

DOI: 10.24850/j-tyca-15-01-09

Artículos

## **Curvas de intensidad-duración-frecuencia para la ciudad de Santa Clara, Cuba**

### **Intensity-duration-frequency Curves for Santa Clara City, Cuba**

Carlos Castillo-García<sup>1</sup>, ORCID: <https://www.orcid.org/0000-0002-6430-2775>

Ismabel Domínguez-Hurtado<sup>2</sup>, ORCID: <https://www.orcid.org/0000-0002-7841-8031>

Yoel Martínez-González<sup>3</sup>, ORCID: <https://www.orcid.org/0000-0002-8023-7897>

Diego Abreu-Franco<sup>4</sup>, ORCID: <https://www.orcid.org/0000-0001-6161-2922>

<sup>1</sup>Universidad Central Marta Abreu de Las Villas, Santa Clara, Cuba,  
[ccgarcia@uclv.cu](mailto:ccgarcia@uclv.cu)

<sup>2</sup>Centro Meteorológico Provincial de Villa Clara, Santa Clara, Cuba,  
[ismabel.dominguez@vcl.insmet.cu](mailto:ismabel.dominguez@vcl.insmet.cu)

<sup>3</sup>Instituto Superior de Tecnologías y Ciencias Aplicadas, La Habana, Cuba,  
[ymg@instec.cu](mailto:ymg@instec.cu)

<sup>4</sup>Empresa de Investigaciones y Proyectos Hidráulicos de Villa Clara, Cuba,  
iph.tecproy14@vc.giat.cu

Autor para correspondencia: Carlos Castillo-García, ccgarcia@uclv.cu

## Resumen

Las curvas de intensidad-duración-frecuencia (IDF) son una representación de fenómenos hidrometeorológicos extremos de la lluvia para su uso en proyectos hidrológicos. En el presente artículo se realizó un análisis de 243 eventos lluviosos convectivos de más de 25 mm ocurridos en la estación meteorológica Yabú de la provincia Villa Clara, Cuba, en el periodo comprendido desde 1990 hasta 2019, con el objetivo de elaborar las curvas IDF de dicha estación. Se elaboró una serie de máximos anuales para las duraciones comprendidas entre los 5 y 4 320 minutos, la cual se sometió a un proceso de imputación de datos faltantes usando el algoritmo de imputación múltiple por regresión lineal; se encontraron valores anómalos y se destacó su tratamiento. Las series resultantes se testearon en pruebas no paramétricas para comprobar su independencia, aleatoriedad y estacionalidad, con lo cual se procedió a ajustarlas a la distribución probabilística de valores extremos Gumbel y posteriormente a una ecuación paramétrica del modelo de Montana. Los resultados obtenidos demostraron que existe un punto donde el ajuste del modelo de Montana empieza a obtener resultados discordantes con la serie ajustada a la distribución Gumbel, por lo cual se proponen dos familias de Curvas IDF para duraciones  $\leq 360$  min y otra para duraciones

> 360 min, con las que se obtienen coeficientes de correlación superiores a los 0.99.

**Palabras clave:** eventos extremos, intensidad de la lluvia, periodo de retorno, precipitación.

## Abstract

The intensity-duration-frequency (IDF) curves are a representation of extreme hydrometeorological phenomena of rainfall to be used in hydrological projects. In this article, an analysis of 243 convective rainy events of more than 25 mm that occurred at the Yabú Meteorological Station in Cuba, Villa Clara province, in the interim period from 1990 to 2019 was carried out with the objective of elaborating the IDF curves of the station aforementioned. A series of annual maximums was elaborated for the durations between 5 and 4 320 minutes, which was subjected to a missing data imputation process using the multiple imputation algorithm by linear regression, anomalous values were found, and their treatment was highlighted. The resulting series were tested in non-parametric tests to verify their independence, randomness and seasonality, with which they were adjusted to the Gumbel probabilistic distribution of extreme values and subsequently to a parametric equation of the Montana model. The results obtained showed that there is a point where the adjustment of the Montana model begins to obtain discordant results with the series adjusted to the Gumbel distribution, for which two families of IDF Curves are proposed: For durations  $\leq 360$  min and another for durations  $> 360$  min, with which correlation coefficients greater than 0.99 are obtained.

**Keywords:** Extreme events, rainfall intensity, return period, precipitation.

Recibido: 11/03/2022

Aceptado: 05/07/2022

Publicado online: 19/07/2022

## Introducción

La tormenta de diseño usualmente necesita para su confección modelos de intensidad-duración frecuencia (IDF) o lámina-duración-frecuencia (DDF, por sus siglas en inglés), los cuales contienen la relación de la probabilidad de ocurrencia de las variables lámina e intensidad con su duración. Ambos son usados como la entrada primaria a modelos de lluvia escorrentía para estimar la magnitud de la avenida de diseño, en particular en cuencas con ausencia de mediciones de gastos. Singh (2017) sostiene que las derivaciones de estas relaciones requieren de un manejo de alta calidad de los datos, ajustándolos a una distribución de valores extremos donde después pueden ser usados para extrapolarlos a una probabilidad de excedencia de interés. Los parámetros duración de la tormenta e intensidad obtenidos de las curvas IDF tienen una gran significancia en el campo de la hidrología, y son elementos básicos para el estudio de grandes avenidas y el desarrollo de infraestructuras urbanas (Yong, Ng, Huang, & Ang, 2021).

Con el desarrollo de las metodologías computacionales y estudios elaborados a partir de experiencias en otras ramas de la estadística ingenieril se identifican dos clasificaciones de curvas IDF; la primera es de acuerdo con la variabilidad en las tendencias de las series de datos empleadas. Autores como Agilan y Umamahesh (2017b); Gregersen, Madsen, Rosbjerg y Arnbjerg-Nielsen (2017), y Soumya, Anjitha, Mohan, Adarsh y Gopakumar (2020) han propuesto la no estacionalidad de las series de datos como una condición de interés para un estudio de lluvias de corta duración; el propio Agilan ha realizado varias contribuciones a esta temática exponiendo que los valores de la serie de datos pueden incrementarse o decrecer de acuerdo con una función lineal, o tendencia lineal, cuyo valor de pendiente puede ser agregado a la función de distribución de valores extremos empleada en su parámetro de posición (Agilan & Umamahesh, 2017a); esta tendencia es contrastada con otra parte de la comunidad científica que aún no se encuentra del todo convencida. Cabe mencionar que los propios Agilan y Umamahesh (2017c) asumen una posición cautelosa sobre los resultados en comparaciones con periodos de ocurrencia de hasta 10 años; sin embargo, otros autores, como Ganguli y Coulibaly (2017), y Yilmaz y Perera (2014), a pesar de comprobar tendencias significativas en la series de máximos anuales (Noor, Ismail, Chung, Shahid, & Sung, 2018) demuestran que los modelos no estacionarios con funciones de valores generalizados extremos (GEV, por sus siglas en inglés) no advierten aún ventajas claras sobre modelos estacionarios similares.

La segunda clasificación se realiza con base en el tipo de serie de datos escogida para el análisis. Una preferencia de los investigadores en

la actualidad es la utilización de series de duración parcial (SDP). Ben-Zvi (2009); Emmanouil, Langousis, Nikolopoulos y Anagnostou (2020), y Chang, Lai y Faridah (2013) han obtenido resultados satisfactorios empleando la distribución generalizada de Pareto (GP); esta metodología consiste en la obtención de una serie de datos con valores superiores a un umbral, que permitiría por cada año obtener valores picos que con una serie de máximos anuales (SMA) quedarían excluidos. Dicha técnica ya ha sido usada con anterioridad en prestaciones similares. Egea-Pérez, Cortés-Molina y Navarro-González (2021) realizan un análisis con la lluvia en localidades con escasos datos anuales; Masseran y Safari (2020) aplican las SDP para obtener la valoración de riesgo de contaminación del aire extrema basada en una aproximación IDF. A pesar de este auge, Sane *et al.* (2018), y Olsson, Södling, Berg, Wern y Eronn (2019) mantienen investigaciones con SMA en países como Senegal y Suecia, respectivamente.

Hay estudios comparativos entre las metodologías con SMA y SDP. Vrbán, Wang, McBean-Edward, Binns y Gharabaghi (2018) demuestran que la SDP para la obtención de tormenta de diseño es más efectiva que la SMA, pues la lluvia excede de un 4 a 10 % y por tanto es más conservador, además de que se obtienen resultados mayores en periodos de retorno de entre 2 y 5 años. Van Campenhout, Houbrechts, Peeters y Petit (2020) utilizan SDP para encontrar una relación con la SMA con respecto al periodo de retorno, aunque el estudio se realiza para series de gastos máximos de escurrimiento. Agilan y Umamahesh (2017b) comparan sus modelos no estacionarios obtenidos con SMA y SDP obteniendo resultados similares a Vrbán *et al.* (2018), aunque afirman

que para cortas duraciones y pequeños periodos de retorno la diferencia es más elevada que para grandes duraciones y mayores periodos de retorno.

Singh (2017) recopila varios trabajos y expone una valoración entre la relación que existe entre el periodo de retorno de las SMA y SDP, donde se llega a la siguiente expresión:

$$T_P = - \frac{1}{\ln\left(1 - \frac{1}{T_A}\right)} \quad (1)$$

Donde  $T_P$  es el periodo de retorno del análisis con SDP y  $T_A$  es el periodo de retorno obtenido en un análisis con SMA. Incluso para periodos de retorno mayores a 10 años, esta expresión se puede reducir a  $T_A = T_P + 1/2$ , lo cual hace que ambos resultados, tanto para SDP como para SMA, sean relativamente iguales.

