

ANÁLISIS HIDROLÓGICO CON `PYTHON` MEDIANTE ESTACIONES CLIMATOLÓGICAS EN EL ESTADO DE SINALOA

Christian Ramos Castillo

14 de diciembre de 2023

Resumen

El *análisis hidrológico* en proyectos de construcción en México desempeña un papel esencial al garantizar no solo la seguridad y sostenibilidad, sino también el cumplimiento normativo. En un entorno donde los eventos climáticos extremos y la gestión eficiente del agua son factores críticos para el desarrollo y el bienestar de las comunidades, la aplicación de prácticas rigurosas se vuelve imperativa.

Para llevar a cabo este análisis, se recurrió a datos provenientes de estaciones climatológicas proporcionados por el Servicio Meteorológico Nacional (SMN) de la Comisión Nacional del Agua (CONAGUA) y el Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE). Se seleccionaron cuatro estaciones específicas: *Bocatoma Sufragio*, *Surutato*, *Yecorato* y *Jaina* de las cuales solo se trabajará en este documento con ***Jaina***. Esta estación ofreció información valiosa sobre las precipitaciones diarias anuales durante el periodo en que estuvieron operativas, sirviendo como base para entender la variabilidad hidrológica en la región.

Adicionalmente, se emplearon *modelos estadísticos inferenciales* para profundizar en el análisis. La Distribución Gumbel, especializada en modelar la distribución del máximo (o mínimo), se utilizó para abordar la complejidad de las lluvias máximas anuales diarias. Este enfoque fue complementado con la comparación frente a la Distribución Log Normal y el Modelo Random Forest Regression. La combinación de estos modelos estadísticos no solo buscó ajustarse a los datos existentes, sino también prever eventos futuros mediante la reducción del error cuadrático.

La proyección de eventos futuros se realizó para periodos acordes con la longitud de datos históricos de acuerdo a la estación, brindando una visión integral de las posibles *condiciones hidrológicas* en el futuro. Estos modelos fueron aplicados considerando las características específicas de la cuenca de estudio. La meta principal era predecir las precipitaciones máximas diarias anuales durante el tiempo de diseño del proyecto. Esta aproximación avanzada proporciona una mayor seguridad en el diseño de cualquier estructura en el sitio del proyecto, considerando no solo las condiciones actuales sino también las potenciales variaciones climáticas a lo largo del tiempo.

El análisis hidrológico no se limitó a la recopilación de datos climatológicos, sino que se enriqueció con la aplicación de modelos estadísticos avanzados. Esta combinación permitió una comprensión más profunda de la hidrología del área de estudio y proporcionó herramientas efectivas para la toma de decisiones en el diseño de infraestructuras, considerando tanto la variabilidad histórica como las proyecciones futuras. En este documento encontrarás toda la información recopilada, el código en Python, y finalmente la conclusión a la que se llegó con el *análisis de datos*.

Índice

1. GENERACIÓN DE LAS SERIES DE PRECIPITACIÓN MÁXIMA ANUAL EN 24H	5
1.1. Limpieza y ordenamiento de datos	5
2. DESCRIPCIÓN ESTADÍSTICA DE LAS SERIES	7
2.1. Tabla de frecuencias	7
3. CURVA DE FRECUENCIA HP MÁXIMA ANUAL EN 24H	9
3.1. Método de Weibull para estimar Eventos Hidroclimatológicos Extremos	9
4. SELECCIONANDO EL MODELO QUE MEJOR SE AJUSTE A LOS DATOS	11
4.1. Método Gumbell	11
4.2. Método LogNormal	14
4.3. Método de Regresión con RandomForest (Machine Learning)	16
5. ¿CUÁL ES EL MODELO QUE MEJOR SE AJUSTA?	18
5.1. Comparación de resultados	18
6. CÓDIGO	19
Referencias	20

Índice de tablas

1.	Frecuencias de datos / Jaina	8
2.	Método Weibull/Estación Jaina	10
3.	Parametros para ajuste/ Jaina	11
4.	Datos Observados vs Calculados / Jaina	12
5.	Método Lognormal/Jaina	15
6.	Método Regresión con RandomForest/Jaina	17
7.	Errores Cuadráticos Mínimos (ECM)/Jaina	18

Índice de figuras

1.	Datos origen / Jaina	5
2.	Precipitaciones máximas diarias mensuales anuales / Jaina	5
3.	Precipitación máxima anual en 24h / Jaina	6
4.	Precipitación máxima mensual histórica / Jaina	7
5.	Distribución Relativa / Jaina	8
6.	Distribución Acumulada / Jaina	8
7.	Sesgo a la derecha / Jaina	8
8.	Curva de frecuencia máxima anual 24h / Jaina	9
9.	Análisis de frecuencias Gumbel / Jaina	13
10.	Análisis de frecuencias Lognormal / Jaina	14
11.	Análisis de frecuencias con Regresión Random Forest / Jaina	16
12.	Comparación de resultados en el análisis de frecuencias / Jaina	19

1 GENERACIÓN DE LAS SERIES DE PRECIPITACIÓN MÁXIMA ANUAL EN 24H

1.1 Limpieza y ordenamiento de datos

En el proceso de *análisis de datos*, previo a cualquier evaluación, es imperativo realizar la manipulación, procesamiento, limpieza y desglose de los mismos. Comúnmente, al trabajar con información, se recurre a herramientas de paquetería, siendo MS *Excel*, una opción frecuente. No obstante, su uso conlleva limitaciones para la manipulación de datos, además de consumir recursos significativos, especialmente cuando lidiamos con conjuntos de datos extensos. Para superar estas limitaciones, optaremos por utilizar *Python* específicamente utilizando el entorno llamado *Jupyter Notebook*, una herramienta que ofrece una amplia variedad de bibliotecas en constante actualización, demostrando ser altamente eficiente en comparación con otras alternativas. Este enfoque nos permitirá llevar a cabo operaciones más complejas y optimizar el manejo de grandes volúmenes de datos, asegurando así un proceso más eficaz y adaptado a las demandas de análisis modernas.

