

Relleno de datos faltantes en series de datos de precipitación para la ciudad de Guanajuato

Arias Hernández Luz Adriana¹, Calderón Vega Felicitas¹
Ramírez Vargas Fátima Lizeth¹, Gutiérrez Belmán Julián Ulises¹, Guerrero Rangel José Carlos¹, Guerrero Carranza Jessica Mariela¹, Rodríguez Cortes Manuel Alejandro¹, Camarillo Llamas Uriel¹
¹Universidad de Guanajuato. División de Ingenierías. Campus Guanajuato.

Resumen

La precipitación es considerada como parte fundamental dentro del ciclo hidrológico desde el punto de vista de ser la base de cálculos de balances hídricos a partir de los cuales sea posible determinar alertas tempranas en la prevención de fenómenos hidrometeorológicos. Es por ello que contar con información de la precipitación exacta es de suma importancia para numerosas aplicaciones, además de la aplicación en pronósticos de máximas avenidas e infraestructura hidráulica. Por lo anterior en este proyecto se obtuvo interés específico en el relleno de datos faltantes en series de datos de precipitación. La metodología consistió en la delimitación del sitio de interés, seguido de la selección de estaciones climatológicas para después aplicar modelos de regresión lineal múltiple para el relleno de datos, finalmente se aplicaron funciones de probabilidad. Como resultado se llevó a cabo el relleno de datos mediante ajustes de regresiones lineales múltiples para estaciones con datos comprendidos en un periodo de 1977-2017. Posteriormente se aplicaron las funciones de probabilidad mediante el Método de Bell y el Método de Chen para la obtención de curvas de precipitación-Intensidad-duración-periodo de retorno.

Palabras clave: Balance Hídricos, Regresión lineal, Modelos hidrológicos, Precipitación.

1. Introducción

El balance hídrico es un principio que se basa en la variación de volumen que hay dentro de un sistema, es decir, el volumen que entra y sale en un determinado periodo de tiempo (Del Toro et al, 2014). Con el objetivo de determinar con precisión la disponibilidad de los recursos hídricos en una determinada zona y los cambios en ellos bajo la influencia de la actividad humana y el cambio climático.

La precipitación origina la fase terrestre del ciclo hidrológico, la cual se manifiesta como una variable aleatoria en función del espacio y del tiempo (Ochoa et al, 2018). Se considera como la variable principal en los estudios hidrológicos, ya que es el punto de partida para el cálculo de balances hídricos y la generación de alertas tempranas por riesgo de sequía en la región (Luna et al, 2015). Estos datos tienen que ser de primera mano, recopilados de una manera especial por estaciones pluviométricas, con años de historia. En ocasiones, en los registros históricos suelen presentarse huecos de información los cuales pueden ser consecuencia de la ausencia de lectura, falla del instrumento de medición, etc., lo que limita su análisis correcto.

Existen métodos para efectuar el completado de valores faltantes en una serie pluviométrica los cuales han sido empleados por autores en diversos estudios donde se hace mención sobre el uso de la guía de prácticas climatológicas de la Organización Meteorológica Mundial, WMO por sus siglas en inglés, en la que se proponen métodos estadísticos

para el relleno de valores faltantes, como regresión lineal, de la razón y la razón normal (Herrera et al, 2017). Estos modelos son los más utilizados en hidrología para distintos propósitos y se usan incluso con diferentes formas de transformación de las variables, pero manteniendo la relación lineal en los parámetros del modelo. Ejemplos de usos de estos modelos en hidrología se dan en aspectos como: predicción hidrológica, relleno y extensión de datos faltantes, modelamiento hidrológico, análisis de relaciones entre variables; entre otras (Luna et al, 2015).

2. Objetivo

Obtener una base de datos de precipitación diaria 24 h para la zona de estudio a través del relleno de datos faltantes diarios mediante el uso de los métodos de regresión lineal múltiple.