Bajo los anteriores criterios el objetivo principal de esta contribución es la obtención de las curvas IDF correspondientes a la estación meteorológica Yabú, cercana a la ciudad de Santa Clara, a través del estudio de 30 años de registros comprendidos entre 1990 y 2019 de eventos lluviosos superiores a 25 mm en 24 h. El estudio se basa en la función de probabilidad de mejor ajuste y en la ecuación paramétrica con similar condición. En tal sentido, serán evaluadas tres funciones de distribución y cuatro ecuaciones paramétricas para el mejor ajuste. Con posterioridad se obtendrán las respectivas curvas IDF, características de

la estación en estudio, para emplearlas en un futuro estudio de regionalización en próximas contribuciones.

## Materiales y métodos

La estación agrometeorológica Yabú (Código 78343) se encuentra en la provincia de Villa Clara, Cuba, localizada en los 22° 26' N y 79° 59' W, a 116.44 m sobre el nivel medio del mar, con presencia de un relieve llano (Figura 1), a aproximadamente 7 km del centro de la ciudad de Santa Clara. Está enmarcada en la margen este de la cuenca del río Sagua La Grande, sistema fluvial más amplio de la provincia, y en donde se encuentran dos de los embalses de mayor importancia económica de la región: Palmarito y Alacranes. Aunque no es la única estación meteorológica de la cuenca, su privilegiada posición en la misma permite tener un certero comportamiento de las variables climáticas que afectan el lugar. Comenzó su funcionamiento el 3 de septiembre de 1976, fecha de su primer registro de medición de todas las variables propias. Posee varios instrumentos adecuados a la actividad meteorológica, entre los cuales se encuentra el pluviómetro y pluviógrafo:

1. Pluviómetro: modelo (USWB), con fecha de inicio de mediciones en 1976. Las mediciones se realizan en milímetros (mm).
2. Pluviógrafo: marca (Estándar), modelo (P-2), serie (281), fabricado en la antigua Unión de Repúblicas Socialistas Soviéticas (URSS), con fecha de inicio de mediciones en 1976. Sin embargo, los registros



desde 1976 hasta 1990 poseen intervalos donde se han producido pérdidas de información que comprometen su análisis y por ello no serán considerados en esta investigación.



**Figura 1.** Ubicación geográfica de la estación meteorológica Yabú, provincia de Villa Clara, Cuba.

Se realiza un análisis de 30 años de registros pluviográficos desde 1990 hasta 2019 con registros solo interrumpidos en periodos que no exceden los tres meses debido a roturas, mantenimiento, mal funcionamiento o mala calidad del pluviograma para su lectura. La Tabla 1 presenta las series de datos obtenidas del procesamiento de las cartas de la estación.

**Tabla 1.** Intensidades máximas en mm/min para distintas duraciones recuperadas de la estación Yabú.

Año	1 hora	2 horas	4 horas	12 horas	24 horas
1990	-	-	-	-	-
1991	1.231	0.780	0.400	0.139	0.061
1992	0.930	0.500	0.250	0.084	0.043
1993	0.730	0.380	0.190	0.083	0.042
1994	0.939	0.563	0.291	0.097	0.051
1995	0.900	0.480	0.240	0.082	0.046
1996	0.720	0.482	0.263	0.136	0.115
1997	0.940	0.500	0.250	0.084	0.042
1998	0.910	0.470	0.255	0.158	0.194
1999	0.840	0.460	0.230	0.134	0.067
2000	1.080	0.570	0.280	0.106	0.053
2001	0.620	0.320	0.190	0.091	0.045
2002	0.870	0.520	0.260	0.094	0.047
2003	0.980	0.650	0.330	0.116	0.058
2004	1.100	0.620	0.310	0.115	0.058
2005	1.050	0.580	0.290	0.162	0.115
2006	0.920	0.480	0.240	0.101	0.051
2007	0.890	0.440	0.220	0.086	0.043
2008	1.200	0.600	0.300	0.153	0.076
2009	-	-	-	-	-
2010	1.000	0.540	0.270	0.114	0.057
2011	1.220	0.740	0.490	0.169	0.085
2012	1.000	0.530	0.260	0.093	0.075
2013	0.980	0.680	0.400	0.138	0.069
2014	1.799	1.035	0.520	0.184	0.092
2015	1.140	0.610	0.310	0.111	0.056
2016	1.020	0.550	0.280	0.095	0.048
2017	0.740	0.370	0.230	0.161	0.064
2018	1.240	0.620	0.310	0.152	0.081
2019	0.680	0.340	0.170	0.063	0.038

En el caso del año 1990, sus registros se extraviaron y en 2009 la cantidad de lluvias analizadas fue insuficiente por problemas en la recopilación de datos del equipo (roturas y mal estado de pluviograma). El pluviograma o carta pluviográfica resultante de la medición analógica del equipo posee una escala mínima de 10 minutos por cada 0.5 cm de papel en la escala horizontal, lo cual permite que en la digitalización se pueda presenciar un sifonado o dos del equipo dentro de un intervalo de tiempo de 10 min, con lo cual se puede reducir al mínimo intervalo de 5 min que se puede identificar visualmente. Las duraciones procesadas son de 5, 10, 20, 40, 60, 90, 120, 150, 240, 300, 720, 1 440, 2 880 y 4 320 minutos; las últimas tres corresponden a 24, 48 y 72 horas de duración, típicas de eventos ciclónicos. En la Tabla 1 se muestra la SMA para 1, 2, 4, 12 y 24 horas.

## **Análisis de datos faltantes**

La tenencia de datos faltantes por diversos motivos en una serie de datos es un problema común para cualquier investigador. Según Molenberghs, Fitzmaurice, Kenward, Tsiatis y Verbeke (2015), y Little y Rubin (1987) existen tres tipos básicos de estos datos:

- Completamente al azar (MCAR, Missing Completely at Random). Representa una situación para la cual la ausencia es independiente de las variables de una investigación. Es una ausencia de datos debida exclusivamente al azar.

- Perdidos al azar (MAR, Missing at Random). Se refiere a que la ausencia de datos está presente en las variables independientes del estudio, pero no a la dependiente. Es el caso más típico de pérdidas de datos y se considera el tipo presente en esta investigación debido a que existen ausencias de datos en la fuente analógica primaria (pluviógrafo) por pérdidas de cartas pluviográficas y deterioro de las mismas. Las intensidades obtenidas de esta fuente entonces se consideran las variables independientes del estudio.
- Un proceso que no es MCAR o MAR es no aleatorio (MNAR, Missing Non at Random).

Singh (2017) propone tres caminos a seguir cuando se presenten ausencias de datos:

- Omitir datos perdidos: solo se analizan los datos existentes sin completar los registros faltantes.
- Imputar datos perdidos: mediante técnicas y metodologías establecidas lograr encontrar un sustituto al dato inexistente que propicie el relleno de la serie. Se analizarán en esta contribución los algoritmos Expectation-Maximization (EM) y Multiple Imputation (IM).
- Acomodar datos perdidos: se realiza con técnicas propias del relleno de datos, pero usando la estadística de la serie de datos.

Los métodos de máxima verosimilitud, como EM, se pueden aplicar en cualquier problema de estimación. En el análisis de datos omitidos, y asumiendo que los datos faltantes siguen un patrón MAR, se demuestra que la distribución marginal de los registros observados está asociada con

una función de verosimilitud para un parámetro desconocido, bajo el supuesto de que el modelo es adecuado para el conjunto de datos completo (Mallol, 2017).

De acuerdo con Little y Rubin (1987), citados en Mallol (2017), a esta función se le conoce como la función de verosimilitud, la cual ignora el mecanismo que generó los datos faltantes. El procedimiento para estimar los parámetros de un modelo utilizando una muestra con datos faltantes se resume a continuación:

- Estimar los parámetros del modelo con los datos completos con la función de máxima verosimilitud.
- Utilizar los parámetros estimados para predecir los valores omitidos.
- Sustituir los datos por las predicciones y obtener nuevos valores del parámetro, maximizando la verosimilitud de la muestra completa.

Un procedimiento eficiente para maximizar la verosimilitud cuando existen datos faltantes es el algoritmo Expectation-Maximization (Miró, Caselles, & Estrela, 2017).

El método de imputación múltiple consiste en realizar varias imputaciones de las observaciones faltantes para luego analizar los conjuntos de datos completados y combinar los resultados obtenidos para obtener una estimación final. El análisis de imputación múltiple está dividido en tres fases: fase de imputación, fase de análisis y fase de puesta en común (Mallol, 2017).

La fase de imputación crea múltiples copias de los conjuntos de datos ( $m$ ), y cada una de ellas contiene diferentes estimaciones de los valores perdidos. Conceptualmente, este paso es una versión iterativa de

la imputación por regresión estocástica, aunque sus fundamentos matemáticos se basan en muchas ocasiones en los principios de estimación bayesiana (Mallol, 2017).

- El objetivo de la fase de análisis, como su nombre indica, es analizar los conjuntos de datos rellenados. Este paso aplica los mismos procedimientos estadísticos que un individuo hubiera utilizado si tuviera todos los datos. La única diferencia es que realizamos cada análisis  $m$  veces, una para cada conjunto de datos imputados.
- La fase de análisis nos lleva a  $m$  conjuntos de estimaciones de parámetros y errores estándar, con lo que el propósito de la fase de puesta en común es combinar todo en un conjunto simple de resultados. Little y Rubin (1987) perfilaron fórmulas relativamente sencillas para poner en común las estimaciones de los parámetros y los errores estándar. Por ejemplo, la estimación del parámetro puesto en común es simplemente la media aritmética de las  $m$  estimaciones de la fase de análisis. Combinar los errores estándar es ligeramente más complejo, pero sigue la misma lógica. El proceso de analizar conjuntos de datos múltiples y poner en común los resultados parece laborioso, pero los paquetes de *software* de imputación múltiple, como SPSS, XLSTAT y R, automatizan completamente el procedimiento. Las  $m$  estimaciones son combinadas en una estimación en conjunto y una matriz de varianzas-covarianzas utilizando las reglas de Rubin, que se basan en la teoría asintótica en un marco bayesiano. La matriz de varianzas-covarianzas combinada incorpora la variabilidad dentro de la imputación (incertidumbre sobre los resultados de unos conjuntos

de datos imputados) y la variabilidad entre las imputaciones (relejando la incertidumbre debido a la información perdida).

