



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

**AJUSTE DE PARÁMETROS EN FUNCIONES DE
DISTRIBUCIÓN DE PROBABILIDAD EMPLEADAS
EN HIDROLOGÍA CON ALGORITMOS GENÉTICOS**

T E S I S

Que para obtener el título de

INGENIERO CIVIL

P R E S E N T A

RAÚL DANIEL HERNÁNDEZ ARUMIR

DIRECTOR DE TESIS

DR. ÓSCAR ARTURO FUENTES MARÍLES



Ciudad Universitaria, Cd. Mx., 2018

DEDICATORIA

A mis padres, por haberme brindado su apoyo, sus experiencias, sus consejos y por haberme dado la oportunidad de ser un profesionalista.

AGRADECIMIENTOS

A mis padres, Raúl Hernández y Juana Arumir, por darme los recursos necesarios de manera incondicional para estudiar.

Al Dr. Óscar Arturo Fuentes Mariles por sus conocimientos, su tiempo, su paciencia y por darme la oportunidad de pertenecer al Instituto de Ingeniería.

A los becarios y colaboradores del Dr. Óscar Arturo Fuentes Mariles por haberme compartido sus experiencias, conocimiento y por haberme incluido en su círculo profesional.

A todas aquellas personas que me ayudaron en el camino al desarrollo de este trabajo y me hicieron crecer en un ámbito multidisciplinario.

Un agradecimiento especial al Ing. Víctor Villegas Castro que me asesoró de manera directa con sus conocimientos en programación.

CONTENIDO

INTRODUCCIÓN	1
CAPÍTULO 1. CONCEPTOS BÁSICOS DE ESTADÍSTICA Y PROBABILIDAD EN HIDROLOGÍA	3
1.1 ESTADÍSTICA	3
1.1.1 <i>Medidas de tendencia central</i>	4
1.1.1.1 Media	4
1.1.1.2 Mediana	4
1.1.1.3 Moda	5
1.1.2 <i>Medidas de dispersión</i>	5
1.1.2.1 Varianza	5
1.1.2.2 Desviación estándar	6
1.1.2.3 Coeficiente de variación	6
1.1.3 <i>Medidas de distribución</i>	6
1.1.3.1 Sesgo	7
1.1.3.2 Curtosis	8
1.2 PROBABILIDAD	9
1.2.1 <i>Variable aleatorias continua</i>	9
1.2.2 <i>Función de densidad de probabilidad</i>	9
1.2.3 <i>Función de distribución de probabilidad</i>	10
1.2.4 <i>Periodo de retorno</i>	11
1.2.4.1 Weibull	11
1.2.4.2 Gringorten	12
1.2.5 <i>Error cuadrático</i>	13
CAPÍTULO 2. MÉTODOS DE AJUSTE	14
2.1 MOMENTOS.....	14
2.2 MÁXIMA VEROSIMILITUD.....	16
2.3 MOMENTOS – L	17
2.4 ALGORITMOS GENÉTICOS	18
2.4.1 <i>Teoría de la evolución</i>	19
2.4.2 <i>Conceptos evolutivos</i>	19
2.4.2.1 Población.....	19
2.4.2.2 Decodificación.....	20
2.4.2.3 Función de adaptación o Fitness	21
2.4.2.4 Selección.....	22
2.4.2.5 Cruza.....	24
2.4.2.6 Mutación	25
2.4.2.7 Criterio de parada del algoritmo genético.....	26
2.4.3 <i>Esquema general de un algoritmo genético</i>	27
2.4.4 <i>Aplicación del algoritmo genético en hidrología</i>	28
CAPÍTULO 3. DISTRIBUCIONES DE PROBABILIDAD PROPUESTAS	30
3.1 NORMAL.....	30
3.1.1 <i>Estimadores por momentos y máxima verosimilitud</i>	32
3.1.2 <i>Estimadores por momentos – L</i>	32
3.1.3 <i>Estimadores por Algoritmos Genéticos</i>	33

3.2	LOG NORMAL 2 PARÁMETROS	34
3.2.1	<i>Estimadores por momentos y máxima verosimilitud</i>	35
3.2.2	<i>Estimadores por Algoritmos Genéticos</i>	36
3.3	LOG NORMAL 3 PARÁMETROS	37
3.3.1	<i>Estimadores por momentos</i>	38
3.3.2	<i>Estimadores por máxima verosimilitud</i>	39
3.3.3	<i>Estimadores por Algoritmos Genéticos</i>	39
3.4	GUMBEL	40
3.4.1	<i>Estimador por momentos</i>	41
3.4.2	<i>Estimador por máxima verosimilitud</i>	41
3.4.3	<i>Estimador por momentos – L</i>	42
3.4.4	<i>Estimador por algoritmos Genéticos</i>	42
3.5	DOBLE GUMBEL	44
3.5.1	<i>Estimador por máxima verosimilitud y algoritmos genéticos</i>	45
CAPÍTULO 4. APLICACIÓN A DISTINTOS TIPOS DE DATOS HIDROLÓGICOS		46
4.1	ESTACIÓN HIDROMÉTRICA DEL RÍO COTAXTLA	48
4.1.1	<i>Distribución Normal</i>	48
4.1.2	<i>Distribución Log Normal 2 Parámetros</i>	51
4.1.3	<i>Distribución Log Normal 3 Parámetros</i>	54
4.1.4	<i>Distribución Gumbel</i>	57
4.1.5	<i>Distribución Doble Gumbel</i>	60
4.2	ESTACIÓN HIDROMÉTRICA DEL RIO JAMAPA	64
4.2.1	<i>Distribución Normal</i>	64
4.2.2	<i>Distribución Log Normal 2 Parámetros</i>	67
4.2.3	<i>Distribución Log Normal 3 Parámetros</i>	70
4.2.4	<i>Distribución Gumbel</i>	73
4.2.4	<i>Distribución Doble Gumbel</i>	76
4.3	ANÁLISIS DE RESULTADOS	80
CAPÍTULO 5. CONCLUSIONES		82
ANEXO 1		83
ANEXO 2		87
ANEXO 3		89
BIBLIOGRAFÍA		90

