

Evaluación de métodos de construcción de curvas IDF a partir de distribuciones de probabilidad y parámetros de ajuste

IDF construction methods' evaluation, from probability distributions and adjustment's parameters

Fecha de Recepción: 19 de septiembre de 2013
Fecha de Aprobación: 07 de noviembre de 2013

Pedro Mauricio Acosta Castellanos*
Lina Xiomara Sierra Aponte**

Resumen

Las curvas intensidad, duración y frecuencia (IDF) son de gran importancia para el diseño de obras hidráulicas; por tal motivo, conocer las metodologías para su construcción es importante. Este artículo presenta una evaluación de las metodologías para la construcción de curvas IDF, aplicando distribuciones de probabilidad como EVI, Pearson Tipo III, log-Pearson Tipo III y GEV. La aplicación de la investigación fue en el departamento de Boyacá, Colombia. La construcción de las curvas IDF se hizo a partir de datos pluviográficos.

De la evaluación realizada se encontró la función de distribución, con la cual se evidencia un mejor comportamiento para la zona de estudio; es la función de Valor Extremo Tipo I. Con resultados similares para periodos de retorno mayores,

Abstract

The intensity, duration and frequency's curves (IDF) are very important for the hydraulic structures' design. For that reason is important to know these curves construction methodologies. This paper presents some methodologies evaluation, for the IDF curves construction, by applying probability distributions as EVI, Pearson type III, log-Pearson Type III and GEV. The research application place was the department of Boyacá, Colombia. The IDF construction's curves was carried out, by using some pluviograph data. From the evaluation carried out, it was found that the distribution function, with the best performance for this area's study, is the Extreme Value Type I. With similar results for higher return periods, is the Generalized Extreme Value function. Also through the analysis it could rule

* M. Sc. Universidad Santo Tomás, Sede Tunja (Colombia), pedro.acosta@usantoto.edu.co

** Universidad Santo Tomás, Sede Tunja (Colombia).

se encuentra la función de Valor Extremo Generalizado. Mediante el análisis se pudo descartar la aplicación de las funciones Pearson y log-Pearson Tipo III.

Palabras clave: Curvas IDF, Obras hidráulicas, Distribuciones de probabilidad, EVI, Pearson Tipo III, log-Pearson Tipo III, GEV, Función de distribución.

out the Pearson functions and log-Pearson Type III's applications.

Keywords: IDF Curves, Hydraulic Structures, Probability Distributions, Evi, Pearson Type Iii, Log-Pearson Type Iii and GEV, Distribution Function.

I. INTRODUCCIÓN

Las curvas intensidad, duración y frecuencia (IDF) son un eslabón fundamental para el diseño y la construcción de obras hidráulicas, como drenajes pluviales, presas, embalses y demás, que buscan controlar inundaciones y aprovechar los recursos hídricos para generar energía eléctrica; además, son de gran valor en el estudio de la erosión del suelo, entre otros usos; es por esto que el conocimiento de cómo se construyen las IDF es fundamental.

Esta investigación dio como resultado una aproximación teórica y práctica de los métodos más utilizados para construir las curvas IDF, determinando cuáles son las diferencias, ventajas y desventajas de tales métodos y su posible aplicación en estaciones pluviométricas del departamento de Boyacá, Colombia, exactamente en las estaciones hidroclimatológicas de los municipios de Aquitania, Pesca, Samacá, Tota, Paipa, Tunja y Tuta. Para el análisis comparativo de metodologías para la construcción de curvas IDF se hizo necesario la recopilación histórica de eventos de precipitación.

Se desarrolló un análisis estadístico continuo, que consiste en parametrizar un conjunto de datos (precipitaciones), con el fin de extrapolar y conocer el comportamiento de los eventos. Inicialmente, la selección de las precipitaciones se ejecutó con la revisión de los registros pluviométricos máximos anuales obtenidos de las estaciones hidroclimatológicas correspondientes al área de estudio. Una vez procesada la información, se realizó una aproximación analítica y estadística para determinar cuáles son las posibles ventajas y desventajas en la aplicación de distribuciones de probabilidad, parámetros de ajuste y en la construcción las curvas.