Para realizar este proceso iterativo y agilizar el proceso de imputación para tomar las decisiones más convenientes se utiliza el *software* SPSS en su versión 22. El procedimiento será la obtención de resultados mediante EM y tres imputaciones aleatorias con IM (mediante la técnica de regresión lineal), y se escogerán los resultados de intensidades máximas más grandes obtenidas en cualquier método, siempre garantizando los siguientes aspectos:

- La media de la serie no puede variar de modo significativo.
- La varianza y covarianzas deben permanecer constantes, con un intervalo de error inferior al 1 %.
- Los resultados obtenidos en los años 1990 y 2009 que se pretenden imputar deberán tener un orden lógico de obtención de intensidades en relación con su duración, es decir, las intensidades de 5 min son mayores que las de 10 min y éstas, a su vez, mayores que las de 20 min y así consecutivamente.
- La serie completada será analizada de forma consecutiva para la obtención de *outliers* (valores fuera de rango); ninguno de los valores imputados puede ser encontrado como *outlier*, y en caso de detectarse en alguno no imputado, se reconfigura la SMA y se vuelve a comenzar el proceso de imputación.



## Análisis de *outliers*

En las ciencias geofísicas, como la hidrología, las observaciones son regularmente obtenidas para su análisis y la captura de cambios en los procesos históricos de un intervalo de tiempo. Los datos hidrológicos muchas veces contienen observaciones extremas o datos anómalos (*outliers*) debidos a eventos reales o factores externos a la medición (Singh, 2017).

Un dato anómalo es aquel que aparece alejado del conjunto de los datos. La presencia de datos anómalos en una muestra de datos puede crear dificultades a la hora de ajustar una distribución a la muestra. En una muestra puede haber datos anómalos de valores elevados o bajos, o de ambos tipos, que pueden influir de diferentes maneras en el análisis de frecuencia. Aunque el tratamiento de los datos anómalos es un tema aún muy debatido, se han utilizado ciertos procedimientos en hidrología para identificarlos y tratarlos, como los descritos en la publicación del Water Resources Council (1981) de EUA para el análisis de frecuencia de crecida o precipitaciones extremas (OMM, 2011).

OMM (2011) y Naghettini (2017) recomiendan el uso del método de US-WRC (United States-Water Resources Council). Para aplicarla será necesario suponer que los logaritmos u otra función de la serie hidrológica están distribuidos normalmente, ya que la prueba solo es aplicable a muestras obtenidas de una población normal. Para efectuar la prueba de US-WRC se calculan las dos expresiones siguientes:



$$X_H = \exp(\bar{x} + K_N s) \quad (2)$$

$$X_L = \exp(\bar{x} - K_N s) \quad (3)$$

Donde  $\bar{x}$  y  $s$  en las ecuaciones (2) y (3) son la media y la desviación típica de los logaritmos naturales de la muestra, respectivamente;  $K_N$ , la estadística de Grubbs y Beck tabulada para varios tamaños de muestra y niveles de importancia;  $N$ , el tamaño de la muestra;  $X_H$ , el límite superior de la prueba, y  $X_L$  es el límite inferior. Para  $5 \leq N \leq 150$ ,  $K_N$  se puede calcular a partir de la Ecuación (4) (Stedinger *et al.*, 1993 citado en OMM, 2011):

$$K_N = -0.9043 + 3.345\sqrt{\log(N)} - 0.4046 \log(N) \quad (4)$$

## Calidad de la serie de datos

Para que los resultados del análisis de frecuencia sean teóricamente válidos, las series de datos deberán satisfacer determinados criterios estadísticos, como los de aleatoriedad, independencia, homogeneidad y estacionalidad (OMM, 2011). Este texto también recomienda las pruebas idóneas a aplicar para comprobar las hipótesis en cada caso, como se muestra en la Tabla 2. Las pruebas indicadas son de tipo no paramétrico, evitando así cualquier supuesto sobre la distribución paramétrica subyacente de los datos.

**Tabla 2.** Pruebas no paramétricas para el análisis de calidad de los datos en la estación meteorológica del Yabú.

Criterios estadísticos	Prueba recomendada	Intervalo de confianza en %
Aleatoriedad	Prueba de Rachas	95
Independencia (1)	Prueba de Mann-Withney,	
Independencia (2)	Prueba de Wald-Wolfowitz	
Estacionalidad	Pruebas de Mann- Kendall, Sen 's Slope	

Por la importancia observada en el análisis del estado del arte de este tópico a las pruebas de tendencia, para confirmar la existencia de estacionalidad o no de una serie de máximos anuales, a continuación se propone el resumen de la teoría que la acompaña.

La prueba de Mann-Kendall es una prueba no paramétrica basada en la correlación de rango que permite evaluar el significado de una tendencia. La hipótesis de tendencia nula  $H_0$  consiste en que una muestra de datos ordenados cronológicamente es independiente y está idénticamente distribuida. La estadística  $S$  se define como sigue (Maity, 2018):

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sgn}(x_j - x_i) \quad (5)$$

Donde:

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{si } x > 0 \\ 0 & \text{si } x = 0 \\ -1 & \text{si } x < 0 \end{cases} \quad (6)$$

Cuando  $n \geq 40$ , la estadística  $S$  presenta una distribución asintóticamente normal de media 0 y varianza dada por la ecuación siguiente:

$$\operatorname{Var}\{S\} = \frac{1}{18[n(n-1)(2n+5) - \sum_t t(t-1)(2t+5)]} \quad (7)$$

Donde  $t$  es el tamaño de un grupo ligado dado y  $\Sigma$  es la suma del conjunto de todos los grupos ligados de la muestra de datos. La estadística de prueba normalizada  $K$  se calcula mediante la ecuación siguiente:

$$K = \frac{S-1}{\sqrt{\operatorname{Var}(S)}}; 0; \frac{S+1}{\sqrt{\operatorname{Var}(S)}} \text{ para: } S > 0, S = 0, S < 0, \text{ respectivamente} \quad (8)$$

La estadística normalizada  $K$  presenta una distribución normal estándar de media igual a 0 y varianza igual a 1. El valor de probabilidad  $P$  de la estadística  $K$  de los datos de la muestra puede estimarse mediante la función de distribución acumulativa normal en la forma:

$$P = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{t^2}{2}} dt \quad (9)$$

Para muestras de datos independientes sin tendencia alguna, el valor de  $P$  deberá ser igual a 0.5. Cuando los datos de la muestra presentan una tendencia positiva acusada, el valor de  $P$  deberá ser próximo a 1, mientras que una tendencia negativa acentuada deberá arrojar un valor  $P$  cercano a 0. Si los datos de la muestra están correlacionados serialmente, será necesario blanquear de manera previa los datos y aplicar una corrección para calcular la varianza (OMM, 2011).

Para la tendencia lineal, la pendiente generalmente se estima calculando la estimación de mínimos cuadrados mediante regresión lineal. Sin embargo, solo es válido cuando no existe una correlación serial y el método es muy sensible a los valores atípicos. Sen (1968), citado en OMM (2011), desarrolló un método más robusto.

La pendiente de una tendencia puede estimarse como sigue:

$$Q = \text{mediana} \left( \frac{x_i - x_j}{i - j} \right); \forall j < i \quad (10)$$

Donde  $Q$  es el valor estimado de la pendiente de la tendencia y  $x_j$  es la  $j$ -ésima observación. Las tendencias ascendentes se representan mediante un valor positivo de  $Q$ , mientras que las descendentes se representan mediante un valor negativo de  $Q$ .

El estimador de pendiente de Sen es la mediana de los valores  $N'$  de  $Q$ . Se sigue el mismo procedimiento ya sea que haya una o varias observaciones por periodo de tiempo.

Sen (1968), citado en OMM (2011), proporciona un método no paramétrico para obtener un intervalo de confianza para esta pendiente, aunque se usa de forma más común un método simple de aproximación normal. Para ello se necesita la desviación estándar del estadístico de Mann-Kendall,  $S$  (OMM, 2011).

## Funciones de distribución de probabilidad

Las distribuciones de probabilidad se utilizan en muy diversos estudios hidrológicos, particularmente en estudios de caudales altos y bajos extremos, crecidas, volúmenes de embalse, cantidades de precipitación de lluvia y modelos de series temporales. Cabe destacar que en el estudio de SMA, las distribuciones de análisis se encuentran bien definidas, y en los últimos años se ha venido usando con más fuerza la distribución de valores extremos generalizada (GVE, Generalized Extreme Value, por sus siglas en inglés) usadas tanto en modelos estacionarios como no estacionarios (Olsson *et al.*, 2019; Yong *et al.*, 2021; Agilan & Umamahesh, 2017a).

La Tabla 3 muestra las distribuciones de mejor ajuste obtenidas por diversos autores, y sus modelos y tipos de relaciones IDF de acuerdo con la clasificación que se explicó con anterioridad.