La información empleada proviene del *Servicio Meteorológico Nacional (SMN)* y del *Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE)* utilizando una estación climatológica llamada *Jaina* con *Clave 25046* específicamente del estado de Sinaloa. En la **Figura 1** se presenta la estructura de los datos originales, los cuales reflejan la precipitación diaria a lo largo del año de registro. La longitud de los datos varía según los años en que estuvo operativa cada estación. En consecuencia, los periodos anuales para estos datos difieren entre las cuatro estaciones climatológicas estudiadas.

	Mes	Día	Año	Datos
0	1	1	1942	0.0
1	1	2	1942	0.7
2	1	3	1942	0.0
3	1	4	1942	0.0
4	1	5	1942	0.0
...
27083	2	25	2016	0.0
27084	2	26	2016	0.0
27085	2	27	2016	0.0
27086	2	28	2016	0.0
27087	2	29	2016	0.0

Figura 1: Datos origen / Jaina

	Año	1	2	3	4	5	6	7	8	9	10	11	12
0	1942	0.7	16.5	0.3	0.0	0.0	0.9	44.2	59.2	47.8	119.4	62.0	15.0
1	1943	2.0	0.0	0.0	0.0	0.0	59.7	33.6	68.6	239.4	159.0	12.8	219.0
2	1944	17.0	24.0	9.4	0.0	0.0	0.8	44.8	47.6	56.0	16.2	35.8	8.5
3	1946	8.0	0.0	2.0	0.0	0.0	51.5	28.5	61.7	41.2	31.5	0.0	19.0
4	1947	8.2	3.8	12.5	0.0	20.0	32.5	89.0	83.5	57.5	2.0	6.2	44.6
...
68	2011	0.0	0.0	0.0	0.0	NaN	7.5	110.0	80.0	64.0	12.0	31.5	12.0
69	2012	0.0	1.5	0.0	0.0	NaN	29.0	71.5	93.0	60.0	21.5	0.0	41.5
70	2013	7.0	0.0	0.0	0.0	NaN	3.0	21.0	60.0	105.0	2.0	44.0	30.5
71	2014	0.0	0.0	3.5	0.0	NaN	35.0	64.5	106.0	37.0	16.0	NaN	0.0
72	2015	7.5	12.5	13.0	3.5	NaN	62.0	59.0	90.5	93.5	24.5	19.5	1.5

Figura 2: Precipitaciones máximas diarias mensuales anuales / Jaina

En la **Figura 2** se construye una tabla pivote o *tabla dinámica* ordenada por año y mes, tomando la precipitación máxima por año, además, se eliminan años en los que hubo ≥ 9 meses de datos *Null* o *NaN*

(datos faltantes). Podemos comparar la **Figura 1** y **Figura 2** con un orden que ahora es mas claro y manipulable.

Finalmente después de limpiar y organizar los datos con los que vamos a trabajar obtenemos la **Figura 3**, esta se conforma por filtrar los datos máximos históricos de precipitación en *mm* de cada año, teniendo así un registro de **73 años**.

Año		Hpmax (mm)
0	1942	119.4
1	1943	239.4
2	1944	56.0
3	1946	61.7
4	1947	89.0
...
68	2011	110.0
69	2012	93.0
70	2013	105.0
71	2014	106.0
72	2015	93.5

Figura 3: Precipitación máxima anual en 24h / Jaina

2 DESCRIPCIÓN ESTADÍSTICA DE LAS SERIES

2.1 Tabla de frecuencias

El objetivo principal de ésta sección es describir las características fundamentales de un conjunto de datos, sin realizar inferencias más allá de la muestra recopilada. La estadística descriptiva proporciona herramientas y técnicas que ayudan a los investigadores, analistas y profesionales a entender la estructura de los datos y comunicar de manera efectiva la información a otras personas.

Para comenzar con la descripción estadística comenzaremos tomando en cuenta el máximo mensual histórico durante los *73 años* analizados, en la **Figura 4** se muestran cuáles fueron en cada uno de los 12 meses.

Mes Datos		
0	1	84.0
1	2	72.0
2	3	33.7
3	4	52.0
4	5	47.8
5	6	126.0
6	7	170.0
7	8	141.8
8	9	248.6
9	10	159.0
10	11	144.0
11	12	219.0

Figura 4: Precipitación máxima mensual histórica / Jaina

Definiendo el número de clase = *12*, comenzamos con la estadística descriptiva que se muestra en la **Tabla 1** de frecuencias y mostrando que la **Figura 7** se encuentra *sesgada hacia la derecha*, y mostrando la distribución acumulada en la **Figura 6**.