3. Metodología

3.1 Delimitación de la zona de estudio

3.1.1 Delimitación con QGIS

QGIS es un Sistema de Información Geográfica (SIG) de Código Abierto. Soporta numerosos formatos y funcionalidades de datos vector, datos ráster y bases de datos. Dispone de un conjunto de herramientas de hidrología para delimitar automáticamente una cuenca hidrográfica usando las herramientas de GRASS. (Acolita, 2020). El proceso es como se marca en la Figura 1.

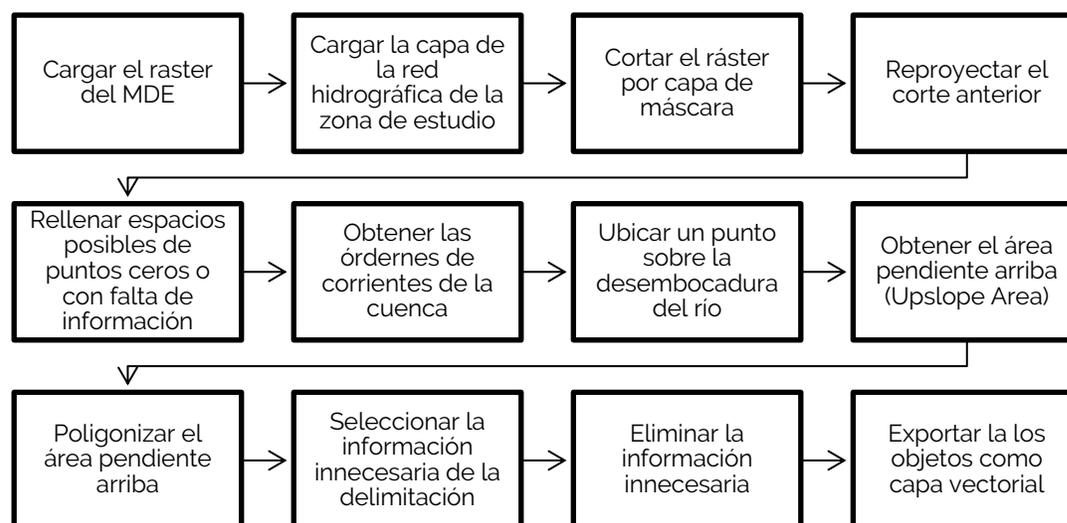


Figura 1. Proceso de delimitación de una cuenca en QGIS

3.1.2 Delimitación con SIATL

El Simulador de Flujos de Agua de Cuencas Hidrográficas (SIATL) es una aplicación geoespacial web gratuita, diseñada para el estudio de cuencas, simulación de flujos de agua, delimitación de cuencas, obtención de la red hidrográfica. Además, integra otros datos, en capas, como cuerpos de agua, curvas de nivel, ortofotos, caminos, localidades, entre otros (INEGI).

3.2 Selección de estaciones

Con el objetivo de identificar las estaciones que serán analizadas, se utilizó una serie de criterios para la selección adecuada de las estaciones, los cuales fueron:

- *La estación se encuentra dentro de la zona de estudio.*
- *La serie histórica cuenta con registros del período 1977-2017.*
- *Los registros de la serie histórica cuentan con al menos 5 años de datos faltantes.*

A partir de los criterios mencionados, se seleccionaron las estaciones mediante Google Earth Pro.

3.3 Relleno de datos faltantes

3.3.1 Regresión lineal múltiple

El método de regresión lineal resulta ser muy útil debido a su eficacia en la estimación de datos faltantes. Se recomienda para la estimación de datos mensuales y anuales de la estación de estudio. Para ello se requiere establecer una regresión y correlación lineal entre una estación patrón y la que tenga carencia de datos, mediante una ecuación lineal.

Con el complemento “Regresión” en Excel, se analizaron las estaciones climatológicas seleccionadas. A partir de los resultados obtenidos, se verifican los resultados con el siguiente criterio arrojados mediante la Tabla de ANOVA:

- a. *El valor de $R^2 \geq 0.90$*

Si el criterio anterior no se cumple, se procede a eliminar estaciones considerando el siguiente criterio:

- b. *En el apartado de “probabilidad”, toda aquella variable x con valor ≤ 0.05 , será descartada del análisis. Esperando así, una mejora en el valor de R^2 .*

Una vez que se cumple el criterio *a* en todas las estaciones, proceder a realizar el relleno de los datos faltantes en las series históricas de cada estación, a partir de la *Ecuación 1*.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \epsilon \quad (1)$$

$$y = \beta_0 + \sum_{i=1} \beta_i x_i + \epsilon_i \quad (2)$$

Donde β_0 es la intercepción; β_1 es el valor obtenido en la regresión; x_1 es el valor de la variable de cada estación.