**Tabla 3.** Recopilación de estudios de IDF y resumen de sus características de mayor relevancia.

Referencia	Curvas estacionarias o no estacionarias	Tipo de serie usada	Distribución de mejor ajuste	Método de ajuste utilizado	Ubicación
Olsson <i>et al.</i> (2019)	Estacionarias	SMA <sup>1</sup>	Generalized Extreme Value	Máxima Vecindad	Suecia
Yong <i>et al.</i> (2021)	Estacionarias	SMA	Gumbel	L-Momentos	Malasia
Agilan y Umamahesh (2017a)	No estacionarias	SMA	Generalized Extreme Value	Máxima Vecindad	Wilmington (EUA) Hyderabad (India)
Agilan y Umamahesh (2017b)	No estacionarias	SMA <sup>1</sup>	Generalized Extreme Value	Máxima Vecindad	Hyderabad (India)
Ganguli y Coulibaly (2017)	Ambos	SMA	Generalized Extreme Value	Markov chain (DE-MC) Monte Carlo simulation Vrbán <i>et al.</i> (2018)	Ontario (Canadá)
Ng, Tiang, Huang, Noh y Al-Mansob (2021)	Estacionarias	SMA <sup>1</sup>	Generalized Extreme Value.	Máxima Vecindad	Kelantan (Malasia)
Sane <i>et al.</i> (2018)	Estacionarias	SMA	Generalized Extreme Value.	L-Momentos	Senegal
Vrbán <i>et al.</i> (2018)	Estacionarias	SMA <sup>1</sup>	Generalized Extreme Value.	L-Momentos	Ontario (Canadá)

<sup>1</sup>También realiza un estudio con PDS con la distribución generalizada de Pareto.

Para este estudio se utilizan tres aproximaciones estadísticas para describir los eventos de corta duración de la lluvia de acuerdo con la teoría de valores extremos: (a) distribución de valores extremos (GEV, Generalized Extreme Value); (b) distribución logarítmica normal bi y tri paramétrica (LN2 y LN3, respectivamente); (c) distribución logarítmica de Pearson tipo III (LP3).

La función acumulativa GEV se expresa de la siguiente manera:

$$F(x; \mu, \sigma, \xi) = \exp \left[ - \left( 1 + \xi * \left( \frac{x - \mu}{\sigma} \right)^{-\frac{1}{\xi}} \right) \right] \quad (11)$$

Donde  $\mu$  es el parámetro de localización;  $\sigma$ , el parámetro de escala, y  $\xi$  es el parámetro de forma. GEV representa una familia de distribuciones dependiendo del valor de  $\xi$ : Gumbel ( $\xi = 0$ ), Fréchet ( $\xi > 0$ ) y Weibull ( $\xi < 0$ ) (Olsson *et al.*, 2019). Para este estudio se utilizará  $\xi = 0$ .

Por lo general, las distribuciones de crecida presentan una asimetría positiva y no se de modo adecuado por una distribución normal. En muchos casos, la variable aleatoria correspondiente al logaritmo de los flujos de crecida estará adecuadamente descrita por una distribución normal. La distribución log-normal paramétrica tiene una función de densidad de probabilidad indicada en la Ecuación (12). Con frecuencia, los logaritmos de una variable aleatoria  $X$  no se ajustan a una distribución normal. En tales casos, el problema puede resolverse introduciendo un

parámetro de contorno  $\tau$  antes de calcular los logaritmos, obteniéndose así una distribución log-normal triparamétrica (OMM, 2011):

$$F(x; \mu, \sigma) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{1}{2} \left( \frac{\ln(x) - \mu}{\sigma} \right)^2 \right] \quad (12)$$

$$F(x; \mu, \sigma, \xi) = \frac{1}{(x-\tau)\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{1}{2} \left( \frac{\ln(x-\tau) - \mu}{\sigma} \right)^2 \right] \quad (13)$$

La distribución log-Pearson de tipo III (LP3) describe una variable  $x$  cuyo logaritmo  $y = \log x$  presenta una distribución Pearson de tipo III. Ésta fue recomendada para la descripción de crecidas en EUA por el Water Resources Council (Consejo de Recursos Hídricos) de ese país, inicialmente en 1966, y después por el Interagency Advisory Committee on Water Data (Comité Consultivo Inter-Organismos sobre Datos Hídricos) en 1982. También fue adoptada en Canadá, entre otros métodos (OMM, 2011):

$$F(x; \beta, \xi, \alpha) = \frac{|\beta| \{ \beta [\ln(x) - \varphi] \}^{\alpha-1} \exp \{ -\beta [\ln(x) - \varphi] \}}{x \Gamma(\alpha)} \quad (14)$$

En esta oportunidad,  $\alpha$  y  $\beta$  son parámetros de escala y forma, mientras  $\varphi$  es un parámetro de localización.



## Estimación de parámetros

Tal vez el planteamiento más simple es el método de los momentos, que permite obtener estimaciones de los parámetros de tal modo que los momentos teóricos de una distribución concuerden con los momentos de la muestra calculados. El procedimiento recomendado para los organismos federales de EUA, referencias citadas todas en OMM (2011), está basado en los momentos de los logaritmos de los flujos de crecida  $X = \log Q$ .

Teegavarapu, Salas y Stedinger (2019) proponen el concepto de  $L$ -momentos como una combinación lineal de los momentos ponderados por probabilidad (PWM, Probability-Weighted Moments). Teegavarapu *et al.* (2019) explican a través de algunos estudios anteriores la ecuación que describe los PWM y los ejemplos de estimadores para cualquier distribución de probabilidad.

## Bondad de ajuste

En hidrología se dispone de varias pruebas estadísticas rigurosas y útiles para determinar si es o no razonable concluir que un conjunto de observaciones dado ha sido obtenido de una familia de distribuciones en particular (OMM, 2011). La prueba de Kolmogorov-Smirnov permite obtener cotas para cada una de las observaciones de una gráfica de probabilidad cuando la muestra ha sido efectivamente obtenida de la distribución supuesta.

Este procedimiento es un test no paramétrico que permite probar si dos muestras provienen del mismo modelo probabilístico. Supóngase que se tienen dos muestras de tamaño total  $N = m + n$  compuestas por observaciones  $x_1, x_2, x_3, \dots, x_n$  e  $y_1, y_2, y_3, \dots, y_m$ . El test supone que las variables  $x$ ,  $y$  son mutuamente independientes y que cada  $x$  proviene de la misma población continua  $P_1$ , y que las variables  $y$  provienen de otra población continua  $P_2$ . La hipótesis nula es que ambas distribuciones son idénticas, es decir, son dos muestras de la misma población.

El test se basa en calcular el estadígrafo  $J$  definido como el valor máximo de la diferencia absoluta entre dos funciones de distribución acumulada.

Entre las ventajas de la prueba Kolmogórov-Smirnov se encuentra su superioridad con respecto a la prueba Chi cuadrado ( $X^2$ ), su facilidad de cálculo y a que no usa una agrupación de datos, además de que el estadístico es independiente de la distribución de frecuencia esperada, solo depende del tamaño de la muestra.

## Modelos de intensidad-frecuencia-duración

Los resultados del análisis de frecuencias se expresan por lo general en términos de las relaciones intensidad-duración-frecuencia en un lugar dado, o se exponen en forma de atlas de frecuencia de precipitación, en los que se indican las alturas acumuladas de precipitación de lluvia para distintas duraciones y periodos de retorno en la región de interés (OMM, 2011).

A partir de la parametrización propuesta por Sherman (1931), citada en Gutiérrez y Barragán (2019), se adopta en todo el mundo la representación matemática y gráfica del cálculo de curvas intensidad (I)-duración (D)-frecuencia (F). Esta formulación es una ecuación racional del tipo:

$$i_d^T = f(x) = \frac{P(T)}{Q(d)} \quad (15)$$

El numerador  $P(T)$  es función del periodo de retorno ( $T$ ) e indica el cuantil para la frecuencia acumulada ( $1-1/T$ ) de una función de distribución de probabilidad de una variable aleatoria. Para el denominador  $Q(d)$ , que también es función del tiempo o duración de la intensidad, se admite que puede expresarse como un polinomio que permite factorización (Gutiérrez & Barragán, 2019). Quedando la siguiente parametrización general:

$$i_d^T = \frac{kT^m}{(d^\theta + c)^n} \quad (16)$$

Donde  $i_d^T$  es la intensidad máxima de precipitación, expresada en mm/min o mm/h;  $T$ , el periodo de retorno en años;  $d$ , la duración de la precipitación en minutos, y  $k$ ,  $m$ ,  $\theta$  y  $n$  son los parámetros de ajuste a estimar. La Ecuación (16) es ampliamente utilizada y varios autores han propuesto diferentes valores de los parámetros  $k$ ,  $m$ ,  $\theta$  y  $n$ . En todos los casos, dichos parámetros se estiman por procedimientos numéricos,

analíticos, numérico lineal, no lineal, estadísticos y de optimización. Se pueden calcular valores óptimos mediante un procedimiento de prueba y error, como citan Gutiérrez y Barragán (2019). Sin embargo, en un ningún caso se tiene evidencia de que cualquiera de estos parámetros tenga significado físico. Es decir, no se ha demostrado que los parámetros de la Ecuación (16) estén relacionados con alguna característica fisiográfica o climatológica del entorno.

De la Ecuación (16) se desprenden varios modelos de ajuste donde los valores de  $k$ ,  $m$ ,  $\theta$  y  $n$  son simplificados o reducidos a 0. En esta contribución los modelos a ajustar serán Montana, Sherman, Bernard y Chow, citados en Gutiérrez y Barragán (2019), de las ecuaciones (17) a (20), respectivamente:

$$i_d^T = \frac{kT^m}{d^{\theta+C}} \quad (17)$$

$$i_d^T = \frac{kT^m}{(d+C)^n} \quad (18)$$

$$i_d^T = \frac{kT^m}{d^n} \quad (19)$$

$$i_d^T = \frac{kT^m}{d+C} \quad (20)$$

## Resultados y discusión

Como se explica en epígrafes anteriores, en los años 1990 y 2009 no se lograron obtener los valores de intensidades máximas anuales de la lluvia debido a distintos factores; para suplir esta carencia de datos se usaron las metodologías de Maximization-Expectation (Miró *et al.*, 2017) e imputación múltiple (Mallol, 2017) con cinco variantes; de estos cinco grupos de datos se escogió el resultado de mayor valor para completar el registro. En la Tabla 4 se expone un resumen de los resultados obtenidos y posteriormente se analizan los mismos.