Tabla 1: Frecuencias de datos / Jaina

N Clase	Eventos	Clases	Frecuencia Relativa	Frecuencia Acumulada
1	17	(57.933, 75.267]	0.233	0.233
2	14	(75.267, 92.6]	0.192	0.425
3	11	(92.6, 109.933]	0.151	0.575
4	11	(109.933, 127.267]	0.151	0.726
5	10	(127.267, 144.6]	0.137	0.863
6	5	(144.6, 161.933]	0.068	0.932
7	2	(161.933, 179.267]	0.027	0.959
8	1	(179.267, 196.6]	0.014	0.973
9	1	(196.6, 213.933]	0.014	0.986
10	1	(213.933, 231.267]	0.014	1.000
11	0		0.000	1.000
12	0		0.000	1.000

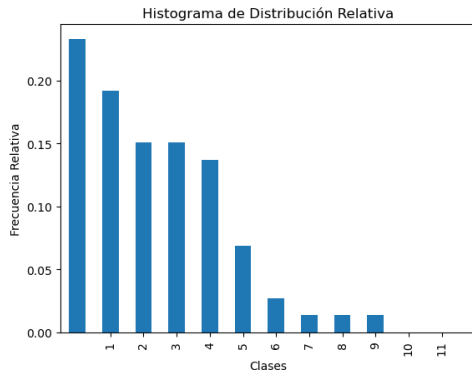


Figura 5: Distribución Relativa / Jaina

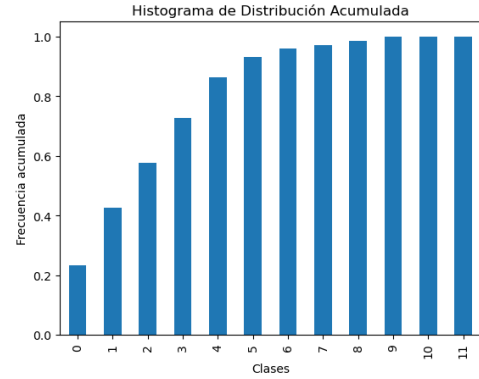


Figura 6: Distribución Acumulada / Jaina

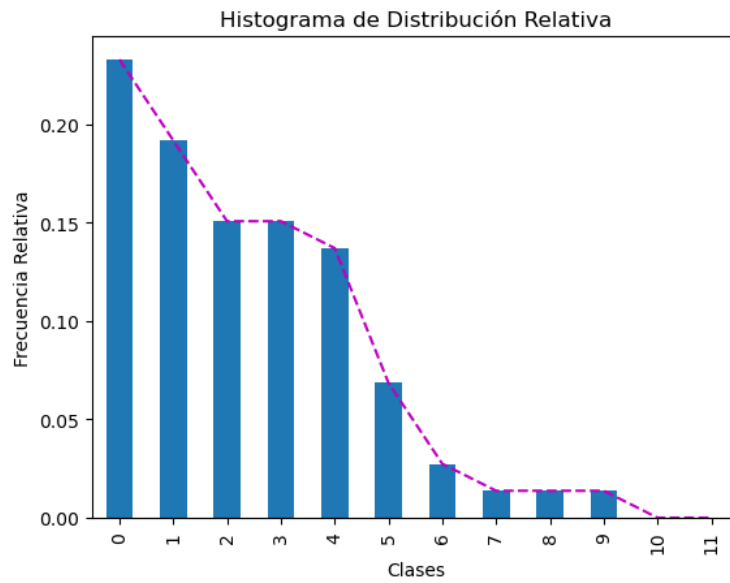


Figura 7: Sesgo a la derecha / Jaina

3 CURVA DE FRECUENCIA HP MÁXIMA ANUAL EN 24H

3.1 Método de Weibull para estimar Eventos Hidroclimatológicos Extremos

(Dodson, 2006) describe que el método de Weibull se aplica en la hidrología para modelar y analizar la distribución de las lluvias, caudales de ríos y otros eventos hidrológicos. Así mismo se utiliza para ajustar las series de datos hidrológicos y estimar parámetros que describen la forma de la distribución. La distribución de Weibull puede ser especialmente útil en hidrología debido a su flexibilidad para adaptarse a diferentes formas de distribución de datos. En el contexto hidrológico, se puede utilizar para modelar eventos extremos como caudales máximos o mínimos, o para describir la duración y la intensidad de las lluvias.

$$Tr = \frac{n+1}{m} \quad (1)$$

$$P(X \geq x) = \frac{1}{Tr} \quad (2)$$

$$F(x) = p(X \leq x) = \frac{Tr - 1}{Tr} \quad (3)$$

Mediante las ecuaciones (1), (2), y (3), se obtuvo como resultado la **Tabla 2**, y así mismo su representación gráfica en la **Figura 8**.