3.3.2 Aplicación de funciones de distribución de probabilidad

Para el análisis con los métodos probabilísticos, es necesario seleccionar las precipitaciones máximas en cada año de cada una de las estaciones. Además, calcular la desviación estándar, media aritmética del número total de datos de cada estación. Para este caso fue considerada la Distribución Gumbel (Aparicio, 1998) para la cual para N muestras a partir de las cuales conociendo el número total de datos (n) en donde a partir de la función de distribución de probabilidad Ecuación 3 se observa que x tiende a:

$$F(x) = e^{-e^{-\alpha(x-\beta)}} \quad (3)$$

donde la función de densidad de probabilidad es entonces como se muestra en la Ecuación 4.

$$f(x) = \alpha e^{-\alpha(x-\beta)-e^{-\alpha(x-\beta)}} \quad (4)$$

Donde α y β son los parámetros de la función.

Tabla 1. Valores de σ_y y μ_y de acuerdo con el número de datos que se cuentan (Aparicio, 1998)

n	μ_y	σ_y
10	0.4952	0.9496
15	0.5128	0.0206
20	0.5236	1.0628
25	0.5309	1.0914
30	0.5362	1.1124
35	0.5403	1.1285
40	0.5436	1.1413
45	0.5463	1.1518
50	0.5485	1.1607
55	0.5504	1.1682
60	0.5521	1.1747
65	0.5535	1.1803
70	0.5548	1.1854
75	0.5559	1.1854
80	0.5569	1.1898
85	0.5578	1.1938
90	0.5586	1.1974
95	0.5593	1.2007
100	0.5600	1.2065

Con los valores de σ_y y μ_y , obtenidos a partir de la Tabla 1 se puede calcular el valor de α y β con las Ecuaciones 5 y 6, respectivamente.

$$\alpha = \frac{\sigma_y}{S} \quad (5)$$

Donde S es la varianza.

$$\beta = \bar{x} - \frac{\mu_y}{\alpha} \quad (6)$$

Donde \bar{x} es la media aritmética.

Con un periodo de retorno (T_r) de 2 años, se calcula el valor de la precipitación para 24 horas en un periodo de 2 años mediante la Ecuación 7.

$$P_{1440}^2 = \beta - \frac{1}{\alpha} \ln \left[\ln \left(\frac{T_r}{T_r - 1} \right) \right] \quad (7)$$

3.3.2.1 Método de Bell

Bell presentó una ecuación (Ecuación 8) tomando en cuenta la precipitación de duración de una hora y período de retorno de 2 años (Campos, 2010). De tal manera que el objetivo de este método es calcular el valor de la intensidad máxima de lluvia (i) en función de su duración (d) y del periodo de retorno (T_r), realizando un ajuste simultaneo de las tres variables (I - D - T_r) por medio de una regresión. Este es el método más común que permite realizar este proceso por las funciones matemáticas mencionadas anteriormente. Conociendo las elevaciones de cada una de las estaciones, se procede a calcular el coeficiente R con la Ecuación 8.

$$R = -0.000000009Elev^2 + 0.0002Elev + 0.3073 \quad (8)$$

Donde $Elev$ es la elevación a la que se encuentra la estación en msnm.