Finalmente, es evidente que las curvas son de gran importancia para la predicción de lluvias venideras, así como para investigar sobre fenómenos que se salen de la normalidad del clima y, en

general, de la hidrología de una zona específica. De tal manera que uno de los objetivos a los que apunta la investigación es el conocimiento de las alternativas que existen para la determinación de curvas IDF.

II. METODOLOGÍA

A. Selección de información

Se tomó como criterio el tipo, la cantidad y la periodicidad de los datos de precipitación registrados con pluviógrafo; para tal fin se obtuvieron los registros históricos superiores a 10 años, y se recopilaron alrededor de 561 pluviogramas correspondientes a las tormentas registradas en 7 estaciones de la zona centro del departamento de Boyacá, Colombia, donde se encuentran 23 municipios y la capital del departamento. Para cada una de las estaciones se recopilaron los mayores eventos de precipitación; a partir de esta información se procedió a analizar los pluviogramas de cada estación, a fin de extraer la información numérica, con el propósito de obtener las precipitaciones máximas para cada duración y convertirlas en intensidades de precipitación.

B. Funciones de distribución

Los datos sistematizados deben ser analizados a la luz de una frecuencia de parámetros hidrológicos, por medio de distribuciones de probabilidad, con el fin de relacionar la magnitud y la frecuencia de eventos extremos. Cada función de distribución tiene diferentes parámetros, estadísticos y matemáticos. En la mayoría de funciones es común la media aritmética, pues es necesaria para determinar el orden de las intensidades. De igual manera, se halla la dispersión o variación de los datos de la muestra, utilizando la desviación estándar de cada muestra. Lo anterior es el primer tratamiento que se hace necesario para aplicar las funciones a los datos.

Una *distribución de probabilidad* es una función que representa la probabilidad de ocurrencia en

una serie pluviográfica. En estadística existe un sinnúmero de funciones de probabilidad. Se eligieron las funciones que mejor se adaptaron a las series de precipitaciones máximas diarias, es decir, se procedió a estimar los parámetros con los datos que se disponían, y se verificó que ellos se pudieran ajustar a una distribución de manera satisfactoria. Las funciones de distribución que se utilizaron en este caso fueron Valor extremo tipo I (EVI), Pearson Tipo III, Log Pearson Tipo III y Valor extremo generalizado (GEV).

1) *Valor Extremo Tipo I (EVI)*: Se utiliza para el análisis de frecuencias de datos extremos, considerando que cada máximo anual es el valor extremo observado en una muestra de un año, y tiene en cuenta dos parámetros, de ubicación (u) y escala (α), y el coeficiente de asimetría es fijo y vale 1,139. Esta distribución solo aplicará a variables aleatorias positivas, es decir, $X > 0$. De igual manera, está definido que los valores máximos anuales se aproximan a una tendencia de distribución específica, cuando el número de muestras es alto [1].

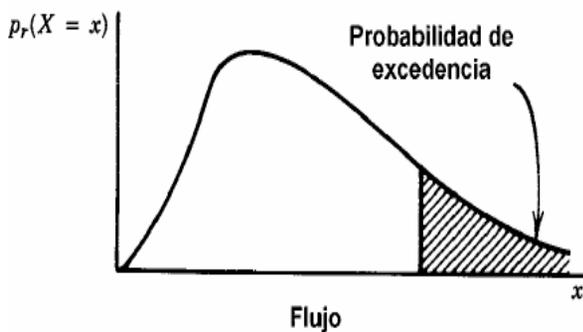


FIG. 1. Tendencia gráfica de la distribución de EVI

Fuente: Torres Abello, Andrés Eduardo. Apuntes de clase sobre hidrología urbana. Bogotá: Pontificia Universidad Javeriana, 2004, p. 41.

2) *Pearson Tipo III*: Normalmente, la distribución Pearson Tipo III es aplicada para describir la distribución de probabilidad de picos decrecientes de máximos anuales. Cuando la información es muy asimétrica positivamente se utiliza una

transformación log para reducir la asimetría [2]. También es llamada distribución gamma de tres parámetros, puesto que introduce tres parámetros, uno más que EVI: parámetro de localización ϵ , parámetro de ubicación β y el parámetro de escala α . Esta es una distribución flexible, ya que puede asumir diferentes formas a medida que ϵ , β y α varían [3].

