
52	1987	36.9	73.8	484.722	-10671.821	-10187.099
53	2004	36.8	73.6	493.568	-10965.304	-10471.736
54	2014	36.2	72.4	548.327	-12839.852	-12291.525
55	1979	35.3	70.6	635.866	-16034.260	-15398.393
56	1989	34.5	69	719.119	-19284.177	-18565.058
57	1990	33.7	67.4	807.491	-22945.994	-22138.502
58	1995	30.6	61.2	1198.295	-41480.641	-40282.346
59	1988	29.9	59.8	1297.181	-46719.767	-45422.586
60	1996	22.8	45.6	2521.686	-126629.985	-124108.299
61	1992	19.2	38.4	3296.642	-189281.308	-185984.665

Bảng 6. *Tính toán theo phân phối Log-Person III*

STT	Năm	Lượng mưa (mm)	R	log R	(log R - avg(log R))^2	(log R - avg(log R))^3	(log R - avg(log R))^2 + (log R - avg(log R))^3
1	2005	81.9	163.8	2.214	0.061	0.015	0.076
2	2006	81.3	162.6	2.211	0.059	0.014	0.074
3	1973	67.7	135.4	2.132	0.027	0.004	0.031
4	1994	67.2	134.4	2.128	0.026	0.004	0.030
5	2008	65.8	131.6	2.119	0.023	0.003	0.027
6	1999	64.9	129.8	2.113	0.021	0.003	0.024
7	1986	62.6	125.2	2.098	0.017	0.002	0.019
8	2010	58.7	117.4	2.070	0.010	0.001	0.012
9	2001	57.8	115.6	2.063	0.009	0.001	0.010
10	2015	56.9	113.8	2.056	0.008	0.001	0.009
11	1968	56.8	113.6	2.055	0.008	0.001	0.008
12	1997	56.6	113.2	2.054	0.007	0.001	0.008
13	1984	56.4	112.8	2.052	0.007	0.001	0.008
14	2002	56.3	112.6	2.052	0.007	0.001	0.008
15	2007	56.2	112.4	2.051	0.007	0.001	0.008
16	2003	55.4	110.8	2.045	0.006	0.000	0.006
17	1967	54.0	108	2.033	0.004	0.000	0.005
18	1964	53.5	107	2.029	0.004	0.000	0.004
19	1998	53.4	106.8	2.029	0.004	0.000	0.004
20	1977	52.6	105.2	2.022	0.003	0.000	0.003
21	1976	52.3	104.6	2.020	0.003	0.000	0.003
22	2021	51.5	103	2.013	0.002	0.000	0.002
23	1993	51.4	102.8	2.012	0.002	0.000	0.002
24	2009	50.7	101.4	2.006	0.001	0.000	0.002
25	2013	49.6	99.2	1.997	0.001	0.000	0.001
26	1991	49.2	98.4	1.993	0.001	0.000	0.001
27	1983	48.8	97.6	1.989	0.000	0.000	0.000

28	1961	48.0	96	1.982	0.000	0.000	0.000
29	1980	48.0	96	1.982	0.000	0.000	0.000
30	1971	47.2	94.4	1.975	0.000	0.000	0.000
31	1963	47.0	94	1.973	0.000	0.000	0.000
32	2018	46.9	93.8	1.972	0.000	0.000	0.000
33	2019	46.7	93.4	1.970	0.000	0.000	0.000
34	1982	46.5	93	1.968	0.000	0.000	0.000
35	1974	46.1	92.2	1.965	0.000	0.000	0.000
36	2016	45.3	90.6	1.957	0.000	0.000	0.000
37	2020	45.1	90.2	1.955	0.000	0.000	0.000
38	2017	45.0	90	1.954	0.000	0.000	0.000
39	1978	44.5	89	1.949	0.000	0.000	0.000
40	1975	44.4	88.8	1.948	0.000	0.000	0.000
41	1966	44.1	88.2	1.945	0.000	0.000	0.000
42	1985	42.5	85	1.929	0.001	0.000	0.001
43	1969	40.5	81	1.908	0.003	0.000	0.003
44	2000	40.0	80	1.903	0.004	0.000	0.004
45	1962	39.9	79.8	1.902	0.004	0.000	0.004
46	2011	39.8	79.6	1.901	0.004	0.000	0.004
47	2012	38.6	77.2	1.888	0.006	-0.001	0.006
48	1972	38.5	77	1.886	0.007	-0.001	0.006
49	1970	38.4	76.8	1.885	0.007	-0.001	0.006
50	1965	37.1	74.2	1.870	0.009	-0.001	0.009
51	1981	36.9	73.8	1.868	0.010	-0.001	0.009
52	1987	36.9	73.8	1.868	0.010	-0.001	0.009
53	2004	36.8	73.6	1.867	0.010	-0.001	0.009
54	2014	36.2	72.4	1.860	0.012	-0.001	0.010
55	1979	35.3	70.6	1.849	0.014	-0.002	0.012
56	1989	34.5	69	1.839	0.017	-0.002	0.014
57	1990	33.7	67.4	1.829	0.019	-0.003	0.017
58	1995	30.6	61.2	1.787	0.033	-0.006	0.027
59	1988	29.9	59.8	1.777	0.036	-0.007	0.029
60	1996	22.8	45.6	1.659	0.095	-0.029	0.066
61	1992	19.2	38.4	1.584	0.147	-0.056	0.091

TÀI LIỆU THAM KHẢO

1. Nguyễn Hữu Khải (2008), *Phân tích thống kê trong thuỷ văn*, NXB Đại học Quốc gia Hà Nội, Hà Nội.

2. Phan Văn Tân (2005), *Phương pháp thống kê trong khí hậu*, NXB Đại học Quốc gia Hà Nội, Hà Nội.

3. Viện khoa học Khí tượng thuỷ văn và Biến đổi khí hậu (2019), *Báo cáo Tổng kết thực hiện Dự án Xây dựng hệ thống đường cong cường độ mưa - thời đoạn - tần suất (IDF) để tăng cường công tác quản lý nhà nước về thông tin khí tượng thuỷ văn phục vụ phát triển kinh tế xã hội*.

4. International Hydrological Programme IHP-VII, Technical Documents in Hydrology, No. 2, 2008, "*Asian Pacific FRIEND, Rainfall Intensity Duration Frequency (IDF), Analysis for the Asia Pacific Region*", Edited by: Trevor M. Daniell and Guillermo Q. Tabios III, Reported by Regional Steering Committee for Southeast Asia and the Pacific, UNESCO Office, Jakarta 2008.

5. Koutsoyiannis D., Kozonis D. and Manetas A. (1998), *A Mathematical Framework for Studying Rainfall Intensity-Duration-Frequency Relationships*. J. Hydrology, 206, 118-135.

6. Rajib Maity (2018), *Statistical Methods in Hydrology and Hydroclimatology*, Springer Transactions in Civil and Environmental Engineering, Springer, Singapore.