Una vez que conocemos el valor del coeficiente R y de la precipitación en T_r de 2 años en 24 horas (P_{1440}^2), podemos obtener (P_{60}^2) a partir de la Ecuación 9.

$$P_{60}^2 = R * (P_{1440}^2) \quad (9)$$

En donde el valor de P_{60}^2 obtenido de la ecuación 9 se sustituye en la Ecuación de Bell (Ecuación 10) expresada a continuación:

$$P_t^{Tr} = [0.35 * \ln(T_r) + 0.76] * (0.54 t^{0.25} - 0.5) * P_{60}^2 \quad (10)$$

3.3.2.2 Método de Chen

Chen propone una ecuación (Ecuación 11) con la cual se puede calcular la precipitación para cualquier duración y período de retorno. (Campos, 2010)

Ya con lo antes mencionado, el objetivo es el mismo que el método de Bell a diferencia de que en algunas partes del proceso cambia la metodología y la ecuación, pero se llega al mismo resultado.

$$P_t^{Tr} = \frac{a * P_{60}^2 * [\log(10^{2-F})(T_r^{F-1})] * t}{60(t+b)^c} \quad (11)$$

4. Resultados

4.1 Delimitación de la cuenca y selección de estaciones

Se obtuvieron las delimitaciones resultantes de cada uno de los softwares, presentando características muy similares, con un área de 492.04 Km² para la cuenca delimitada con QGIS y 488.82 Km² para la delimitación mediante SIATL. Si bien la obtención de la delimitación a través de SIATL es fácil y sencilla y proporciona una red hidrográfica completa, se seleccionó la delimitación obtenida con QGIS ya que, además de ser un software libre permite mayor manejo de información y procesamiento de datos.

La cuenca delimitada se presenta en la Figura 2. A través de la delimitación obtenida, se seleccionaron las estaciones ubicadas tanto dentro de la cuenca como en un buffer aplicado de 5 Km alrededor de la de cuenca, posteriormente se aplicaron los criterios de selección mencionados en la metodología, obteniendo un total de 7 estaciones comprendidas en un periodo de 1977-2017, mostradas también en la Figura 2, las cuales fueron las siguientes: 11070, 11103, 11024, 11007, 11121, 11124 y 11037.