Figura 1 Los tres estados del sesgo	7
Figura 2 Los tres estados de la curtosis	8
Figura 3 Función de densidad de probabilidad	10
Figura 4 Función de distribución de probabilidad	10
Figura 5 Elementos que integran una población.....	20
Figura 6 Función de adaptación	22
Figura 7 Selección directa.....	23
Figura 8 Selección por ruleta.....	23
Figura 9 Selección por torneo.....	24
Figura 10 Cruza en un punto	25
Figura 11 Mutación de la población	25
Figura 12 Estructura de ejecución de un algoritmo genético.....	27
Figura 13 Función de distribución normal.....	30
Figura 14 Función de distribución lognormal.....	35
Figura 15 Comparación entre valores históricos y calculados. Función Normal – Weibull.....	50
Figura 16 Comparación entre valores históricos y calculados. Función Normal – Gringorten.....	51
Figura 17 Comparación entre valores históricos y calculados. Función Log Normal 2P – Weibull	53
Figura 18 Comparación entre valores históricos y calculados. Función Log Normal 2P – Gringorten	54
Figura 19 Comparación entre valores históricos y calculados. Función Log Normal 3P – Weibull	56
Figura 20 Comparación entre valores históricos y calculados. Función Log Normal 3P – Gringorten	57
Figura 21 Comparación entre valores históricos y calculados. Función Gumbel – Weibull	59
Figura 22 Comparación entre valores históricos y calculados. Función Gumbel – Gringorten.....	60
Figura 23 Comparación entre valores históricos y calculados. Función Doble Gumbel – Weibull.....	62
Figura 24 Comparación entre valores históricos y calculados. Función Doble Gumbel – Gringorten	63
Figura 25 Comparación entre valores históricos y calculados. Función Normal – Weibull.....	66
Figura 26 Comparación entre valores históricos y calculados. Función Normal – Gringorten.....	67
Figura 27 Comparación entre valores históricos y calculados. Función Log Normal 2P – Weibull	69
Figura 28 Comparación entre valores históricos y calculados. Función Log Normal 2P – Gringorten	70
Figura 29 Comparación entre valores históricos y calculados. Función Log Normal 3P – Weibull	72
Figura 30 Comparación entre valores históricos y calculados. Función Log Normal 3P – Gringorten	73
Figura 31 Comparación entre valores históricos y calculados. Función Gumbel – Weibull	75
Figura 32 Comparación entre valores históricos y calculados. Función Gumbel – Gringorten.....	76
Figura 33 Comparación entre valores históricos y calculados. Función Doble Gumbel – Weibull.....	78
Figura 34 Comparación entre valores históricos y calculados. Función Doble Gumbel – Gringorten	79
Tabla 1 Tabla para la obtención del valor de a en la expresión de Gringorten	12
Tabla 2 Gastos máximos anuales de la estación “Paso del toro”	48
Tabla 3 Resultados obtenidos de la distribución Normal para estación “Paso del toro”	48
Tabla 4 Valores calculados. Función Normal – Weibull y Gringorten.....	49
Tabla 5 Resultados obtenidos de la distribución Log Normal 2P para estación “Paso del toro”	51
Tabla 6 Valores calculados. Función Log Normal 2P – Weibull y Gringorten	52
Tabla 7 Resultados obtenidos de la distribución Log normal 3P para estación “Paso del toro”	54
Tabla 8 Valores calculados. Función Log Normal 3P – Weibull y Gringorten	55

Tabla 9 Resultados obtenidos de la distribución Gumbel para estación “Paso del toro”	57
Tabla 10 Valores calculados. Función Gumbel – Weibull y Gringorten	58
Tabla 11 Resultados obtenidos de la distribución Doble Gumbel para estación “Paso del toro”	60
Tabla 12 Valores calculados. Función Doble Gumbel – Weibull y Gringorten	61
Tabla 13 Gastos máximos anuales de la estación “El Tejar”	64
Tabla 14 Resultados obtenidos de la distribución Normal para estación “El Tejar”	64
Tabla 15 Valores calculados. Función Normal – Weibull y Gringorten	65
Tabla 16 Resultados obtenidos de la distribución Log Normal 2P para estación “El Tejar”	67
Tabla 17 Valores calculados. Función Log Normal 2P – Weibull y Gringorten	68
Tabla 18 Resultados obtenidos de la distribución Log normal 3P para estación “El Tejar”	70
Tabla 19 Valores calculados. Función Log Normal 3P – Weibull y Gringorten	71
Tabla 20 Resultados obtenidos de la distribución Gumbel para estación “El tejar”	73
Tabla 21 Valores calculados. Función Gumbel – Weibull y Gringorten	74
Tabla 22 Resultados obtenidos de la distribución Doble Gumbel para estación “El Tejar”	76
Tabla 23 Valores calculados. Función Doble Gumbel – Weibull y Gringorten	77
Ilustración 1 Interfaz Spyder Anaconda.....	28
Ilustración 2 Ubicación de las estaciones hidrométricas, Google Earth.....	46
Ilustración 3 Programa AX (Jiménez, 1996).....	47
Ilustración 4 Gráfica del Ln (L) en cada generación para la función de distribución Normal	50
Ilustración 5 Gráfica del Ln (L) en cada generación para la función de distribución Log Normal 2P	53
Ilustración 6 Gráfica del Ln (L) en cada generación para la función de distribución Log Normal 3P	56
Ilustración 7 Gráfica del Ln (L) en cada generación para la función de distribución Gumbel	59
Ilustración 8 Gráfica del Ln (L) en cada generación para la función de distribución Doble Gumbel	62
Ilustración 9 Resultados globales del programa AX	63
Ilustración 10 Gráfica del Ln (L) en cada generación para la función de distribución Normal	66
Ilustración 11 Gráfica del Ln (L) en cada generación para la función de distribución Log Normal 2P	69
Ilustración 12 Gráfica del Ln (L) en cada generación para la función de distribución Log Normal 3P	72
Ilustración 13 Gráfica del Ln (L) en cada generación para la función de distribución Gumbel	75
Ilustración 14 Gráfica del Ln (L) en cada generación para la función de distribución Doble Gumbel.....	78
Ilustración 15 Resultados globales del programa AX	79

“Olvidamos que el ciclo del agua y el ciclo de la vida son uno mismo”

Jacques-Yves Cousteau

Introducción

El agua es un elemento no renovable y esencial en la vida cotidiana del ser humano. Esto la vuelve objeto de estudio para su cuidado, tratamiento y distribución. El uso y consumo del agua son factores que influyen de manera directa en el desarrollo de una sociedad. Es entonces que surge una ciencia natural que estudia el agua, su ocurrencia, su circulación y su distribución en la superficie terrestre, la hidrología. En particular existe una rama de esta ciencia que estudia la ingeniería, la hidrología aplicada, esta rama es la responsable del estudio del agua en la superficie, para el diseño y operación de obras civiles hidráulicas que tienen como objetivo controlar y aprovechar el agua.