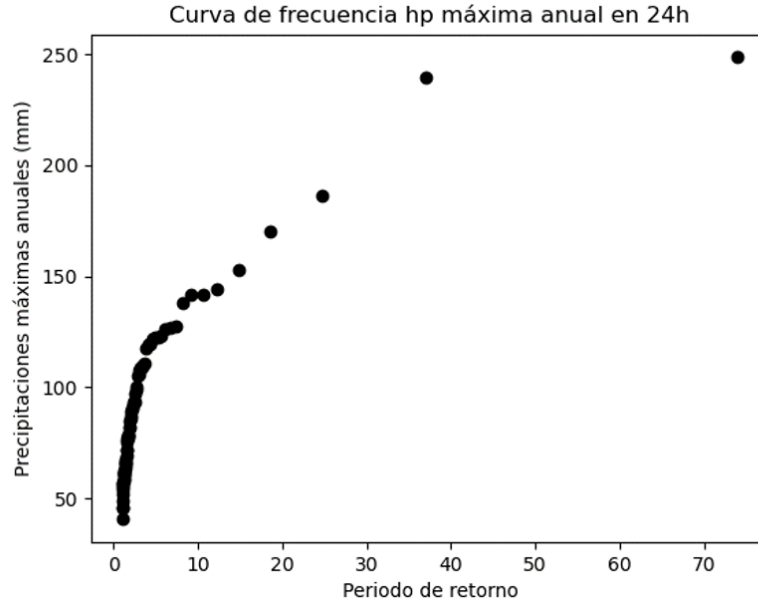


Figura 8: Curva de frecuencia máxima anual 24h / Jaina

Tabla 2: Método Weibull/Estación Jaina

m	Hpmax Ordenada (mm)	Tr	Probabilidad de excedencia	Probabilidad de no excedencia
1	248.600	74.000	0.014	0.986
2	239.400	37.000	0.027	0.973
3	186.500	24.667	0.041	0.959
4	170.000	18.500	0.054	0.946
5	153.000	14.800	0.068	0.932
6	144.000	12.333	0.081	0.919
7	141.900	10.571	0.095	0.905
8	141.800	9.250	0.108	0.892
9	138.000	8.222	0.122	0.878
10	127.800	7.400	0.135	0.865
11	127.000	6.727	0.149	0.851
12	126.000	6.167	0.162	0.838
13	123.000	5.692	0.176	0.824
14	122.400	5.286	0.189	0.811
15	122.400	4.933	0.203	0.797
16	122.000	4.625	0.216	0.784
17	119.400	4.353	0.230	0.770
18	119.300	4.111	0.243	0.757
19	117.600	3.895	0.257	0.743
20	110.800	3.700	0.270	0.730
21	110.000	3.524	0.284	0.716
22	109.000	3.364	0.297	0.703
23	109.000	3.217	0.311	0.689
24	108.000	3.083	0.324	0.676
25	106.000	2.960	0.338	0.662
26	105.000	2.846	0.351	0.649
27	100.500	2.741	0.365	0.635
28	99.000	2.643	0.378	0.622
29	97.300	2.552	0.392	0.608
30	93.500	2.467	0.405	0.595
31	93.400	2.387	0.419	0.581
32	93.000	2.312	0.432	0.568
33	91.100	2.242	0.446	0.554
34	90.500	2.176	0.459	0.541
35	89.000	2.114	0.473	0.527
36	87.300	2.056	0.486	0.514
37	86.000	2.000	0.500	0.500
38	84.500	1.947	0.514	0.486
39	82.600	1.897	0.527	0.473
40	81.800	1.850	0.541	0.459
41	78.900	1.805	0.554	0.446
42	78.400	1.762	0.568	0.432
43	78.000	1.721	0.581	0.419
44	77.200	1.682	0.595	0.405
45	76.900	1.644	0.608	0.392
46	75.800	1.609	0.622	0.378
47	72.000	1.574	0.635	0.365
48	71.800	1.542	0.649	0.351
49	69.500	1.510	0.662	0.338
50	67.500	1.480	0.676	0.324
51	66.200	1.451	0.689	0.311
52	66.000	1.423	0.703	0.297
53	65.800	1.396	0.716	0.284
54	65.000	1.370	0.730	0.270
55	64.000	1.345	0.743	0.257
56	62.600	1.321	0.757	0.243
57	61.700	1.298	0.770	0.230
58	61.200	1.276	0.784	0.216
59	61.000	1.254	0.797	0.203
60	60.800	1.233	0.811	0.189
61	60.300	1.213	0.824	0.176
62	59.000	1.194	0.838	0.162
63	58.100	1.175	0.851	0.149
64	57.100	1.156	0.865	0.135
65	57.000	1.138	0.878	0.122
66	56.000	1.121	0.892	0.108
67	55.000	1.104	0.905	0.095
68	54.000	1.088	0.919	0.081
69	52.000	1.072	0.932	0.068
70	49.200	1.057	0.946	0.054
71	46.000	1.042	0.959	0.041
72	45.500	1.028	0.973	0.027

4 SELECCIONANDO EL MODELO QUE MEJOR SE AJUSTE A LOS DATOS

4.1 Método Gumbell

(Orduña, 2021) señala que en el ámbito de la hidrología, se emplea el método Gumbel para llevar a cabo análisis de valores extremos y calcular cuantiles extremos. Este análisis implica ajustar la distribución de Gumbel a los valores extremos observados en un conjunto de datos hidrológicos, utilizando técnicas estadísticas como el método de máxima verosimilitud para estimar los parámetros de la distribución.

Una vez que la distribución de Gumbel ha sido ajustada, es posible realizar inferencias sobre eventos extremos, como determinar la probabilidad de que ciertos niveles de caudal máximo ocurran en un periodo de retorno específico. Este tipo de análisis resulta fundamental en la planificación hidrológica y en el diseño de infraestructuras destinadas a enfrentar condiciones hidrológicas extremas, como las inundaciones.