**Tabla 4.** Resultados de la imputación de valores a los datos faltantes de 1990 y 2009 para duraciones de 1, 2, 4 y 12 h.

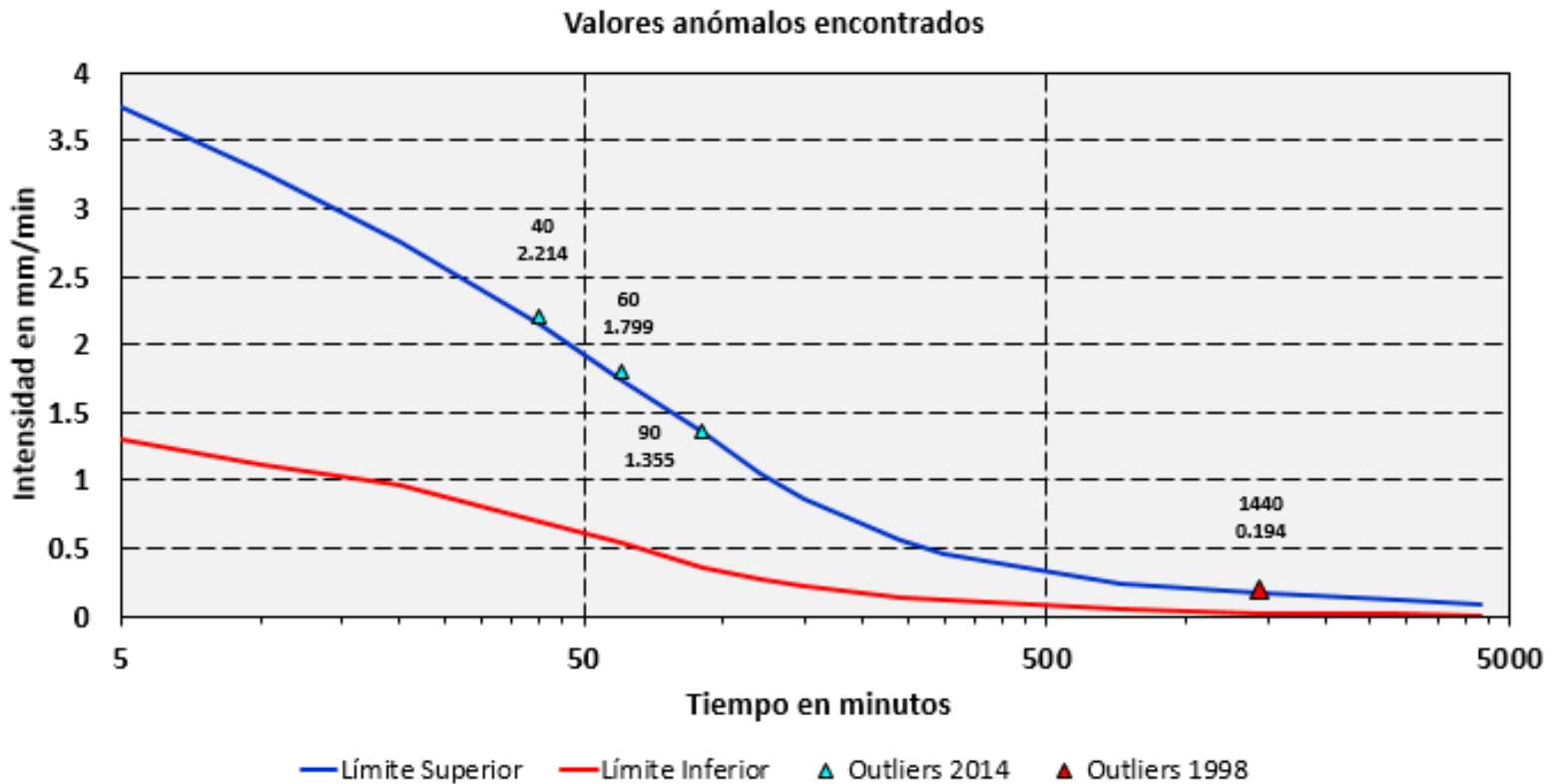
Hora	Valor imputado 1990 mm/min	Valor imputado 2009 mm/min	Media mm/min		Desviación estándar mm/min		Curtosis		Coeficiente de asimetría	
			S/I	C/I	S/I	C/I	S/I	C/I	S/I	C/I
1	1.008	0.856	0.99	0.99	0.23	0.22	4.66	4.91	1.48	1.45
2	0.572	0.488	0.55	0.55	0.15	0.14	3.45	3.58	1.29	1.24
4	0.287	0.298	0.29	0.29	0.08	0.08	2.36	2.26	1.46	1.38
12	0.118	0.107	0.12	0.12	0.03	0.03	0.98	0.83	0.37	0.36

C/I: con imputación; S/I: sin imputación.

Se resumen los estadísticos descriptivos de la muestra antes y después de la imputación. Como se puede apreciar, no existe una diferencia significativamente relevante entre los resultados para las duraciones seleccionadas, cumpliendo así con las condiciones propuestas anteriormente.

## Procesamiento de la serie de datos

La aplicación del método US-WRC demuestra la obtención de cuatro datos anómalos, tres de ellos ocurridos el 2 de agosto de 2014 con un evento lluvioso convectivo de 124.6 mm en 120 minutos; el otro valor se presentó el 24 de septiembre de 1998 con el paso del huracán George, el cual dejó un registro de 274 mm en 24 horas. En la Figura 2 se muestran estos resultados.



**Figura 2.** Datos anómalos de las series de datos.

Tras la revisión por especialistas y considerando criterios de experiencias con otras estaciones, los valores anómalos que presentan la serie se aceptan, ya que no exceden en un 10 % los límites superiores del modelo US-WRC para un nivel de confianza del 95 %.

Los resultados de las pruebas de calidad aplicadas a las series de máximos anuales recomendadas en OMM (2011), Rachas, Mann-Whitney (M-W), Wald-Wolfowitz (W-W) y Mann-Kendall (M-K) se resumen en la Tabla 5.

La palabra Sí en la tabla 5 significa que se acepta la hipótesis nula de que:

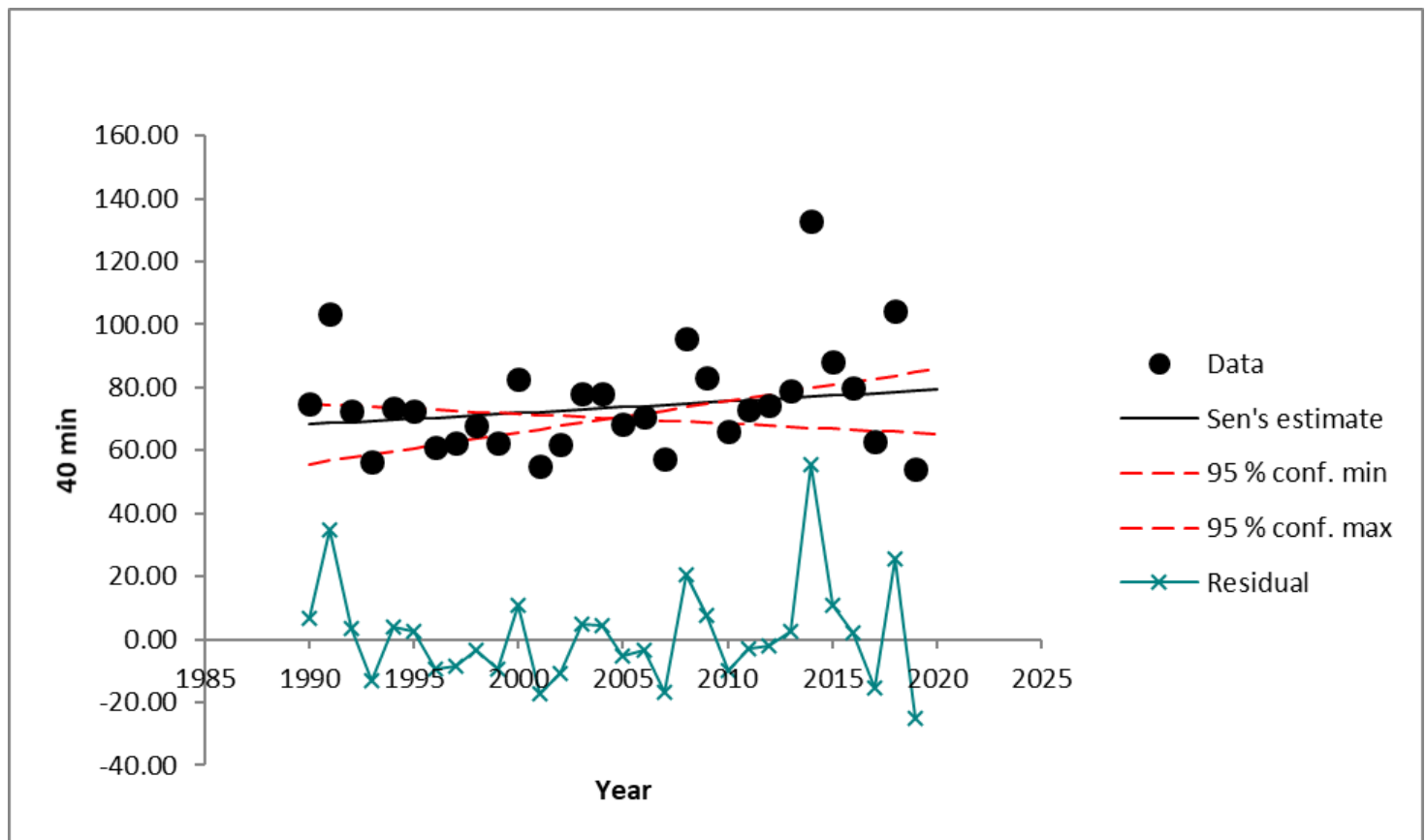
1. La serie es aleatoria para una significancia del 5 % (prueba de Rachas).
2. La serie es independiente para una significancia del 5 % (pruebas M-W, W-W)
3. La serie es estacional para una significancia del 5 % (prueba de M-K).