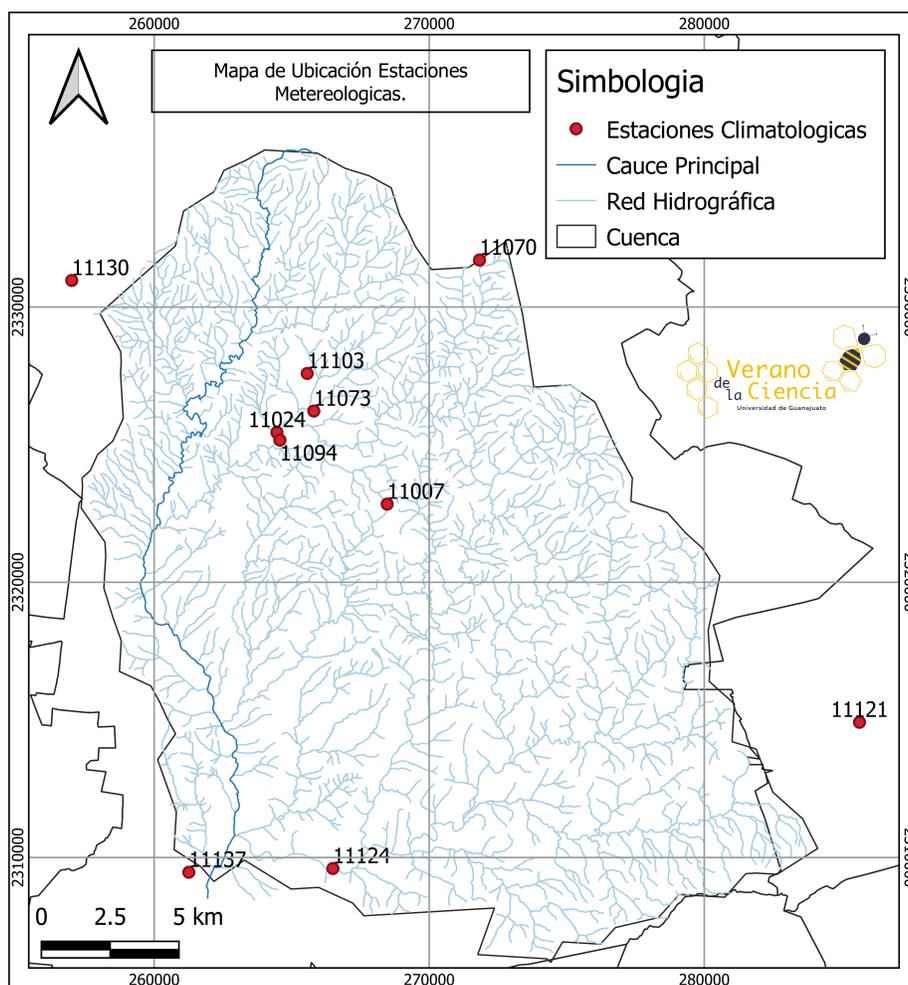


Figura 2. Delimitación de la zona de estudio obtenida con QGIS

4.2 Relleno de datos faltantes

La aplicación de los métodos de regresión lineal ha sido utilizada en diversas investigaciones en el relleno de datos faltantes (Herrera et al, 2017). Para este proyecto mediante el empleo de la Ecuación 1 y a través del ajuste del modelo de regresión para el análisis de los datos estadísticos de los coeficientes de regresión y probabilidad cercana al valor 0.05 de la tabla de ANOVA de la regresión, permitió obtener el relleno de datos faltantes en la serie histórica de las 7 estaciones seleccionadas, el cual se presenta en la Tabla 2.

Tabla 3. Resultados del relleno de datos faltantes

AÑO	PRECIPITACIONES MÁXIMAS						
	ESTACIÓN						
	11007	11024	11070	11103	11121	11124	11137
1977	78.0	94.7	62.4	103.6	61.0	88.6	54.6
1978	63.0	75.9	55.4	52.0	43.0	67.0	68.8
1979	72.0	48.8	61.2	64.0	55.0	29.5	27.0
1980	63.0	56.3	53.0	52.0	49.0	46.0	44.0
1981	60.0	54.8	36.7	62.0	60.0	58.7	51.0
1982	43.5	90.0	48.8	35.0	29.0	47.6	138.1
1983	76.0	60.0	61.5	62.8	62.5	65.3	63.7
1984	35.6	43.0	58.0	53.3	58.0	53.5	49.4
1985	50.2	52.0	79.0	63.0	51.0	33.5	38.0
1986	50.0	48.0	45.3	49.5	64.9	85.0	44.5