Se ha intentado definir de manera simple y práctica a la Hidrología y se ha dicho que es una ciencia interpretativa, ya que en general sus resultados están basados en el procesamiento estadístico de la información hidrométrica, pluviográfica y pluviométrica disponible.

En un país único por su orografía como lo es México se presenta una gran diversidad de condiciones de lluvia, en consecuencia, resulta en una desigual distribución del líquido vital sobre la superficie. Además, las precipitaciones y los escurrimientos nunca se presentan de la misma manera, y las características peculiares de cada caso son determinantes para la construcción de obras hidráulicas. En este tipo de proyectos se busca responder a una pregunta primordialmente. ¿Cuánta agua?

Debido a que se habla de fenómenos naturales, estos son impredecibles si lo que se busca es exactitud. Sin embargo, en la inagotable necesidad del ser humano por predecir fenómenos naturales o modelar la naturaleza misma, se ha logrado desarrollar la teoría necesaria mediante análisis estadísticos y probabilísticos para asemejarse a la realidad. En este largo camino donde múltiples personajes han hecho sus aportaciones directa o indirectamente a la hidrología surgen modelos que se pueden adaptar a la escasa información que se tiene de eventos naturales como lo son las lluvias.

Es entonces cuando las funciones de distribución como la *Normal*, *Log Normal*, *Gumbel*, *Doble Gumbel*, entre otras, aparecen para tratar responder la pregunta anterior. Existe una amplia selección de estos modelos que participarán para determinar cuál es el más adecuado para describir el comportamiento de un conjunto de datos históricos dados.

Como se ha enfatizado estos modelos matemáticos no son fieles al 100 % de lo que ocurre en la realidad. Se presentan errores en cada uno de ellos y es deber del especialista determinar cuál de todos ellos proporciona el menor error. Lo ideal, es que el error que se

presenta en las funciones de distribución sea nulo, en otras palabras, que el modelo describa completamente el comportamiento de los eventos.

Para reducir los errores que presentan los algoritmos determinísticos, en este trabajo se propone la utilización de algoritmos no determinísticos que están fundamentados en una teoría puramente probabilística. Modelos como los algoritmos genéticos permiten explorar espacios a los que métodos convencionales no pueden ingresar por su naturaleza.

El objetivo de utilizar estos algoritmos es reducir el error que se presenta en las funciones de distribución, en consecuencia, se precisará en la estimación de eventos futuros que podrían reducir costos y tiempo en la construcción de obras hidráulicas, y tener un criterio más amplio en cuanto a la toma de decisiones en proyectos de hidrología.

CAPÍTULO 1. CONCEPTOS BÁSICOS DE ESTADÍSTICA Y PROBABILIDAD EN HIDROLOGÍA

1.1 Estadística

El campo de la estadística tiene que ver con la recopilación, presentación, análisis y uso de datos para tomar decisiones y resolver problemas. La importancia de la estadística en la ingeniería ha quedado manifestada al relacionarla con la mejoría en la calidad de la obtención de resultados. La estadística aplicada a la información hidrológica es una herramienta fundamental para organizar, analizar e interpretar las características de una población a partir de una muestra.

El término población se refiere a la recolección de mediciones de todos los elementos del universo con respecto al cual se quieren obtener conclusiones o tomar decisiones. En la mayoría de las aplicaciones de la estadística consisten de una muestra de la población de interés. Esta muestra es solo un subconjunto de observaciones seleccionadas de una población.

La estadística se clasifica en *univariable* o *multivariable*, dependiendo de la cantidad de variables con las que se estén trabajando, esto se refiere a la relación de las variables de una muestra.

La estadística también se puede clasificar en *descriptiva* o *inferencial*. En la primera se puede resumir y organizar datos tanto de una población como de una muestra, mientras que en la inferencial únicamente trabaja con muestras a partir de las cuales intenta extraer conclusiones sobre la población, es decir, de manera intuitiva.

En la hidrología es utilizada la estadística descriptiva. Por la naturaleza de los datos de precipitaciones que se tienen es posible extraer información, organizarla y hacer un resumen de forma cuantitativa.

La estadística descriptiva es compuesta principalmente por tres técnicas:

- Distribuciones de frecuencias
- Gráficas
- Medidas numéricas

1.1.1 Medidas de tendencia central

Algunas de las medidas numéricas de la estadística descriptiva son las medidas de tendencia central que dado un conjunto de datos agrupados buscan dar a conocer su característica principal. Dicho de otra manera, pretenden resumir en un solo valor a un conjunto de valores. Las medidas de tendencia central más utilizadas son: la media, la mediana y la moda.

1.1.1.1 Media

Debido a que en la mayoría de los casos se toma a los datos como la *muestra* nos referimos a la media aritmética como la media de la *muestra*. La media de la *muestra* representa el valor promedio de todas las observaciones en dicho conjunto. Donde n es el tamaño de la muestra y las observaciones son x_1, x_2, \dots, x_n , entonces la media de la *muestra* es:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n} \quad (1.1)$$

La media de la *población* está representada por la letra griega μ .

1.1.1.2 Mediana

Es el punto en el cual la *muestra* se divide en 2 mitades iguales. Siendo $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ los elementos de una muestra arreglada en orden creciente, es decir, que $x_{(1)}$ representa el valor más pequeño de toda la muestra, $x_{(2)}$ el segundo valor más pequeño y $x_{(n)}$ el valor más grande de la muestra. Entonces la mediana se define matemáticamente como:

$$\mathbf{Mediana} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & n \text{ impar} \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\left[\frac{n}{2}\right]+1\right)}}{2} & n \text{ par} \end{cases} \quad (1.2)$$

1.1.1.3 Moda

La moda es la observación que se presenta con más frecuencia. Si existen dos o más valores con la misma frecuencia máxima se dice entonces que se tiene una distribución multimodal. También puede ocurrir el caso en el que no exista un valor que se repita más veces que los otros.