Los parámetros que mostramos en la **Tabla 3** se utilizaron para ajustar el método *Gumbel* utilizando las ecuaciones (4),(5), (6), (7), y finalmente el modelo estadístico Gumbel (8),(9) que se muestran en la **Tabla 4**.

$$\alpha = \frac{1.2825}{S} \quad (4)$$

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} \quad (5)$$

$$\beta = \bar{X} - 0.45S \quad (6)$$

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (7)$$

$$F(x) = e^{-e^{-\alpha(x-\beta)}} \quad (8)$$

$$x_T = \beta - \frac{1}{\alpha} \ln \left[\ln \left(\frac{T_r - 1}{T_r} \right) \right] \quad (9)$$

Tabla 3: Parametros para ajuste/ Jaina

Desviación Estándar	Alfa	Beta	Media	Kurtosis	Coficiente de asimetría
40.167	0.032	76.274	94.349	3.350	1.576

Tabla 4: Datos Observados vs Calculados / Jaina

Tr	Hpmax Observada	Hpmax Calculada Gumbel
74.000	248.600	210.861
37.000	239.400	188.937
24.667	186.500	176.020
18.500	170.000	166.790
14.800	153.000	159.579
12.333	144.000	153.643
10.571	141.900	148.586
9.250	141.800	144.173
8.222	138.000	140.250
7.400	127.800	136.713
6.727	127.000	133.488
6.167	126.000	130.519
5.692	123.000	127.765
5.286	122.400	125.194
4.933	122.400	122.780
4.625	122.000	120.501
4.353	119.400	118.342
4.111	119.300	116.287
3.895	117.600	114.325
3.700	110.800	112.446
3.524	110.000	110.641
3.364	109.000	108.902
3.217	109.000	107.224
3.083	108.000	105.601
2.960	106.000	104.027
2.846	105.000	102.499
2.741	100.500	101.011
2.643	99.000	99.562
2.552	97.300	98.146
2.467	93.500	96.762
2.387	93.400	95.407
2.312	93.000	94.078
2.242	91.100	92.773
2.176	90.500	91.490
2.114	89.000	90.227
2.056	87.300	88.982
2.000	86.000	87.753
1.947	84.500	86.539
1.897	82.600	85.338
1.850	81.800	84.148
1.805	78.900	82.968
1.762	78.400	81.797
1.721	78.000	80.633
1.682	77.200	79.474
1.644	76.900	78.320
1.609	75.800	77.168
1.574	72.000	76.018
1.542	71.800	74.867
1.510	69.500	73.714
1.480	67.500	72.557
1.451	66.200	71.395
1.423	66.000	70.226
1.396	65.800	69.047
1.370	65.000	67.857
1.345	64.000	66.653
1.321	62.600	65.431
1.298	61.700	64.190
1.276	61.200	62.925
1.254	61.000	61.632
1.233	60.800	60.307
1.213	60.300	58.943
1.194	59.000	57.534
1.175	58.100	56.070
1.156	57.100	54.542
1.138	57.000	52.936
1.121	56.000	51.232
1.104	55.000	49.406
1.088	54.000	47.423
1.072	52.000	45.229
1.057	49.200	42.737
1.042	46.000	39.792
1.028	45.500	36.062
1.014	40.600	30.562

Con la **Tabla 4**, podemos graficar la comparación entre los *datos observados* y los *calculados* que se muestran en la Figura 9.

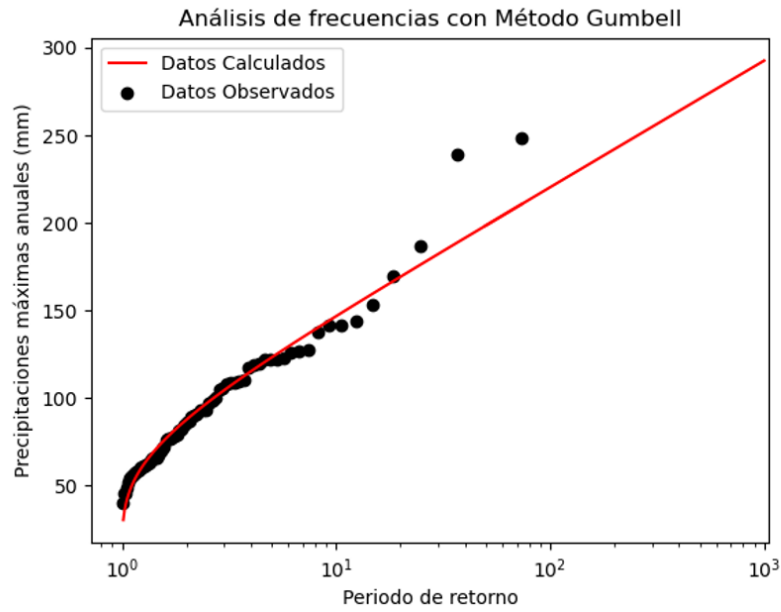


Figura 9: Análisis de frecuencias Gumbel / Jaina

4.2 Método LogNormal

De acuerdo con (Béjar, 2006) la aplicación del método lognormal en hidrología implica ajustar la distribución lognormal a los datos observados utilizando técnicas estadísticas, como el método de máxima verosimilitud. Esto permite estimar los parámetros de la distribución (media y desviación estándar en escala logarítmica) para describir mejor la variabilidad de las variables hidrológicas y facilitar la predicción de eventos futuros.