1987	58.2	52.0	34.3	83.6	45.0	47.0	54.3
1988	60.4	49.0	60.6	48.5	42.0	45.0	64.0
1989	67.0	63.0	61.7	61.3	32.6	50.5	67.4
1990	57.0	63.0	63.0	48.0	33.9	58.5	38.3
1991	59.9	86.0	72.5	82.0	33.8	51.0	36.6
1992	80.0	79.0	49.8	88.8	30.2	37.5	48.6
1993	85.0	81.0	92.9	83.0	16.4	44.0	54.0
1994	38.0	62.0	53.5	52.6	29.5	66.0	70.3
1995	58.5	69.0	69.5	75.5	29.3	38.5	46.2
1996	66.0	53.0	43.2	61.7	63.0	42.5	24.3
1997	50.0	52.6	43.4	45.9	45.0	79.0	64.4
1998	48.0	45.0	74.2	62.9	30.0	46.5	55.4
1999	63.0	77.0	56.0	32.7	45.0	71.0	69.5
2000	82.0	67.9	51.2	81.2	42.0	24.0	34.0
2001	63.0	72.6	84.3	55.6	33.0	69.5	41.8
2002	49.0	52.2	51.1	49.3	38.0	46.5	42.5
2003	60.0	111.8	82.6	67.6	52.0	57.0	70.5
2004	73.0	50.6	54.9	73.5	50.0	58.5	52.0
2005	33.0	56.6	34.5	58.3	44.0	49.5	36.7
2006	82.0	40.4	55.0	81.2	57.0	39.0	40.3
2007	48.5	64.5	78.9	79.5	42.0	80.0	70.6
2008	49.0	76.9	38.3	69.0	90.0	81.0	73.2
2009	42.5	45.6	53.9	54.4	51.0	75.0	65.6
2010	90.0	95.6	85.0	93.5	66.0	98.0	84.7
2011	84.0	52.6	37.4	34.6	35.0	100.0	76.0
2012	53.0	51.3	48.8	61.8	67.0	38.0	81.7
2013	73.5	46.3	49.2	38.5	89.0	40.0	52.4
2014	53.0	68.8	50.4	46.5	55.0	63.5	45.6
2015	65.5	60.0	88.2	95.0	55.0	60.0	58.0
2016	65.0	55.0	35.7	46.2	39.0	49.0	40.0
2017	60.0	39.3	56.6	40.0	142.0	51.0	72.7