Caso 1: Sean x_1, x_2, x_3, x_3, x_4 las observaciones de la muestra, entonces:

$$\mathbf{Moda} = x_3$$

Caso 2: Sean $x_1, x_1, x_2, x_3, x_3, x_4$ las observaciones de la muestra, entonces:

$$\mathbf{Moda} = x_1 \text{ y } x_3$$

Caso 3: Sean x_1, x_2, x_3, x_4 las observaciones de la muestra, entonces:

$$\mathbf{Moda} = \emptyset$$

1.1.2 Medidas de dispersión

Los valores de las medidas de tendencia central no necesariamente proporcionan información útil, por lo que es necesario recurrir a las medidas de dispersión que aportan información sobre qué tan lejos están los valores de las medidas de tendencia central.

1.1.2.1 Varianza

Es el promedio de las distancias de los valores observados x_i elevadas al cuadrado respecto a la media aritmética. Se representa con σ^2 ó S^2 , esto dependiendo del conjunto de datos con el que se cuenta. Se divide entre n cuando se cuenta con la *población* y entre $n-1$ cuando se tiene la *muestra*.

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (1.3)$$

1.1.2.2 Desviación estándar

Es la raíz cuadrada positiva de la varianza y es representado con σ para un conjunto de datos poblacional y con la letra S cuando se cuenta con la muestra.

$$\begin{aligned}\sigma &= \sqrt{\sigma^2} \\ S &= \sqrt{S^2}\end{aligned}\tag{1.4}$$

1.1.2.3 Coeficiente de variación

Cuando se desea expresar la variación como una fracción de la media. Se utiliza una medida adimensional de variación relativa. El coeficiente de variación es útil cuando se compara la variabilidad de 2 o más conjuntos de datos que difieren de manera considerable en la magnitud de las observaciones. Se calcula como el cociente de la desviación estándar entre la media y se representa para datos *poblacionales* y *muestrales* de la siguiente forma respectivamente:

$$\begin{aligned}Cv &= \frac{\sigma}{\bar{\mu}} \\ Cv &= \frac{S}{\bar{x}}\end{aligned}\tag{1.5}$$

1.1.3 Medidas de distribución

Son aquellas que nos permiten conocer en qué forma se alejan o se acercan los datos de acuerdo a una representación gráfica que está relacionada con la frecuencia con la que se hallen dentro de la información.

1.1.3.1 Sesgo

Esta medida nos permite identificar si los datos se distribuyen de forma uniforme alrededor de la media aritmética. La asimetría presenta tres estados diferentes, cada uno de los cuales define de forma concisa como están distribuidos los datos respecto al eje de asimetría. Se calcula de la siguiente manera dependiendo de qué tipo de datos se tengan:

$$\begin{array}{ll}
 \text{Datos poblacionales} & \gamma = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{(S^2)^{(3/2)}} \\
 \text{Datos muestrales} & g = \frac{n^2}{(n-1)(n-2)} * \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{(\sigma^2)^{(3/2)}}
 \end{array} \tag{1.6}$$

Si $g < 0$, la curva es asimétricamente negativa por lo que los valores se tienden a reunir más en la parte derecha de la media.

Si $g = 0$, se acepta que la distribución es simétrica, es decir, existe aproximadamente la misma cantidad de valores a los dos lados de la media.

Si $g > 0$, la curva es asimétricamente positiva por lo que los valores se tienden a reunir más en la parte izquierda que en la derecha de la media.

En la *Figura 1* se representa de manera gráfica los 3 casos anteriores.

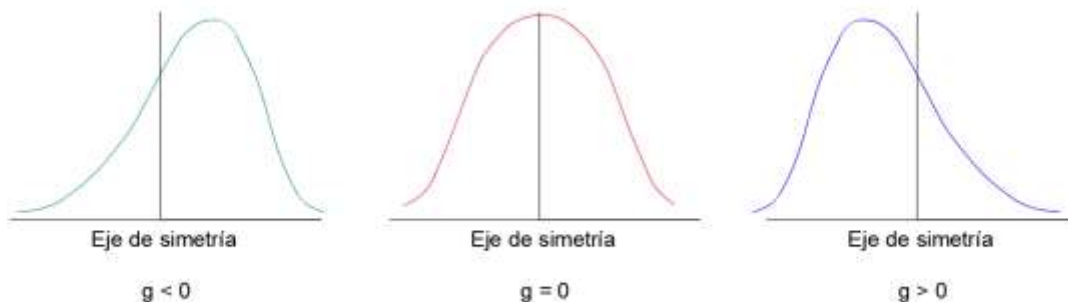


Figura 1 Los tres estados del sesgo

1.1.3.2 Curtosis

Esta medida determina el grado de concentración que presentan los valores en la región central de la distribución. Se calcula de la siguiente manera dependiendo del tipo de datos que se tengan:

$$\begin{aligned}
 \text{Datos poblacionales} \quad k &= \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{(S^2)^2} \\
 \text{Datos muestrales} \quad K &= \frac{n^3}{(n-1)(n-2)(n-3)} * \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{(S^2)^2}
 \end{aligned}
 \tag{1.7}$$

Si $k > 0$, la distribución es Leptocúrtica

Si $k = 0$, la distribución es Mesocúrtica

Si $k < 0$, la distribución es Platicúrtica

En la *Figura 2* se representa de manera gráfica los 3 casos anteriores.