Debido a los tres parámetros que tiene la distribución, es necesario utilizar el coeficiente de asimetría, que para el caso de Pearson, al ser una distribución simétrica, la media de la distribución es igual a la moda.

La flexibilidad de esta distribución está dada, claramente, por los tres parámetros estadísticos. Cuando el coeficiente de asimetría es cero, pasará a ser una distribución normal.

La ecuación (1) representa la función de probabilidad Pearson tipo III; contiene los parámetros propios de la función y los parámetros estadísticos mencionadas en los párrafos anteriores.

$$f(x) = \frac{1}{\alpha \Gamma(\beta)} * \left\{ \frac{x-\epsilon}{\alpha} \right\}^{\beta-1} * e^{-\frac{x-\epsilon}{\alpha}} \quad (1)$$

3) *Log-Pearson Tipo III*: Se utiliza principalmente para análisis de probabilidad de eventos extremos; consiste, principalmente, en transformar los valores extremos en logaritmo en base 10, con el fin de reducir la asimetría de los datos; se calcula con la misma función de Pearson tipo III, utilizando la ecuación (2), $F(x) = \log x$. Los parámetros de esta distribución se calculan de la misma manera que en Pearson tipo III, con la diferencia de que los parámetros estadísticos media, desviación estándar y coeficiente de asimetría se hallan con los logaritmos en base 10 de los datos originales.

En el caso de la hidrología, el uso de Log-Pearson tipo III se limita para el tratamiento de

datos de picos de inundación; esta distribución es la distribución estándar para el análisis de la crecida máxima en los Estados Unidos [4]. Esta distribución infiere que si en la muestra de máximos anuales la frecuencia alcanza un valor de límite inferior, como, por ejemplo, cero, en tal caso será apropiado utilizar otra distribución [5].

4) *Distribución de Valor Extremo Generalizada (GEV)*: Las tres distribuciones anteriores son casos especiales de una distribución única que podría llamarse distribución de valor extremo generalizado; es decir, que combina los tres tipos de distribuciones más comunes en una sola forma, permitiendo un rango continuo de formas posibles [6].

La expresión (2) representa la función de probabilidad para (GEV); ella relaciona, al igual que las funciones antes mencionadas, los parámetros propios de la función, añadiendo uno propio (k) y los de ubicación y escala.

$$F(x) = \exp \left[- \left(1 - k \frac{x-u}{\alpha} \right)^{\frac{1}{k}} \right] \quad (2)$$

Se ha demostrado que la distribución de valores extremos seleccionados de conjuntos de muestras de cualquier distribución se aproxima a una de las tres distribuciones de valores extremos, llamadas Tipo I, Tipo II y Tipo III, indicando que el número de datos seleccionados es mayor. La teoría de los

valores extremos proporciona técnicas simples para estimar las probabilidades de ocurrencia de niveles extremos originados por datos históricos. La mayor parte de métodos estadísticos estudian principalmente lo que ocurre en el centro de una distribución estadística, y no prestan particular atención a las colas de la distribución, es decir, a los valores extremos, ya sea en el extremo alto o bajo [7].

C. Prueba de bondad de ajuste

Una prueba de bondad de ajuste determina si la muestra o los datos se ajustan a la distribución de probabilidad que se está utilizando; puede utilizarse tanto para hipótesis simple o compuesta. Este paso es fundamental para determinar si es viable utilizar una distribución determinada en una serie de datos; en este caso, para la muestra de máximos anuales se utilizó el método de Kolmogorov-Smirnov, que permite calcular la relación que existe entre una distribución de una muestra de datos y una distribución teórica [8].

La prueba de bondad de ajuste requiere que el valor estadístico de prueba de la muestra (D_n) sea menor que el valor tabulado ($D_n \alpha$) para un nivel de probabilidad requerido, a fin de que la prueba sea aceptada. En la Tabla 1 se muestran los resultados de la prueba para cada una de las estaciones, con la función de distribución EVI.