4.3 Aplicación de funciones de distribución de probabilidad

A continuación, en la Figura 3, se presentan gráficas de las curvas de Lámina de Precipitación-Duración-Período de retorno y las curvas de Intensidad-Duración-Periodo (I-D-Tr) de retorno (5, 10, 25, 50, 100, 200, 500 y 1000 años) de la estación 11007, obtenidas con el método de Bell.

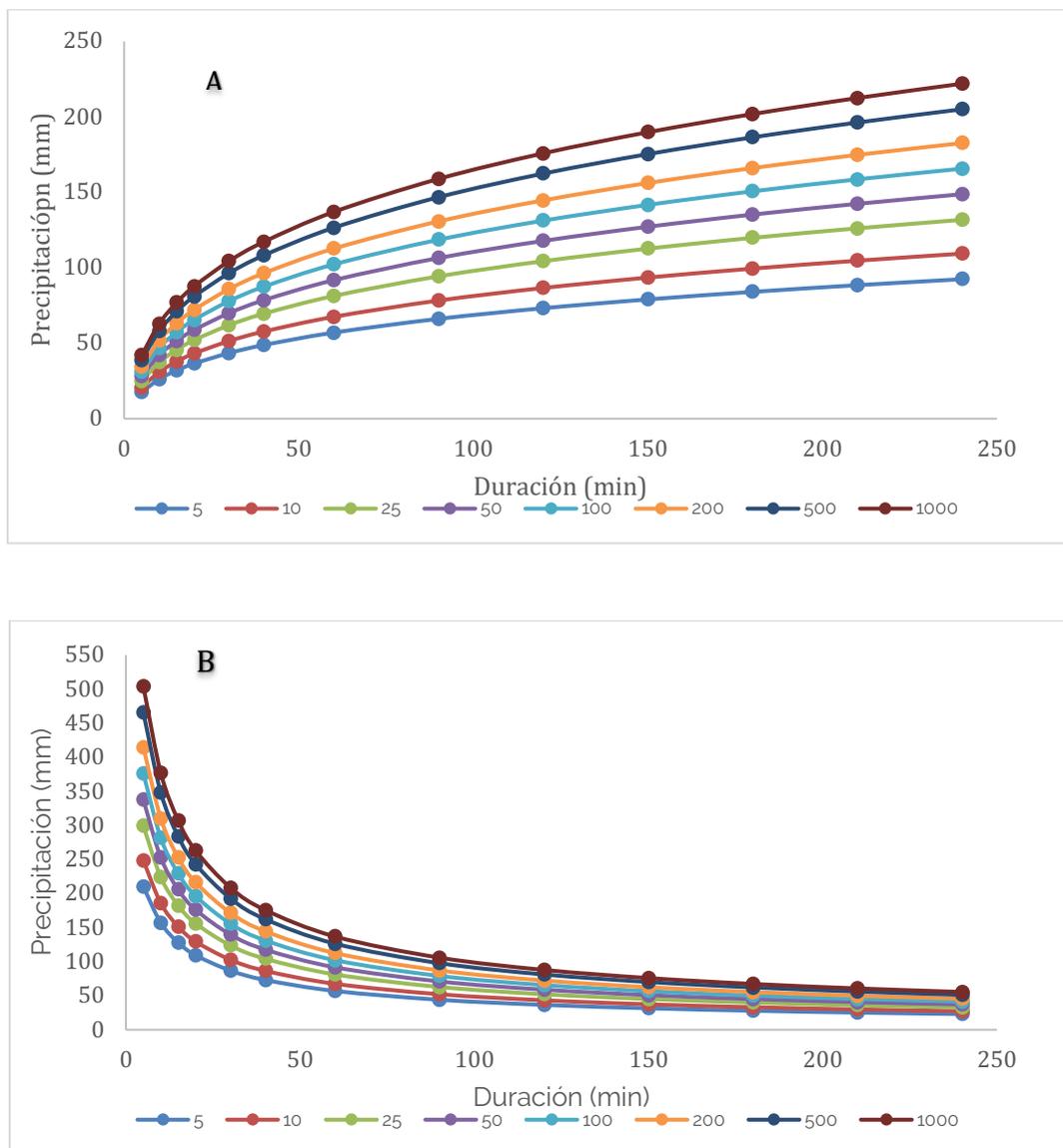


Figura 3. A)Curvas de Precipitación-Duración-Periodo . **B)** Curvas I-D-Tr. (Método de Bell, estación 11007)

Cabe mencionar que se presenta, a manera de ejemplo, sólo para una estación. Sin embargo, para las demás estaciones se observa el mismo comportamiento de las curvas del resto de las estaciones en comparación con las presentadas en la Figura 3. A) y B). De igual manera se presentan gráficas de las curvas de Precipitación-Duración-Periodo de retorno y las curvas de I-D-Tr de la estación 11007, obtenidas con el método de Chen (Figura 4).