Las ecuaciones utilizadas son (10), (11), que representa llevar a α y β a escala logarítmica.

$$\alpha = \sum_{i=1}^n \frac{\ln x_i}{n} \quad (10)$$

$$\beta = \left[\sum_{i=1}^n \frac{(\ln x_i - \alpha)^2}{n} \right]^{\frac{1}{2}} \quad (11)$$

Finalmente obtenemos la **Figura 10** que se grafica mediante los datos de la **Tabla 5**.

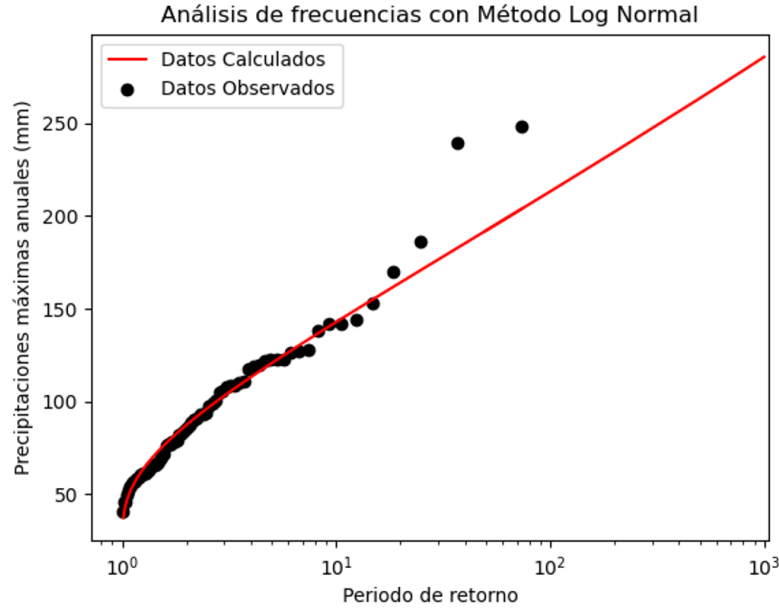


Figura 10: Análisis de frecuencias Lognormal / Jaina

Tabla 5: Método Lognormal/Jaina

Tr	Hpmax Observada	Hpmax Calculada Gumbel	Hpmax Calculada LogNormal
74.000	248.600	210.861	203.915
37.000	239.400	188.937	182.825
24.667	186.500	176.020	170.505
18.500	170.000	166.790	161.736
14.800	153.000	159.579	154.902
12.333	144.000	153.643	149.286
10.571	141.900	148.586	144.507
9.250	141.800	144.173	140.339
8.222	138.000	140.250	136.636
7.400	127.800	136.713	133.299
6.727	127.000	133.488	130.258
6.167	126.000	130.519	127.460
5.692	123.000	127.765	124.865
5.286	122.400	125.194	122.443
4.933	122.400	122.780	120.169
4.625	122.000	120.501	118.025
4.353	119.400	118.342	115.992
4.111	119.300	116.287	114.059
3.895	117.600	114.325	112.214
3.700	110.800	112.446	110.448
3.524	110.000	110.641	108.752
3.364	109.000	108.902	107.119
3.217	109.000	107.224	105.544
3.083	108.000	105.601	104.021
2.960	106.000	104.027	102.545
2.846	105.000	102.499	101.112
2.741	100.500	101.011	99.719
2.643	99.000	99.562	98.361
2.552	97.300	98.146	97.037
2.467	93.500	96.762	95.743
2.387	93.400	95.407	94.477
2.312	93.000	94.078	93.236
2.242	91.100	92.773	92.019
2.176	90.500	91.490	90.823
2.114	89.000	90.227	89.647
2.056	87.300	88.982	88.489
2.000	86.000	87.753	87.347
1.947	84.500	86.539	86.219
1.897	82.600	85.338	85.105
1.850	81.800	84.148	84.003
1.805	78.900	82.968	82.911
1.762	78.400	81.797	81.829
1.721	78.000	80.633	80.754
1.682	77.200	79.474	79.686
1.644	76.900	78.320	78.624
1.609	75.800	77.168	77.565
1.574	72.000	76.018	76.509
1.542	71.800	74.867	75.455
1.510	69.500	73.714	74.401
1.480	67.500	72.557	73.345
1.451	66.200	71.395	72.287
1.423	66.000	70.226	71.224
1.396	65.800	69.047	70.154
1.370	65.000	67.857	69.077
1.345	64.000	66.653	67.990
1.321	62.600	65.431	66.890
1.298	61.700	64.190	65.775
1.276	61.200	62.925	64.643
1.254	61.000	61.632	63.489
1.233	60.800	60.307	62.310
1.213	60.300	58.943	61.101
1.194	59.000	57.534	59.857
1.175	58.100	56.070	58.572
1.156	57.100	54.542	57.235
1.138	57.000	52.936	55.838
1.121	56.000	51.232	54.364
1.104	55.000	49.406	52.796
1.088	54.000	47.423	51.106
1.072	52.000	45.229	49.253
1.057	49.200	42.737	47.172
1.042	46.000	39.792	44.746
1.028	45.500	36.062	41.731
1.014	40.600	30.562	37.415

4.3 Método de Regresión con RandomForest (Machine Learning)

De acuerdo con (Grus, 2023) Random Forest Regression (o RandomForestRegressor) es un algoritmo de aprendizaje supervisado que se utiliza para tareas de regresión en machine learning. Forma parte de la familia de algoritmos conocida como Random Forest, que también incluye Random Forest Classification para problemas de clasificación. El Random Forest es una técnica basada en ensambles que combina múltiples modelos más simples, llamados árboles de decisión, para mejorar el rendimiento predictivo y la robustez del modelo.