TABLA 1
APLICACIÓN DE LA PRUEBA DE BONDAD DE AJUSTE

Estación	Nivel de significancia	Dn. Máx.	Dn. α
Tunguavita - Paipa	0,05	0,14	0,2895
Potrerito - Aquitania	0,05	0,13	0,309
Villa Carmen - Samacá	0,05	0,14	0,309
UPTC – Tunja	0,05	0,16	0,287
Tota	0,05	0,15	0,349
Pesca	0,05	0,14	0,361
Azulejos – Tuta	0,05	0,18	0,301

D. Estimación de parámetros de ajuste a una distribución de probabilidad

Como se mostró en los literales anteriores, cada distribución, independiente del tipo que sea, tiene diferentes parámetros; cada uno de ellos con características propias y dependientes de la distribución a la que se relacionen.

Existen diferentes métodos o estimadores para el cálculo de los parámetros de una función de distribución; es necesario que los estimadores cumplan con las propiedades de ser no sesgados, regulares, eficientes, suficientes y consistentes [9]. Cada distribución de probabilidad tiene parámetros, que deben ser ajustados por métodos que cumplan con las propiedades anteriores. De la misma manera, debe buscarse que cada método pueda ajustar los parámetros de la distribución de probabilidad que se vaya a utilizar.

Entre los métodos de estimación de parámetros pueden mencionarse, en orden ascendente de

eficiencia de manejo: el método de los momentos ordinarios (OM), el método de los momentos ponderados por probabilidad (PWM) y el método de máxima verosimilitud (ML) [10].

1) *Método de los momentos ordinarios (OM)*: Este método considera que una estimación adecuada de los parámetros de una función de probabilidad es aquella para la cual los momentos de la función de densidad de probabilidad alrededor del origen son iguales a los momentos correspondientes a la información de la muestra [2]. Los estimadores calculados mediante el método de los momentos son asintóticamente eficientes, particularmente en distribuciones asimétricas, que son las más usadas en hidrología; esta eficiencia hace que este método se use en muchos casos como una primera aproximación en la estimación de los parámetros [11].

En la Tabla 2 se muestra el parámetro α calculado a partir de momentos ordinarios para la Estación Tanguavita, de la ciudad de Paipa.

TABLA 2
PARÁMETRO α , ESTACIÓN TUNGUAVITA, ESTIMADO POR MOMENTOS ORDINARIOS PARA EVI

Parámetro	Alfa								
	10	20	30	60	120	180	220	280	320
Duración (min.)									
Valor α	35,3	17,8	14,5	9,54	9,42	9,80	9,97	10,22	10,33

2) *Método de los momentos ponderados por probabilidad (PWM) o L-Moments*: Este método es similar al de los momentos ordinarios, y se basa en que una función de distribución puede ser caracterizada por momentos ponderados por la probabilidad; por lo tanto, basta con calcular los momentos ponderados por la probabilidad a partir de los datos muestrales.

El PWM tiene ventaja teórica sobre el método de los momentos convencional, pues con él se caracteriza una amplia gama de funciones de distribución y, además, es un método más robusto ante la presencia de puntos extraordinarios en la muestra “Outliers”, cuando se calculan los parámetros a partir de los datos muestrales [12].

TABLA 3
PARÁMETRO ξ , PARA GEV, ESTACIÓN PESCA

Parámetro	ξ								
Duración (min.)	10	20	30	60	120	180	220	280	320
Valor ξ	17,7	983	75,8	3298	2,3	0,6	0,5	0,38	0,32

3) *Método de máxima verosimilitud (ML)*: Este método busca que el valor de un parámetro en una distribución de probabilidad sea el valor que maximizará la verosimilitud o probabilidad conjunta de ocurrencia de una muestra observada; por tal motivo, representa una mayor precisión desde el punto de vista estadístico, pero puede presentar imprecisiones para muestras de menor tamaño, comparado con los demás métodos; adicionalmente, es de mayor laboriosidad desde el punto de vista de análisis de datos [13].