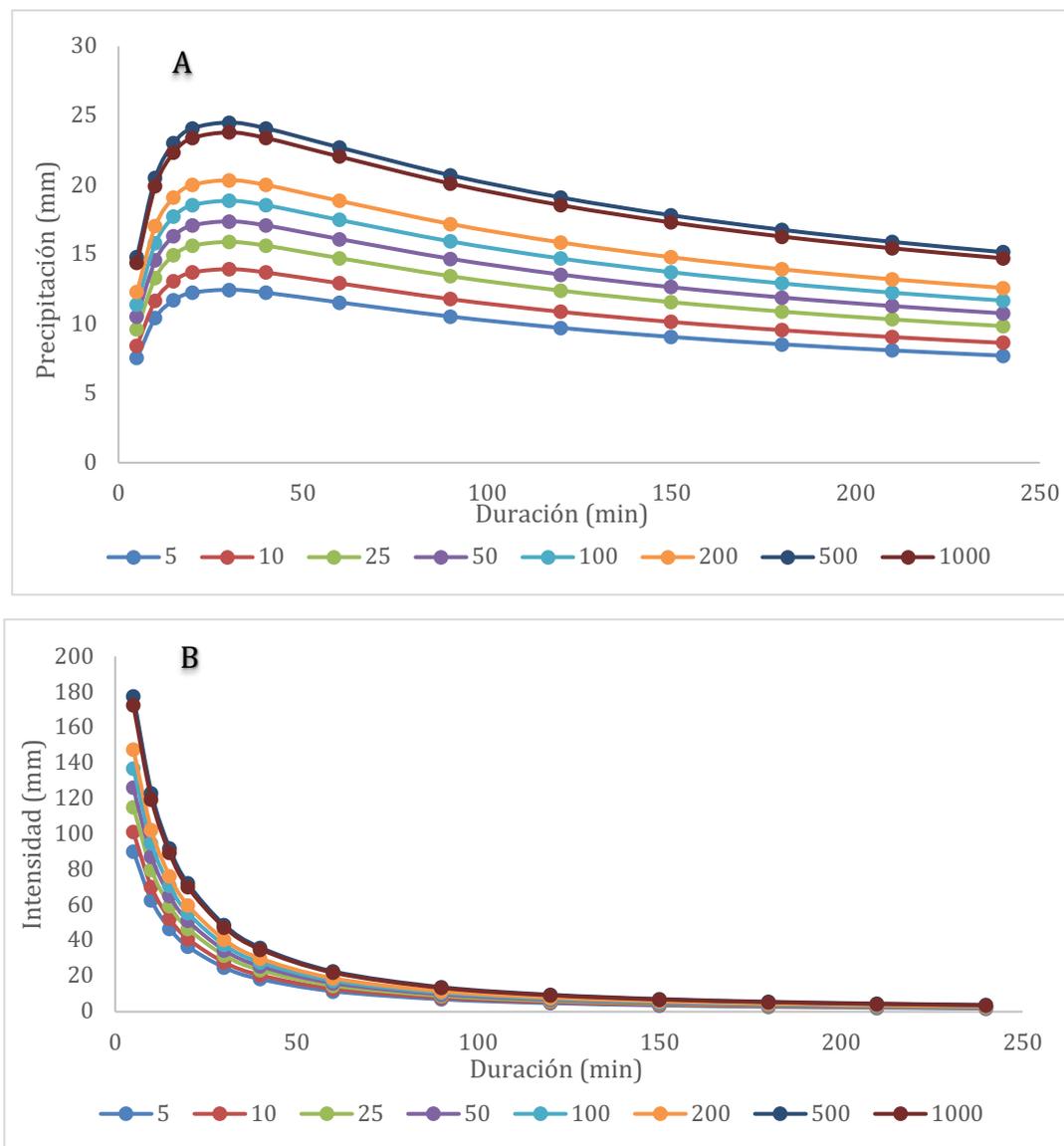


Figura 4. A) Curvas de Precipitación-Duración-Periodo de Retorno. B) Curvas I-D-Tr. (Método de Chen, estación 11007)

5. Conclusión

Mediante la delimitación de la Cuenca a través de los softwares empleados se estableció la selección de estaciones en la zona de estudio de interés.

Se estimaron los datos faltantes en los registros históricos de la precipitación durante 24 horas de 7 estaciones climatológicas comprendidas en la zona de estudio para un periodo comprendido entre 1977-2017 mediante un análisis de regresión lineal múltiple.

Se obtuvieron las curvas Precipitación-Duración-Periodo de retorno e Intensidad-Duración-Periodo de retorno aplicando los métodos de distribución probabilística de Bell y Chen.

Los resultados obtenidos en el presente trabajo proporcionan una base de datos de precipitación máxima de 24 horas completa, homogénea para el periodo de análisis la cual puede ser usada para complementar estudios hidrológicos.

Referencias

Del Toro Guerrero, F. J. & Kretschmar, T. & Hinojosa C. A. (2014) Estimación del balance hídrico en una cuenca semiárida, El Mogor, Baja California, México. *Tecnología en Ciencias del Agua*. 5(6), 69 – 81.

Acolita. (27 de enero de 2020). Obtenido de <https://acolina.com/delimitar-automaticamente-una-cuenca-hidrografica-en-qgis-3/>.
Carrera, D. & Guevara, P.& Tamayo, L., Balarezo, A.& Narváez, C., & Morocho, D. (2016) Relleno de series anuales de datos meteorológicos mediante métodos estadísticos en la zona costera e interandina del Ecuador, y cálculo de la precipitación media. *Idesia*, 81-90.

Luna, E., & Lavado, W. (2015). Evaluación de métodos hidrológicos para la completación de datos faltantes de precipitación en estaciones de la cuenca Jetepeque, Perú. *Revista Tecnológica ESPOL – RTE*, 42-52.

Herrera Oliva, C. S. & Campos Gaytán, J. R.& Carrillo González F.M. (2017) Estimación de datos faltantes de precipitación por el método de regresión lineal: Caso de estudio Cuenca Guadalupe, Baja California, México. *Investigación y Ciencia*. 25(71). 34 – 44.

Ochoa, C., C.G., N. B., Zimmermann, E., Vives, L., Blanco, M., & Cazenave, G. (2018). Relleno de series de precipitación diaria para largos períodos de tiempo en zonas de llanura. Caso de estudio Cuenca Superior del Arroyo del Azul. GEOACTA. INEGI Obtenido de https://antares.inegi.org.mx/analisis/red_hidro/siat/

Campos Aranda, D. F. (2010) *Introducción a la hidrología urbana*. (1ª ed) San Luis Potosí, México: Printego, Pedro Moreno.

Aparicio Mijares, F. J. (1992) *Fundamentos de Hidrología de superficie*. D.F. México: Limusa.