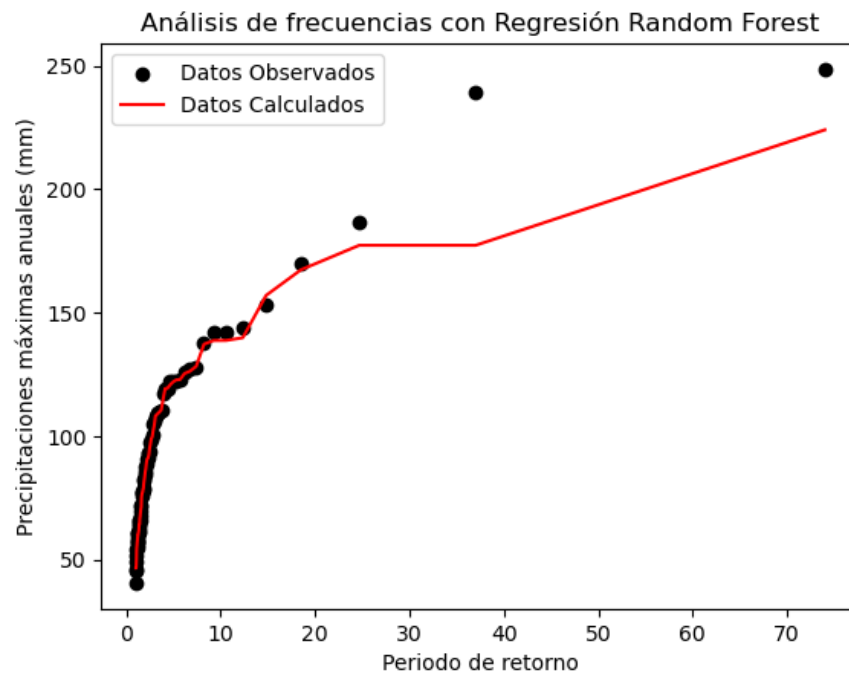


Figura 11: Análisis de frecuencias con Regresión Random Forest / Jaina

Tabla 6: Método Regresión con RandomForest/Jaina

Tr	Hpmax Observada	Hpmax Calculadas Gumbel	Precipitaciones Calculadas LogNormal	Hpmax Calculada RandomForestRegression
74.000	248.600	210.861	203.915	224.070
37.000	239.400	188.937	182.825	177.300
24.667	186.500	176.020	170.505	177.300
18.500	170.000	166.790	161.736	167.400
14.800	153.000	159.579	154.902	157.000
12.333	144.000	153.643	149.286	139.860
10.571	141.900	148.586	144.507	138.790
9.250	141.800	144.173	140.339	138.750
8.222	138.000	140.250	136.636	137.260
7.400	127.800	136.713	133.299	128.280
6.727	127.000	133.488	130.258	126.300
6.167	126.000	130.519	127.460	125.400
5.692	123.000	127.765	124.865	122.960
5.286	122.400	125.194	122.443	122.780
4.933	122.400	122.780	120.169	121.940
4.625	122.000	120.501	118.025	120.730
4.353	119.400	118.342	115.992	119.380
4.111	119.300	116.287	114.059	119.310
3.895	117.600	114.325	112.214	115.050
3.700	110.800	112.446	110.448	110.640
3.524	110.000	110.641	108.752	109.880
3.364	109.000	108.902	107.119	109.280
3.217	109.000	107.224	105.544	108.700
3.083	108.000	105.601	104.021	108.200
2.960	106.000	104.027	102.545	104.450
2.846	105.000	102.499	101.112	102.280
2.741	100.500	101.011	99.719	99.730
2.643	99.000	99.562	98.361	98.960
2.552	97.300	98.146	97.037	96.870
2.467	93.500	96.762	95.743	95.170
2.387	93.400	95.407	94.477	92.520
2.312	93.000	94.078	93.236	91.270
2.242	91.100	92.773	92.019	90.980
2.176	90.500	91.490	90.823	90.410
2.114	89.000	90.227	89.647	88.830
2.056	87.300	88.982	88.489	86.670
2.000	86.000	87.753	87.347	85.620
1.947	84.500	86.539	86.219	83.740
1.897	82.600	85.338	85.105	82.280
1.850	81.800	84.148	84.003	80.930
1.805	78.900	82.968	82.911	78.800
1.762	78.400	81.797	81.829	78.450
1.721	78.000	80.633	80.754	77.320
1.682	77.200	79.474	79.686	77.060
1.644	76.900	78.320	78.624	76.360
1.609	75.800	77.168	77.565	75.300
1.574	72.000	76.018	76.509	71.240
1.542	71.800	74.867	75.455	70.700
1.510	69.500	73.714	74.401	69.580
1.480	67.500	72.557	73.345	68.450
1.451	66.200	71.395	72.287	67.260
1.423	66.000	70.226	71.224	66.710
1.396	65.800	69.047	70.154	64.210
1.370	65.000	67.857	69.077	64.210
1.345	64.000	66.653	67.990	64.210
1.321	62.600	65.431	66.890	61.930
1.298	61.700	64.190	65.775	61.420
1.276	61.200	62.925	64.643	60.950
1.254	61.000	61.632	63.489	60.530
1.233	60.800	60.307	62.310	59.910
1.213	60.300	58.943	61.101	59.910
1.194	59.000	57.534	59.857	59.080
1.175	58.100	56.070	58.572	58.060
1.156	57.100	54.542	57.235	57.310
1.138	57.000	52.936	55.838	56.600
1.121	56.000	51.232	54.364	55.600
1.104	55.000	49.406	52.796	55.000
1.088	54.000	47.423	51.106	53.720
1.072	52.000	45.229	49.253	50.160
1.057	49.200	42.737	47.172	48.560
1.042	46.000	39.792	44.746	46.640
1.028	45.500	36.062	41.731	46.640
1.014	40.600	30.562	37.415	46.640

5 ¿CUÁL ES EL MODELO QUE MEJOR SE AJUSTA?