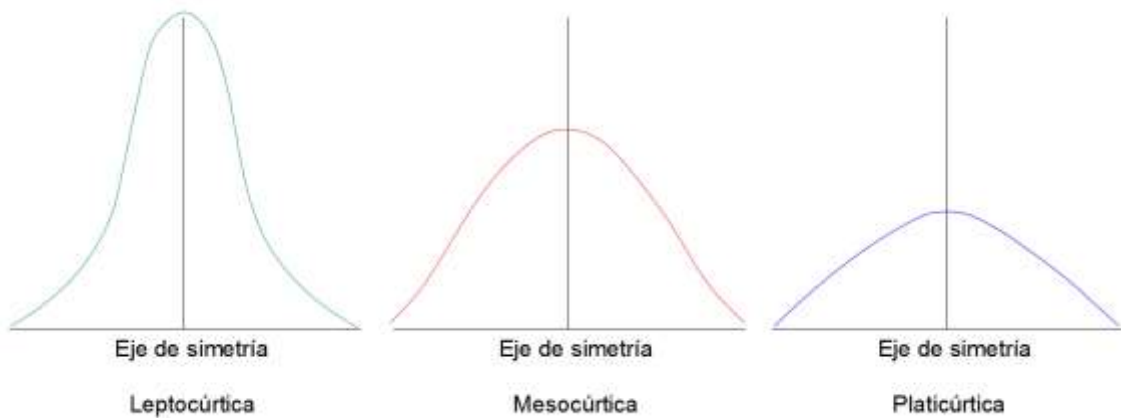


Figura 2 Los tres estados de la curtosis

1.2 Probabilidad

La probabilidad ha alcanzado un amplio uso para cuantificar el grado de confianza en un evento de interés. La Probabilidad es la mayor o menor posibilidad de que ocurra un determinado suceso. En otras palabras, su noción viene de la necesidad de medir o determinar cuantitativamente la certeza o duda de que un suceso dado ocurra o no.

Un experimento aleatorio es aquel que proporciona diferentes resultados aun cuando se repita siempre de la misma manera. El conjunto de los posibles resultados de un experimento aleatorio recibe el nombre de espacio muestral del experimento. Un evento es un subconjunto del espacio muestral de un experimento aleatorio.

1.2.1 Variable aleatorias continua

Si el rango de una variable aleatoria X contiene un intervalo, ya sea finito o infinito de números reales, entonces X es una variable aleatoria continua. Las variables aleatorias continuas cumplen con las siguientes características.

1. $f(x) \geq 0$ para toda $x \in R_x$
2. $\int_{R_x} f(x)dx = 1$
3. $f(x)$ es un tramo continuo
4. $F(x) = 0$ si x no esta en el rango de R_x

1.2.2 Función de densidad de probabilidad

La función de densidad de probabilidad generalmente denotada como $f(x)$ es el lugar geométrico que representa a la variable aleatoria, es decir, describe la probabilidad según la cual dicha variable aleatoria tomará determinado valor. Esta función es positiva a lo largo de su dominio y su integral sobre el espacio es de valor unitario. Debe cumplir que para cualquier intervalo de números reales ocurre que:

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $P(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(u)du$

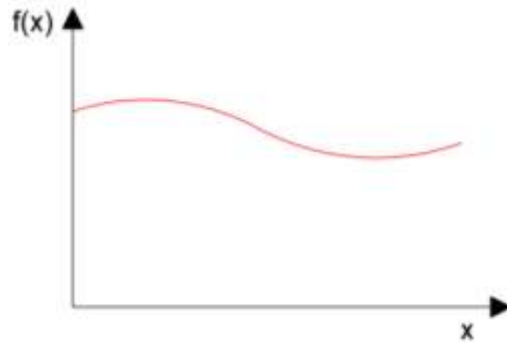


Figura 3 Función de densidad de probabilidad

1.2.3 Función de distribución de probabilidad

La función de distribución de probabilidad es el área bajo la curva del lugar geométrico de $f(x)$. Es una función que asigna a cada suceso definido sobre la variable aleatoria la probabilidad de que dicho suceso ocurra. Tomando a x minúscula como un valor particular de la variable aleatoria de tal modo que $(X = x)$, (X, x) , $(X \leq x)$ son eventos en el espacio del rango de la variable aleatoria X , donde x es un número real. La probabilidad del evento $(X \leq x)$ puede expresarse como función de x en la forma:

$$F(x) = P(X \leq x) \quad (1.8)$$

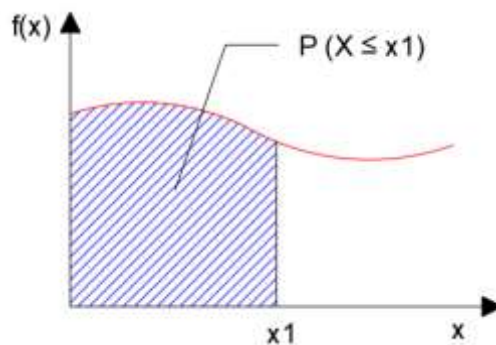


Figura 4 Función de distribución de probabilidad

1.2.4 Periodo de retorno

Se puede establecer que si un evento hidrológico X igual o mayor que x ocurre una vez en promedio en un lapso de Tr años, entonces el cociente $1/Tr$ corresponderá a su probabilidad de excedencia $P(X > x)$. El intervalo de recurrencia o de repetición promedio en años, como el inverso de la probabilidad de excedencia, o bien como el inverso de uno menos la probabilidad de no excedencia.

$$T = \frac{1}{P(x \geq x_0)} \quad (1.9)$$

$$P(x < x_0) = 1 - P(x \geq x_0) \quad (1.10)$$

El periodo de retorno es una forma de expresar la probabilidad de excedencia, por ello se dice la lluvia o intensidad de 10 años o la creciente de 100 años en lugar de decir, los eventos cuyas probabilidades de excedencia son 10% y 1 % en cada año, correspondiendo a posibilidades de 1 en 10 y 1 en 100. El periodo de retorno no significa que un evento de Tr años ocurrirá cada Tr años, sino que más bien existe una probabilidad de $1/Tr$ de que tal evento ocurra en cada año.

$$P(x \geq x_0) = \frac{1}{T} \quad (1.11)$$

Existen diversos autores que han propuesto expresiones para estimar el periodo de retorno Tr , en este trabajo se abordarán las expresiones desarrolladas por los autores Weibull y Gringorten con el objetivo de visualizar las diferencias más significativas obtenidas en los resultados.