**Tabla 5.** Resultados de las pruebas de calidad para todas las series de datos.

Serie	Rachas	M-W	W-W	M-K
5 min	Sí	Sí	Sí	Sí
10 min	Sí	Sí	Sí	Sí
20 min	Sí	Sí	Sí	Sí
40 min	Sí	Sí	Sí	Sí
60 min	Sí	Sí	Sí	Sí
90 min	Sí	Sí	Sí	Sí
120 min	Sí	Sí	Sí	Sí
150 min	Sí	Sí	Sí	Sí
240 min	Sí	Sí	Sí	Sí
300 min	Sí	Sí	Sí	Sí
720 min	Sí	Sí	Sí	Sí
1 440 min	Sí	Sí	Sí	Sí
2 880 min	Sí	Sí	Sí	Sí
4 320 min	Sí	Sí	Sí	Sí



Al comprobar que todas las series son estacionales, se investiga sobre las tendencias de dichas series. La Figura 3 muestra el resultado de las pruebas de Mann-Kendall y Sen's Slope para la SMA de 40 minutos, en la cual se puede apreciar que aunque hay una tendencia al aumento por el estimador lineal, aún es insuficiente para considerarla dentro del análisis.



**Figura 3.** Análisis de tendencia para la serie de 40 minutos cuya pendiente fue de las más elevadas de las series estudiadas.

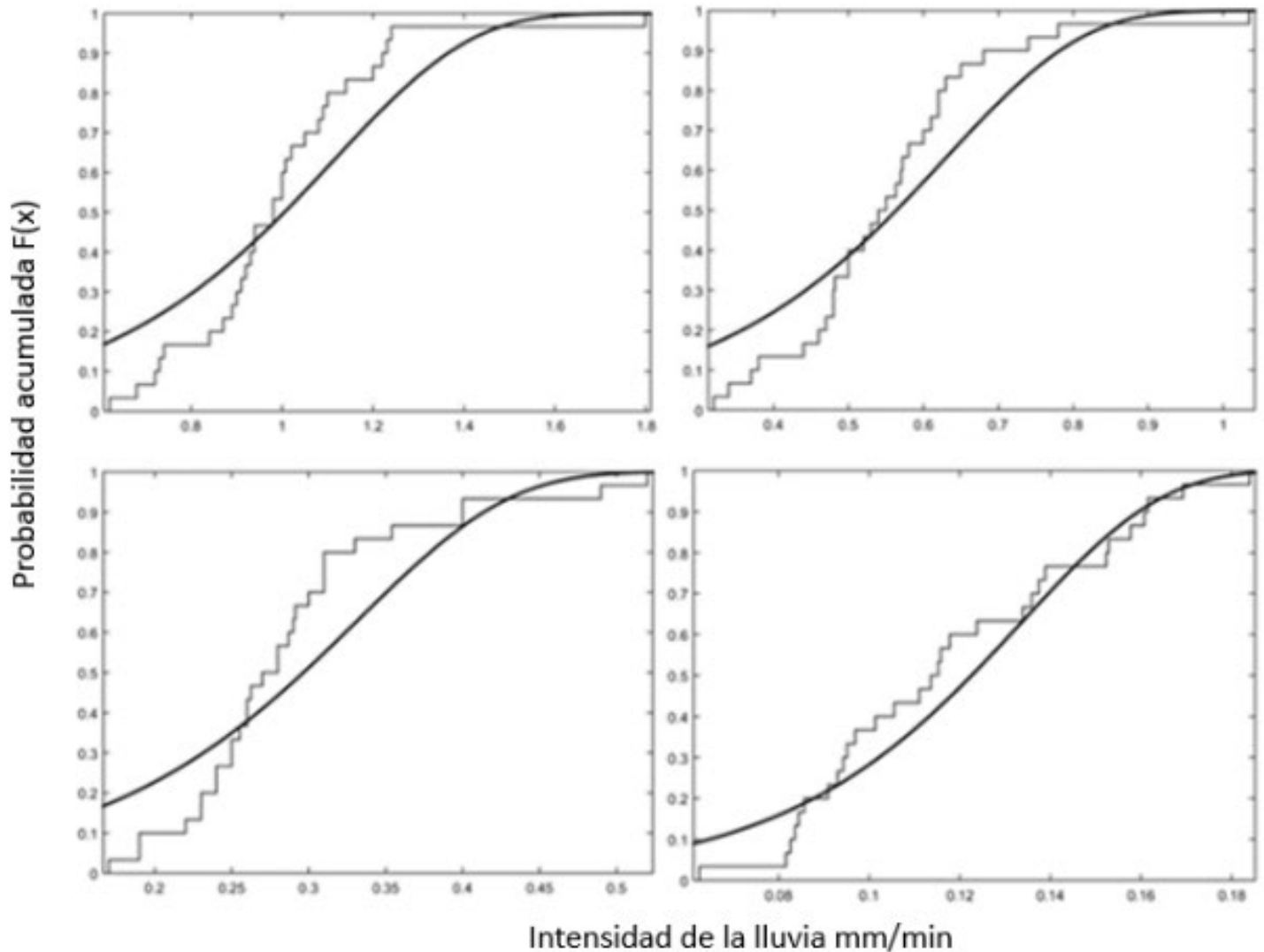
Los resultados obtenidos en las pruebas de calidad demuestran que las series de datos para la elaboración de las curvas IDF de la estación meteorológica del Yabú son aptas para su procesamiento probabilístico, destacando que se pueden usar modelos estacionarios para su representación sin necesidad de acudir a modelos no estacionarios.

El resumen del análisis con la distribución de Gumbel ( $\xi = 0$ ) con un ajuste por  $L$ -momentos se observan en la Figura 4 para 1, 2, 4 y 12 h. La Tabla 6 muestra los resultados de los parámetros de posición y escala obtenidos.

**Tabla 6.** Parámetros de la distribución de probabilidad de Gumbel obtenidos para las series de 1, 2, 4, 12 h.

Serie	Parámetro de localización $\mu$	Error estándar	Parámetro de escala $\sigma$	Error estándar
1 hora	1.11481	0.05835	0.29973	0.03425
2 horas	0.63042	0.03536	0.18169	0.02117
4 horas	0.33243	0.01902	0.09768	0.01192
12 horas	0.13387	0.00595	0.03069	0.00415

En la Figura 4 se grafica el ajuste para las duraciones antes señaladas usando la forma de probabilidad acumulada.



**Figura 4.** Ajuste a la función de Gumbel de probabilidad acumulada.

Para la obtención de la función de probabilidad de mejor ajuste se realiza el test de bondad de ajuste de Kolmogorov-Smirnov para los resultados de las funciones; (a) distribución de valores extremos ( $\xi = 0$ ); (b) distribución logarítmica normal bi y tri paramétrica, lo cual, para un

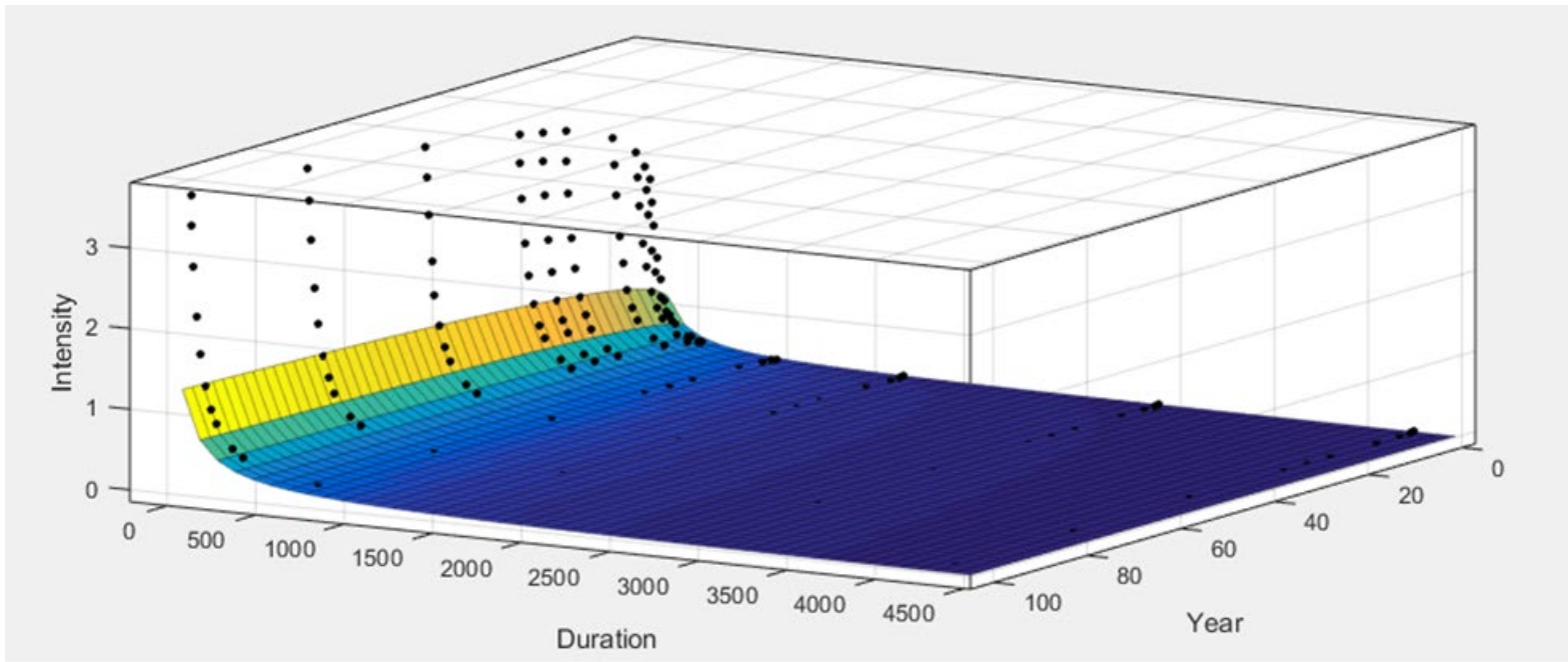
nivel de significancia del 5 %, demostró que el ajuste por Gumbel es el más efectivo.