E. Periodo de retorno

El periodo de retorno de un evento con una magnitud dada puede definirse como el intervalo de recurrencia promedio entre eventos que igualan o exceden una magnitud especificada. Puede ocurrir que para un valor umbral de intensidad media haya más de un valor que lo supera en un mismo año; con el modelo de máximos anuales, solo uno de ellos, el mayor, conformará la serie de observaciones, por consiguiente, quedarán por fuera los demás, incluso cuando pudieran ser mayores que los máximos de otros años [14]. Lo anterior implica que el período de retorno asociado deba ser corregido, para lo cual se utilizó la ecuación (3), denominada expresión de Langbein.

$$TAM = \frac{1}{1 - e^{-\frac{1}{T}}} \quad (3)$$

F. Variable reducida

Una variable reducida puede definirse como la relación entre la probabilidad y el periodo de retorno; se utiliza en conjunto con los parámetros de ajuste de la distribución de probabilidad y el

periodo de retorno, a fin de determinar las curvas intensidad, duración y frecuencia (IDF).

De esta manera, si se tiene que la ecuación (3) representa el periodo de retorno, y la ecuación (4), la distribución de probabilidad de EVI, al fusionar estas dos matemáticamente se obtiene la variable reducida, representada por la ecuación (5), donde a partir de los parámetros de escala y ubicación y el periodo de retorno se puede hallar el valor intensidad, para una duración de evento determinada.

$$F(x) = e^{-e^{-\alpha(x-u)}} \quad (4)$$

$$x(T) = U + \alpha \left\{ -\ln \left[-\ln \left(1 - \frac{1}{T} \right) \right] \right\} \quad (5)$$

III. RESULTADOS

Se construyeron las curvas IDF para siete estaciones pluviográficas del departamento de Boyacá, teniendo en cuenta los tres tipos de distribuciones mencionadas en el numeral II. De la misma manera, se ajustaron los parámetros de cada distribución, por momentos ordinarios, máxima verosimilitud y momentos ponderados por probabilidad. De los resultados obtenidos en la aplicación de estas distribuciones de probabilidad, y partiendo de las intensidades máximas de precipitación para cada una de las estaciones que se estudiaron, se generaron 21 graficas: 7 correspondientes a la aplicación de la distribución de Valor Extremo Tipo I EVI-Gumbel, mediante el ajuste por el método de los Momentos Ordinarios (OM); 7 para la aplicación de la distribución de Valor Extremo Tipo I EVI-Gumbel, mediante el ajuste por el método de

Máxima Verosimilitud (ML), y, finalmente, 7 para la distribución de Valor Extremo Generalizado GEV, mediante el ajuste por el método de los Momentos Ponderados por Probabilidad (PWM). Cada grafica corresponde a una curva intensidad, duración y frecuencia.

Se realizó un proceso comparativo, analizando las ventajas, desventajas, diferencias y aplicabilidad de cada distribución a la zona de estudio. En el análisis de resultados se plasman las principales conclusiones del proceso comparativo.

IV. ANÁLISIS DE RESULTADOS

En la obtención de los valores de intensidades máximas para distintos periodos de retorno se presentó una variabilidad entre distribuciones de probabilidad, siendo la función de log-Pearson la que presenta valores de intensidad más variables, sobreestimándolos. La función Pearson Tipo III presenta variaciones menores y cercanas a las de EVI.

En el análisis de eventos extremos de intensidades, los parámetros de cada una de las funciones de distribución fueron estimados de acuerdo con los métodos de momentos ordinarios, máxima verosimilitud y momentos ponderados por probabilidad. A pesar de no obtenerse diferencias significativas en los parámetros estimados, se pudo observar que el ajuste más adecuado comparativamente fue obtenido por el método de momentos ordinarios.

La función de distribución de Pearson Tipo III presenta la calidad más baja de ajustes, pudiéndose concluir que esta función no es la recomendable para estimar el comportamiento de las intensidades medias de esta región. Adicionalmente, la variable reducida es de larga laboriosidad matemática.

Para ajustar distribuciones de tres parámetros se necesita estimar el coeficiente de asimetría de la

distribución, para lo cual es necesario disponer de una serie con longitud de registros larga, preferiblemente mayor de 50 años.

Se puede concluir que las distribuciones cuyos extremos disminuyen exponencialmente, como la normal, llevan a las distribuciones del tipo I. Los extremos que disminuyen, como un polinomio, conducen a las tipo II. Las distribuciones cuyos extremos tienden a ser finitos, como el beta, conducen a distribuciones tipo III.