1.2.4.1 Weibull

La expresión propuesta por Waloddi Weibull involucra el número de datos con los que se cuenta y un método que ordena dichos datos en orden ascendente.

$$Tr = \frac{n + 1}{m} \quad (1.12)$$

Donde:

n Es el tamaño de la muestra

m Es el número de orden de la serie de datos

1.2.4.2 Gringorten

Otro autor ha propuesto otra expresión para calcularlos tomando en cuenta el tamaño de la muestra que se tiene. Irving Gringorten propone la siguiente expresión general:

$$Tr = \frac{n + 1 - 2a}{m - a} \quad (1.13)$$

Donde:

n Es el tamaño de la muestra

m Es el número de orden de la serie de datos

a Es un parámetro que depende de n como se muestra en la tabla siguiente

n	10	20	30	40	50
a	0.448	0.443	0.442	0.441	0.44
n	60	70	80	90	100
a	0.44	0.44	0.44	0.439	0.439

Tabla 1 Tabla para la obtención del valor de a en la expresión de Gringorten

En general $a = 0.4$ es recomendado usar en la ecuación de Gringorten. Si la distribución es aproximadamente normal, $a = 0.375$. Un valor de $a = 0.44$ en los datos ajustados a una distribución Gumbel.

1.2.5 Error cuadrático

Para encontrar la función de distribución que cuenta con el mejor ajuste, existen distintos criterios para considerarla, en este trabajo se emplea el error cuadrático. Esta técnica fue propuesta por Kite en 1988. Se obtiene como:

$$EE = \sqrt{\frac{\sum_{i=1}^n (X_i - X_{Ti})^2}{n - m_p}} \quad (1.14)$$

Donde:

- X_i Son los eventos ordenados de mayor a menor con un periodo de retorno asignado
- X_{Ti} Son los eventos estimados con alguna función de probabilidad para cada periodo de retorno
- n El tamaño de la muestra
- m_p Número de parámetros de la distribución ajustada

La distribución de mejor ajuste será aquella que proporcione el mínimo valor del error cuadrático. Si una o más distribuciones tienen valores similares del error cuadrático, entonces se deberá optar por aquella distribución que tenga el menor número de parámetros.

CAPÍTULO 2. MÉTODOS DE AJUSTE

Los problemas en hidrología requieren del uso de modelos que pueden ser determinísticos, paramétrico o estocástico, que permiten proporcionar un valor aproximado de un parámetro de la población, a partir de los datos de la muestra.

Un modelo determinístico es aquel que se obtiene a través de las relaciones físicas y no requiere de datos experimentales para su aplicación.

Un modelo paramétrico se puede considerar como un determinístico dado que los parámetros de entrada que se le dan al modelo siempre arrojarán las mismas salidas. Sin embargo también se puede decir que es estocástico en el sentido de que los parámetros dependen de los datos observados, es decir, que si los datos observados cambian, los parámetros también lo harán.

Un modelo estocástico es aquel en el que las entradas llegan a ser predecibles solo en un sentido estadístico. En este tipo de modelos, el empleo repetido de entradas generan salidas que no se repetirán, sin embargo, siguen un modelo estadístico.

2.1 Momentos

Es un método de obtención de estimadores poblacionales relativamente muy sencillo. Su procedimiento consiste en plantear un sistema de ecuaciones, cuyo tamaño dependerá del número de parámetros a estimar. El sistema se obtiene al igualar los momentos poblacionales con los muestrales.

Los momentos poblacionales se pueden obtener respecto a la media en variables continuas con la expresión:

$$M_r^\mu = \int_{-\infty}^{\infty} (x - \mu)^r f(x) dx \quad (2.1)$$

O con respecto al origen:

$$M_r = E[x^r] = \int_{-\infty}^{\infty} x^r f(x) dx \quad (2.2)$$

Es posible obtener el valor de la media de la variable aleatoria continua X con el momento respecto al origen de orden 1, también conocido como esperanza matemática o valor esperado de X .

$$\mu = E(x) = \int_{-\infty}^{\infty} xf(x)dx \quad (2.3)$$

Con ayuda del momento respecto a la media de segundo orden se obtiene la expresión para la varianza.

$$\sigma^2 = Var(x) = E[(x - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx \quad (2.4)$$

Con el tercer momento respecto a la media y dividida entre un factor para obtener un valor adimensional se tiene la expresión para el sesgo.

$$\gamma = \frac{E[(x - \mu)^3]}{\sigma^3} = \frac{\int_{-\infty}^{\infty} (x - \mu)^3 f(x)dx}{\sigma^3} \quad (2.5)$$

De forma análoga al sesgo la curtosis es el cuarto momento respecto a la media.

$$k = \frac{E[(x - \mu)^4]}{\sigma^4} = \frac{\int_{-\infty}^{\infty} (x - \mu)^4 f(x)dx}{\sigma^4} \quad (2.6)$$

Los momentos muestrales, son los también conocidos como estadísticos muestrales y se obtienen como se menciona en el capítulo 1.

2.2 Máxima verosimilitud

Uno de los mejores métodos para obtener un estimador puntual de un parámetro es el de máxima verosimilitud. Al tomar una muestra aleatoria de una población, se sabe que los elementos de dicha muestra son variables aleatorias y que se obtienen con cierta probabilidad, la cual varía de acuerdo con el valor del parámetro poblacional desconocido.

Sea $f(x; a_1, a_2, \dots, a_m)$ una función de densidad de probabilidad de x con a_i parámetros, $i = 1, \dots, m$ y x_1, x_2, \dots, x_n los valores observados en una muestra de tamaño n . La función de verosimilitud de la muestra es

$$L = \prod_{i=1}^n f(x_i; a_1, a_2, \dots, a_m) \quad (2.7)$$

El método de máxima verosimilitud estima los parámetros al maximizar la función de verosimilitud para encontrar el valor para el cual sea máxima la probabilidad de haber obtenido los valores muestrales encontrados. Frecuentemente $\ln L$ es usado para simplificar los cálculos, entonces, la función de verosimilitud puede expresarse como:

$$\ln L = \ln \prod_{i=1}^n f(x_i; a_1, a_2, \dots, a_m) \quad (2.8)$$

Para obtener el máximo de la función es necesario aplicar la diferenciación parcial con respecto a cada parámetro e igualar a cero. En consecuencia, se generarán m ecuaciones con m incógnitas y así poder resolverse para m parámetros desconocidos.