La aplicación de los modelos Montana, Sherman, Bernard y Chow para el ajuste de los resultados de la función de probabilidad de Gumbel en la serie de máximos anuales se resume en la Tabla 7, donde se muestran los valores de  $k$ ,  $m$ ,  $\theta$ ,  $C$  y  $n$  para cada modelo, además de utilizar el coeficiente de correlación de Pearson para hallar cuál de ellos es el de mejor ajuste.

**Tabla 7.** Parámetros obtenidos y resultados de la correlación de Pearson para los modelos aplicados.

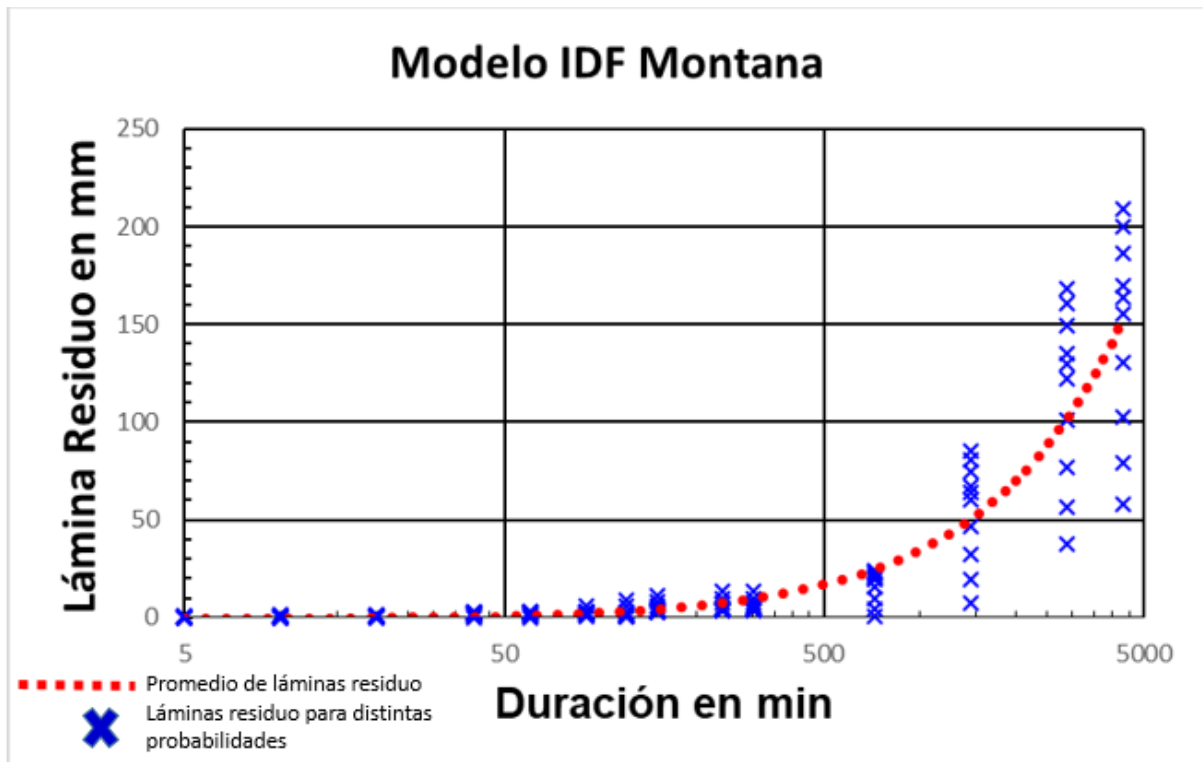
Modelo	$k$	$m$	$\theta$	$C$	$n$	Pearson
Montana	120.304	0.131	1.065	54.101	-	0.999194
Sherman	176.351	0.131	-	48.229	1.124	0.999192
Bernard	4.500	0.131	-	-	0.424	0.975301
Chow	89.042	0.131	-	38.757	-	0.999010

El análisis realizado muestra que el modelo de Montana es el que mejor se ajusta a los resultados de la distribución de probabilidad de Gumbel, con lo que se escoge para su análisis. La Figura 5 ilustra la función de ajuste de Montana para la estación meteorológica Yabú.



**Figura 5.** Modelo de Montana ajustado a los datos obtenidos por el modelo probabilístico de Gumbel.

Para profundizar más en el comportamiento del modelo es necesario convertir las curvas IDF obtenidas con Montana en curvas de precipitación-duración-frecuencia (PDF) con lo cual se puede obtener el residuo del modelo y se puede apreciar con claridad el rango de validez del mismo. En la Figura 6 se muestra un gráfico donde se plotea la lámina residuo en mm de precipitación contra la duración en minutos de la precipitación.



**Figura 6.** Lámina residuo encontrada para las distintas duraciones de la lluvia empleando los parámetros de la Tabla 7 en el modelo de Montana.

Los resultados obtenidos en la Figura 6 demuestran que el modelo de Montana presenta un comportamiento de los residuos lineal con pendiente ascendente que se evidencia con mayor claridad a partir de los 500 mm para todas las probabilidades. Esta observación permite inferir que puede existir un modelo Montana de IDF que describa las intensidades para duraciones inferiores a 360 min y otro para duraciones mayores a dicho valor, la Tabla 8 expone estos nuevos ajustes y la Ecuación (21) el resultado final para la estación de análisis.

**Tabla 8.** Parámetros y correlación encontrados para el modelo de Montana con duraciones inferiores a 360 min y superiores a dicho valor.

Modelo	$k$	$m$	$\theta$	$C$	$n$	Pearson
Montana (-6 horas)	135.5	0.1309	1.092	61.58	-	0.9981
Montana (+6 horas)	4.348	0.1826	0.5816	-2.905	-	0.9911

$$I = \begin{cases} \frac{135.5T^{0.1309}}{(D^{1.092} + 61.58)} & \text{para toda } D \leq 360 \\ \frac{4.348T^{0.1826}}{(D^{0.5816} - 2.905)} & \text{para toda } D > 360 \end{cases} \quad (21)$$

Donde:

$I$ : intensidad de la precipitación media en mm/min.

$T$ : periodo de retorno en años.

$D$ : duración de la tormenta en minutos.

Los valores obtenidos con la Ecuación (21) nunca superan el 20 % del valor obtenido de la función de probabilidad de Gumbel y las mayores discordancias (de 10 a 20 %) son resultados de los periodos de retorno más bajos (2 y 3 años), lo cual también desde el punto de vista del diseño ingenieril es favorable, al tener un porcentaje de seguridad para su uso. La Tabla 9 muestra los valores que se encuentran entre un 10 y 20 % superiores al valor obtenido en la función de probabilidad.

**Tabla 9.** Comparación entre los valores obtenidos con la Ecuación (19) y los valores de la distribución de probabilidad de Gumbel para los resultados más desfavorables.

<b>Función de probabilidad de Gumbel (mm/min)</b>	<b>Periodo de retorno (años)</b>	<b>Duración (min)</b>	<b>Resultado del modelo de Montana (mm/min)</b>
0.031	2	4 320	0.039
0.040	2	2 880	0.049
0.061	2	1 440	0.075
0.276	2	240	0.323
0.427	2	150	0.495
0.226	2	300	0.261
0.037	3	4 320	0.042
0.531	2	120	0.598
0.073	3	1 440	0.081

## Conclusiones

Tras el estudio y análisis de los resultados obtenidos durante esta investigación, los autores arribamos a las siguientes conclusiones:



- Se procesaron los datos de 30 años de registros pluviográficos de la estación meteorológica del Yabú, obteniendo las series de máximos anuales para duraciones de 5, 10, 20, 40, 60, 90, 120, 150, 240, 300, 720, 1 440, 2 880 y 4 320 minutos, las últimas tres corresponden a 24, 48 y 72 horas de duración, típicas de eventos ciclónicos.
- Hubo problemas en las mediciones de los años 1990 y 2009, sin embargo se aplicó el algoritmo de imputación múltiple mediante regresión lineal y se obtuvieron los datos faltantes sin evidenciar cambios significativos en las medidas de tendencia central de las series de datos.
- Se encontraron cuatro valores considerados *outliers* por el método de US-WRC, los cuales se aceptaron, ya que provenían de eventos lluviosos reales bien documentados, uno de ellos un huracán.
- La serie de datos se analizó para comprobar su aleatoriedad, independencia y tendencia, con lo cual se confirmaron las hipótesis y se concluyó que un modelo estacionario IDF representaría con claridad el fenómeno.
- La distribución de Gumbel se convino a las series de datos y se comprobó mediante el método de Kolmogorov-Smirnov que la distribución se ajusta a los mismos, siendo también la de mejores resultados en comparación con las distribuciones logarítmicas testeadas.