5.1 Comparación de resultados

El análisis de los Errores Cuadráticos Mínimos (ECM) proporciona una visión importante sobre la calidad de ajuste de los modelos utilizados para predecir las precipitaciones máximas anuales en el período de retorno Tr. Vamos a examinar los resultados y realizar una conclusión de acuerdo a la **Tabla 7** y la **Figura 12** :

- ECM Gumbell: 8.433
- ECM Lognormal: 9.180
- ECM RandomForestRegression: 8.010

En general, un ECM más bajo indica un mejor ajuste del modelo a los datos observados. El modelo de Regresión RandomForest tiene el ECM más bajo (8.010), lo que sugiere que tiene un mejor rendimiento en términos de ajuste a los datos históricos de precipitaciones máximas anuales. Por otro lado, el modelo Gumbell también presenta un ECM relativamente bajo (8.433), indicando un buen ajuste, pero ligeramente inferior al RandomForest. El modelo LogNormal tiene el ECM más alto (9.180), sugiriendo un ajuste menos preciso.

Es esencial considerar que, aunque el ECM proporciona una medida cuantitativa de ajuste, **no es la única métrica a tener en cuenta**. Además, factores como la interpretación física del modelo, la simplicidad y la consistencia con el conocimiento previo también son cruciales.

Dado que el objetivo final es realizar un análisis hidrológico con las precipitaciones máximas anuales históricas, se podría seleccionar el modelo RandomForestRegression como el más adecuado en base a los resultados del ECM. Sin embargo, **se recomienda una evaluación más completa considerando otros aspectos del modelado y la hidrología para tomar decisiones informadas**.

Tabla 7: Errores Cuadráticos Mínimos (ECM)/Jaina

ECM Gumbell	ECM Lognormal	ECM RandomForestRegression
8.433	9.180	8.010

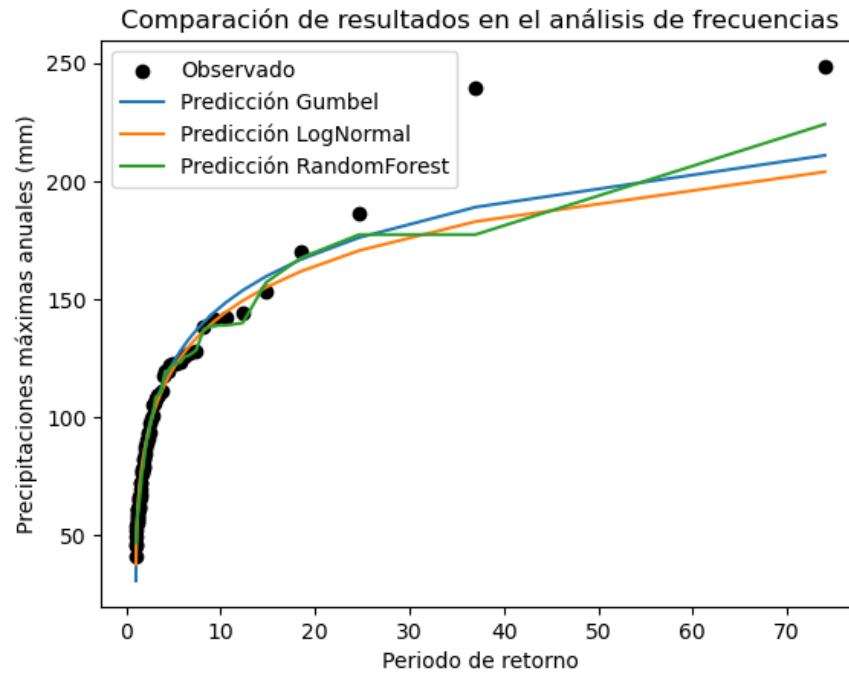


Figura 12: Comparación de resultados en el análisis de frecuencias / Jaina

6 CÓDIGO

Para obtener los archivos con los que se trabajaron se recomienda entrar a [GITHUB](#) y leer el repositorio ["/analishidrologico"](#).

Referencias

- Béjar, M. (2006). *Hidrología estadística*. Instituto Tecnológico de Costa Rica. <https://books.google.com.mx/books?id=BpnGDwAAQBAJ>
- Dodson, B. (2006). *The Weibull Analysis Handbook* (2.^a ed.). ASQ Quality Press.
- Grus, J. (2023). Árboles de decisión. En A. Multimedia (Ed.), *Ciencia de datos desde cero. Principios básicos con Python* (2.^a ed., pp. 282-284). Madrid.
- Orduña, L. (2021). *Hidrología (Acceso)*. Ediciones Paraninfo, S.A. <https://books.google.com.mx/books?id=V7G8zgEACAAJ>