La estación Tinguavita, del municipio de Paipa, es la que permite evidenciar un mejor comportamiento de las precipitaciones máximas anuales; así mismo, el comportamiento gráfico de las curvas IDF es adecuado al realizarlas por EVI, GEV y Pearson tipo III. Esto se debe a que la estación tiene continuidad en el registro histórico de precipitaciones y no presenta vacíos o datos faltantes.

La función de Gumbel, ajustada por el método de los Momentos Ordinarios y Máxima Verosimilitud, presenta mayor flexibilidad para su aplicación; en general, es la más apropiada para ajustarse a precipitaciones máximas, en distintos periodos de retorno para la región de estudio; además, es recomendable su uso por su rapidez y facilidad de cálculo.

En la estimación de los parámetros por el método de Máxima Verosimilitud se observó que, específicamente en EVI, para determinar el parámetro α era necesario tener un parámetro α de referencia. Así, para utilizar este método de manera adecuada puede ser necesario calcular previamente cada parámetro con otro método, a fin de que sirva de referencia.

El método de Valor Extremo Generalizado (GEV) proporciona una técnica sencilla para estimar probabilidades de futuros niveles extremos en un proceso de datos históricos, lo cual es fundamental si se pretende efectuar un análisis para obras hidráulicas de gran impacto, como

presas e hidroeléctricas, puesto que la estimación de intensidades para periodos de retorno altos es más adecuada para el modelo de máximos anuales.

En relación con la construcción de las curvas IDF para las distintas estaciones, se puede concluir que, en general, no se presentaron grandes problemas, salvo con las estaciones de Pesca y Tota, en las cuales fue necesario elaborar una revisión detallada de la acumulación de las duraciones de precipitaciones máximas para cada año, esto con el objeto de obtener una gráfica lo más cercana al modelo técnico. Debido al vacío de datos en el registro histórico de precipitaciones.

REFERENCIAS

- [1] R. Muñoz, A. Ritter, *Hidrología Agroforestal*, Canarias: Mundi Prensa.
- [2] V.T. Chow, “Estadística hidrológica”, *Hidrología aplicada*, Bogotá: McGraw Hill, 2000.
- [3] D. R. Maidment, “Frequency Analysis of Extreme Events”, *Handbook of Hydrology*, Texas: McGraw Hill, 1992.
- [4] P.J. Reddy, *Stochastic Hydrology*, New Delhi: Laxmi Publications, 1997.
- [5] V. W. Griffis, J. R. Stedinger, T. A. Cohn. “LP3 quantile estimators with regional skew information and low outlier adjustments”. *Water Resources Research*, vol. 40, W07503.
- [6] R.W. Katz, M.C. Parlange. “Statics of Extreme in Hydrology”. *Advances in Water Resources*, vol. 25, pp. 1124-1134, 2002.
- [7] O.S. Muhaisen, *Simulación basada de la teoría de valor extremo y cópula para estimación de caudales en sistema de sanitaria*, Tesis Doctoral. Universidad de Granada, 2006.
- [8] R. Simard, P. L’Ecuyer. “Computing the Two-Sided Kolmogorov-Smirnov Distribution”. *Journal of Statistical Software*, vol. 39, Núm. 11, 2011.
- [9] W. Yepes. *Estudio de las propiedades de invarianza de las precipitaciones máximas puntuales en el departamento de Antioquia*. Medellín: Universidad Nacional de Colombia, 2001.
- [10] E.J. Gumbell, *Statistics of extremes*. New York: Columbia University, Dover edition, 2004.
- [11] D. Koutsoyiannis, “Statistics of extremes and estimation of extreme rainfall: I Theoretical investigation”. *Hydrological Sciences-Journal-des Sciences Hydrologiques*, 2004, pp. 575-590.
- [12] E.G. Pulgarín. *Fórmulas regionales para la estimación de curvas intensidad-frecuencia-duración basadas en las propiedades de escala de la lluvia (Región Andina Colombiana)*. Medellín: Universidad Nacional de Colombia. 2009.
- [13] S. Chavoshi, W.N. Azmin. “Development of L-Moment based models for extreme flood events”. *Malaysian Journal of Mathematical Sciences*, vol. 3, pp. 281-296, 2009.
- [14] D.L. Fitzgerald. “Estimation of Point Rainfall Frequencies”. *Technical Note 61*. Irish Meteorological Service, pp. 42, 2007.