- El modelo de Montana parametrizó con mayor correlación los datos obtenidos de la distribución probabilística; pero se encontró que para duraciones mayores a 500 minutos existían residuos significativos que derivaron en que se definiera truncar la serie y elaborar un nuevo ajuste.
- Las ecuaciones propuestas para duraciones menores a 360 min y mayores a esa cifra se representaron por un modelo de Montana de dos ecuaciones con parámetros distintos; es de esperar que esta situación se repita para las demás estaciones meteorológicas de la provincia y por tanto se recomienda el análisis de esta contribución para su implementación.

### Agradecimientos

Los autores desean transmitir su más profundo agradecimiento al Centro Meteorológico Provincial de Villa Clara, cuyo apoyo en el desarrollo de las investigaciones ha sido encomiable y digno de destacar, a pesar de las duras condiciones impuestas por la pandemia de COVID-19 que afecta al país. A los estudiantes de Ingeniería Hidráulica de la Universidad Central Marta Abreu de Las Villas: José Solís Quintana, Alejandro Marrero, José Manuel Alba Bacallao y Roberto Fernández.

## Referencias

- Agilan, V., & Umamahesh, N. V. (2017a). Modelling nonlinear trend for developing non-stationary rainfall intensity-duration-frequency curve. *International Journal of Climatology*, 37(3), 1265-1281. Recuperado de <https://doi.org/https://doi.org/10.1002/joc.4774>
- Agilan, V., & Umamahesh, N. V. (2017b). Non-stationary rainfall intensity-duration-frequency relationship: A comparison between annual maximum and partial duration series. *Water Resources Management*, 31(6), 1825-1841. Recuperado de <https://doi.org/10.1007/s11269-017-1614-9>
- Agilan, V., & Umamahesh, N. V. (2017c). What are the best covariates for developing non-stationary rainfall Intensity-Duration-Frequency relationship? *Advances in Water Resources*, 101, 11-22. Recuperado de <https://doi.org/https://doi.org/10.1016/j.advwatres.2016.12.016>
- Ben-Zvi, A. (2009). Rainfall intensity-duration-frequency relationships derived from large partial duration series. *Journal of Hydrology*, 367(1), 104-114. Recuperado de <https://doi.org/https://doi.org/10.1016/j.jhydrol.2009.01.007>
- Chang, K. B., Lai, S. H., & Faridah, O. (2013). RainIDF: Automated derivation of rainfall intensity-duration-frequency relationship from annual maxima and partial duration series. *Journal of Hydroinformatics*, 15(4), 1224-1233. Recuperado de <https://doi.org/10.2166/hydro.2013.192>

- Egea-Pérez, R., Cortés-Molina, M., & Navarro-González, F. J. (2021). Analysis of rainfall time series with application to calculation of return periods. *Sustainability*, 13(14). Recuperado de <https://doi.org/10.3390/su13148051>
- Emmanouil, S., Langousis, A., Nikolopoulos, E. I., & Anagnostou, E. N. (2020). Quantitative assessment of annual maxima, peaks-over-threshold and multifractal parametric approaches in estimating intensity-duration-frequency curves from short rainfall records. *Journal of Hydrology*, 589, 125151. Recuperado de <https://doi.org/https://doi.org/10.1016/j.jhydrol.2020.125151>
- Ganguli, P., & Coulibaly, P. (2017). Does non-stationarity in rainfall require nonstationary intensity-duration-frequency curves? *Hydrology and Earth System Sciences*, 21(12), 6461-6483. Recuperado de <https://doi.org/10.5194/hess-21-6461-2017>
- Gregersen, I. B., Madsen, H., Rosbjerg, D., & Arnbjerg-Nielsen, K. (2017). A regional and nonstationary model for partial duration series of extreme rainfall. *Water Resources Research*, 53(4), 2659-2678. Recuperado de <https://doi.org/https://doi.org/10.1002/2016WR019554>
- Gutiérrez, A., & Barragán, R. (2019). Ajuste de curvas IDF a partir de tormentas de corta duración. *Tecnologías y ciencias del agua*, 10, 1-24. Recuperado de <https://doi.org/10.24850/j-tyca-2019-06-01>
- Little, R. J., & Rubin, D. B. (1989). El análisis de datos de ciencias sociales con valores faltantes. *Métodos e Investigación Sociológicos*, 18(2-3), 292-326.

- Maity, R. (2018). *Statistical methods in hydrology and hydroclimatology*. Berlin, Germany: Springer. Recuperado de <https://doi.org/https://doi.org/10.1007/978-981-10-8779-0>
- Mallol, P. (2017). *Importancia del tratamiento de datos perdidos. Aplicación en estudios longitudinales pequeños*. Barcelona, España: Universitat Oberta de Catalunya. Recuperado de <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/64105/6/pmallolrTFM0617memoria.pdf>
- Masseran, N., & Safari, M. A. M. (2020). Risk assessment of extreme air pollution based on partial duration series: IDF approach. *Stochastic Environmental Research and Risk Assessment*, 34(3), 545-559. Recuperado de <https://doi.org/10.1007/s00477-020-01784-2>
- Miró, J., Caselles, V., & Estrela, M. (2017). Multiple imputation of rainfall missing data in the Iberian Mediterranean context. *Atmospheric Research*, 197, 313-330. Recuperado de <https://ui.adsabs.harvard.edu/abs/2017AtmRe.197..313M/abstract>
- Molenberghs, G., Fitzmaurice, G., Kenward, M. G., Tsiatis, A., & Verbeke, G. (2015). *Handbook of missing data methodology*. London, UK: Chapman & Hall/CRC.
- Naghetini, M. (2017). *Fundamentals of statistical hydrology*. Berlin, Germany: Springer. Recuperado de <https://doi.org/DOI10.1007/978-3-319-43561-9>

- Ng, J. L., Tiang, S. K., Huang, Y. F., Noh, N. I. F. M., & Al-Mansob, R. A. (2021). Analysis of annual maximum and partial duration rainfall series. *IOP Conference Series: Earth and Environmental Science*, 646(1), 012039. Recuperado de <https://doi.org/10.1088/1755-1315/646/1/012039>
- Noor, M., Ismail, T., Chung, E.-S., Shahid, S., & Sung, J. H. (2018). Uncertainty in rainfall intensity duration frequency curves of peninsular Malaysia under changing climate scenarios. *Water*, 10(12). Recuperado de <https://doi.org/10.3390/w10121750>
- Olsson, J., Södling, J., Berg, P., Wern, L., & Eronn, A. (2019). Short-duration rainfall extremes in Sweden: A regional analysis. *Hydrology Research*, 50(3), 945-960. Recuperado de <https://doi.org/10.2166/nh.2019.073>
- OMM, Organización Meteorológica Mundial. (2011). *Guía de prácticas hidrológicas: gestión de recursos hídricos y aplicación de prácticas hidrológicas* (6a. ed.). Vol. II. Ginebra, Suiza: Organización Meteorológica Mundial.
- Sane, Y., Panthou, G., Bodian, A., Vischel, T., Lebel, T., Dacosta, H., Quantin, G., Wilcox, C., Ndiaye, O., Diongue-Niang, A., & Diop Kane, M. (2018). Intensity-duration-frequency (IDF) rainfall curves in Senegal. *Natural Hazards and Earth System Sciences*, 18(7), 1849-1866. Recuperado de <https://doi.org/10.5194/nhess-18-1849-2018>

- Singh, V. (2017). *Handbook of applied hydrology* (2<sup>nd</sup> ed. to replace the classic 1963 edition edited by Ven Te Chow). New York, USA: McGraw-Hill Education.
- Soumya, R., Anjitha, U. G., Mohan, S., Adarsh, S., & Gopakumar, R. (2020). Incorporation of non-stationarity in precipitation intensity-duration-frequency curves for Kerala, India. *IOP Conference Series: Earth and Environmental Science*, 491, 012013. Recuperado de <https://doi.org/10.1088/1755-1315/491/1/012013>
- Teegavarapu, R., Salas, J., & Stedinger, J. (2019). *Statistical analysis of hydrologic variables. Methods and Applications*. Reston, USA: American Society of Civil Engineers. Recuperado de <https://doi.org/10.1061/9780784415177>.
- Van Campenhout, J., Houbrechts, G., Peeters, A., & Petit, F. (2020). Return period of characteristic discharges from the comparison between partial duration and annual series, application to the Walloon Rivers (Belgium). *Water*, 12(3). Recuperado de <https://doi.org/10.3390/w12030792>
- Vrban, S., Wang, Y., McBean Edward, A., Binns, A., & Gharabaghi, B. (2018). Evaluation of stormwater infrastructure design storms developed using partial duration and annual maximum series models. *Journal of Hydrologic Engineering*, 23(12), 04018051. Recuperado de [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001712](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001712)

- Yilmaz, A., & Perera, B. (2014). Extreme rainfall non-stationarity investigation and intensity–frequency–duration relationship. *Journal of Hydrologic Engineering*, 19, 1160-1172. Recuperado de [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0000878](https://doi.org/10.1061/(ASCE)HE.1943-5584.0000878)
- Yong, S. L. S., Ng, J. L., Huang, Y. F., & Ang, C. K. (2021). Assessment of the best probability distribution method in rainfall frequency analysis for a tropical region. *Malaysian Journal of Civil Engineering*, 33(1). Recuperado de <https://doi.org/10.11113/mjce.v33.16253>