$$\begin{aligned} \frac{\partial L(a_1, a_2, \dots, a_m)}{\partial a_1} &= 0 \\ \frac{\partial L(a_1, a_2, \dots, a_m)}{\partial a_2} &= 0 \\ &\vdots \\ \frac{\partial L(a_1, a_2, \dots, a_m)}{\partial a_m} &= 0 \end{aligned} \quad (2.9)$$

2.3 Momentos – L

Los momentos – L son similares a los anteriormente mencionados con la diferencia de que estos son capaces de caracterizar a un mayor número de distribuciones, además están virtualmente libres de sesgo aún para muestras pequeñas. (Escalante Sandoval & Reyes Chávez, 2008)

El primer es la media definido como:

$$l_1 = E[X] \quad (2.10)$$

Sea $x_{(i|n)}$ la i -ésima observación en una muestra de tamaño n , ordenada de mayor a menor, para cualquier distribución de probabilidad el segundo momento – L es una descripción de escala basada en la diferencia esperada entre dos observaciones seleccionadas de forma aleatoria. (Escalante Sandoval & Reyes Chávez, 2008)

$$l_2 = \frac{1}{2} E[X_{(1|2)} - X_{(2|2)}] \quad (2.11)$$

De forma análoga, el sesgo y la curtosis se obtienen como:

$$l_3 = \frac{1}{2} E[X_{(1|3)} - 2X_{(2|3)} + X_{(3|3)}] \quad (2.12)$$

$$l_4 = \frac{1}{2} E[X_{(1|4)} - 3X_{(2|4)} + 3X_{(3|4)} - X_{(4|4)}] \quad (2.13)$$

Los momentos – L pueden estimarse en función de los momentos de probabilidad pesada.

$$\beta_r = E\{X[F(X)]^r\} \quad (2.14)$$

Los primeros momentos – L son:

$$l_1 = \beta_0 \quad (2.15)$$

$$l_2 = 2\beta_1 - \beta_0 \quad (2.16)$$

$$l_3 = 6\beta_2 - 6\beta_1 + \beta_0 \quad (2.17)$$

$$l_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \quad (2.18)$$

Donde

$$\beta_r = M_{i,r,0} = \int_0^1 xF^r dF \quad (2.19)$$

2.4 Algoritmos genéticos

Los algoritmos genéticos son una técnica de resolución de problemas de búsqueda y optimización que surgen entre los años 50's y 70's, inspirada en la teoría de la evolución de las especies y la selección natural, aunque recientemente la computación evolutiva (CE) ha sido reconocida y el término fue inventado en los 90's. De acuerdo con la teoría de Darwin la evolución de las especies se debe al principio de selección natural que favorece la supervivencia y la multiplicación de aquellas especies que están mejor adaptadas a las condiciones de su entorno. Tienen la característica de trabajar la búsqueda aleatoria con la búsqueda dirigida, que surgen de los individuos mejor adaptados. Esta cualidad les permite ingresar a otras regiones del espacio de búsqueda que exploran las soluciones de una forma distinta a los métodos puramente aleatorios. Comúnmente utilizan codificación binaria. (Araujo & Cervigón, 2009)

Proporcionan un esquema general para la resolución de problemas. Tenemos el algoritmo diseñado para el problema que nos ocupa, únicamente hay que especificar la forma de ciertos componentes. La parte que resulta más dependiente del problema específico considerado es la definición de la función de adaptación. Se debe tener presente que los algoritmos evolutivos no garantizan una solución exacta al problema abordado, es una aproximación cuya calidad dependerá del tiempo y memoria disponible. (Araujo & Cervigón, 2009)

2.4.1 Teoría de la evolución

De acuerdo con la teoría de Darwin de la evolución de las especies se debe al principio de *selección* natural, que favorece la supervivencia y la multiplicación de aquellas especies que están mejor adaptadas a su *entorno*. Otro elemento que Darwin señaló como relevante para la evolución son las *mutaciones* o pequeñas variaciones que introducen diferencias en las características físicas y tipos de respuesta de los padres y de los hijos. Mientras la *población* crece exponencialmente, en consecuencia se presentan situaciones de escasez de recursos en el *entorno* y los *individuos* “mejor adaptados” tienen mayores posibilidades de sobrevivir y de dejar descendencia. (Araujo & Cervigón, 2009)

Dando lugar a teorías neodarwinistas se puede decir que las características físicas de un individuo, su *fenotipo*, son la consecuencia de su información genética o *genotipo*, cadenas de genes con complejas interacciones. Los genes pueden modificarse puntualmente por *mutaciones*. La replicación en las cadenas de genes no siempre es perfecta. A veces aunque con una frecuencia extremadamente baja, se producen errores en los procesos de copia y ocurren *mutaciones*. Sin embargo existen mecanismos para mitigar o agravar esta situación. (Araujo & Cervigón, 2009)

Los individuos son la consecuencia del *genotipo* y su interacción con el medio, constituyendo las unidades de *selección*. Lo que evoluciona es el conjunto de individuos que constituyen la *población*, que representa a un conjunto de *genes* comunes a sus individuos. Estas ideas están en la base de los algoritmos evolutivos y se debe tener presente que la naturaleza evoluciona a lo largo de millones de años, mientras que en los problemas de optimización se requieren resultados en periodos de tiempo muy cortos.

2.4.2 Conceptos evolutivos

2.4.2.1 Población

La *población* es un conjunto de *individuos* generados de manera completamente aleatoria candidatos a ser la mejor solución, pero no la solución misma al problema en cuestión. Cada uno de los *individuos* está definido como un *cromosoma* que contiene la información genética llamado *genotipo* en forma de cadenas de bits. Mientras que el *fenotipo* del individuo es la representación de esa información genética. De esta forma el individuo no representa el valor de la solución que aporta, sino el valor que toma cada una de las variables que dan la solución al problema. El tamaño de la población es arbitraria, pero se debe tener en cuenta que una población grande proporciona más posibilidades de encontrar una solución.

La codificación binaria de los individuos resulta muy flexible al trabajar con múltiples variables y es manipulable la longitud del cromosoma del individuo para brindar un mayor campo de soluciones, además que es propensa a recibir de manera directa las variaciones de una *mutación* y hacer más diversa a la población.

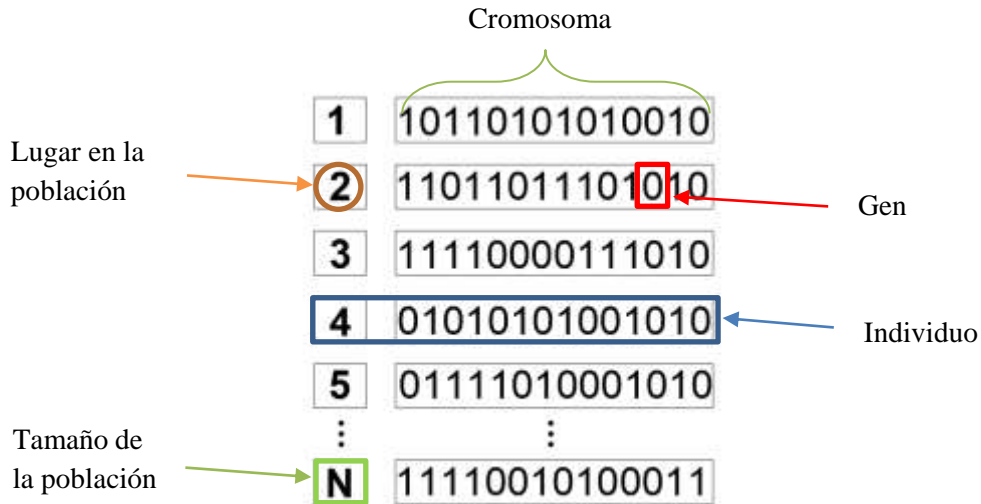


Figura 5 Elementos que integran una población

2.4.2.2 Decodificación

El cromosoma de cada individuo está codificado en sistema binario y estos a su vez pueden mapear un valor en el espacio del sistema a optimizar en cuestión. Para conocer dicho valor es necesario realizar un proceso de decodificación, este proceso es el proceso de conversión del sistema binario al decimal y posteriormente un escalamiento al conjunto de valores que puede tomar el sistema a optimizar.

Por ejemplo, se genera una población binaria de 5 individuos con una longitud de cromosoma de 6 bits y los valores que puede tomar el sistema a optimizar están dados por [0, 1].

$$Población_{(binaria)} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

A continuación, se realiza la conversión de binario a decimal de cada individuo y se obtiene el siguiente arreglo con números decimales.

$$Población_{(decimal)} = \begin{bmatrix} 35 \\ 41 \\ 30 \\ 55 \\ 62 \end{bmatrix}$$

Finalmente, se realiza una conversión a la escala deseada en la que se encuentra el espacio del sistema en discurso, se transforma al intervalo que pueden tomar los individuos ($[I_{max}, I_{min}]$) previamente establecido con la siguiente expresión:

$$P_{dec} = (I_{max} - I_{min}) \frac{P_{10}}{2^L - 1} + I_{min} \quad (2.20)$$

Donde P_{10} es el individuo en decimal, L es la longitud del cromosoma y P_{dec} es el individuo escalado en el intervalo predeterminado. De esta manera se obtiene la población de individuos decodificado en la misma naturaleza que el espacio de búsqueda del sistema de optimización.

$$Población_{(decodificada)} = \begin{bmatrix} 0.5555555555555556 \\ 0.650793650793651 \\ 0.476190476190476 \\ 0.873015873015873 \\ 0.984126984126984 \end{bmatrix}$$

A pesar de obtener este arreglo de valores decodificados, se debe tener presente que el algoritmo genético actúa sobre la población binaria y que la decodificación es una herramienta para obtener el fenotipo de la genética del individuo.

2.4.2.3 Función de adaptación o Fitness

Tal y como sucede en la naturaleza, los individuos que mejor se adaptan a su entorno son los que tienen más oportunidad de reproducirse y que su material genético se propague en las siguientes generaciones. Esto es lo que se intenta simular en los algoritmos genéticos y la función de adaptación es la que permite decir qué tan “bueno” o “malo” es un individuo.

Existen distintos métodos para evaluar a los individuos de una población, tales como normalizar el valor de la solución entre 0 y 1 en función de los valores obtenidos por todos y cada uno de los individuos hasta considerar penalizaciones para los menos aptos, sin embargo la más frecuentemente empleada por su sencillez es asignar un valor real que defina cómo es de bueno un individuo comparado con los demás considerando los objetivos del problema.

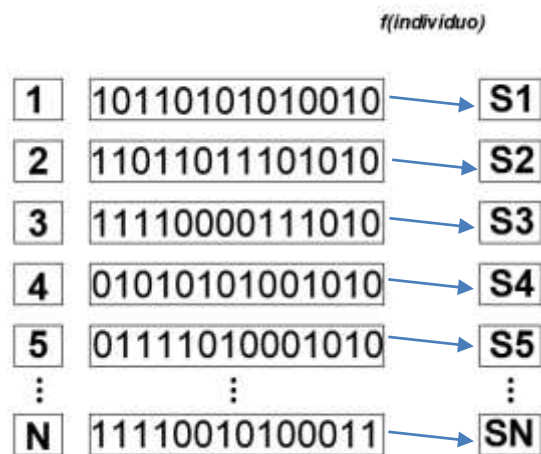


Figura 6 Función de adaptación

2.4.2.4 Selección

La selección tiene como objetivo elegir a los individuos que jugarán el papel de padres de la siguiente generación. A pesar de que se busca transferir la información genética de los individuos mejor adaptados a la siguiente generación no se seleccionan dichos individuos ya que se caería en un problema muy común en los algoritmos genéticos que es la convergencia a un valor prematuramente, esto a consecuencia de una falta de diversidad en las poblaciones. La diversidad genética hace referencia a la distribución de las soluciones en el espacio de búsqueda.

Existen distintos métodos para seleccionar a los individuos y entre los más comunes están los siguientes:

- *Selección directa*: Este método consiste en ordenar a los individuos de menor a mayor adaptados para poder seleccionar un número arbitrario de los mejor adaptados, el problema más recurrente de este método es el de la convergencia prematura que podría compensarse con otras técnicas de implementación en el algoritmo.

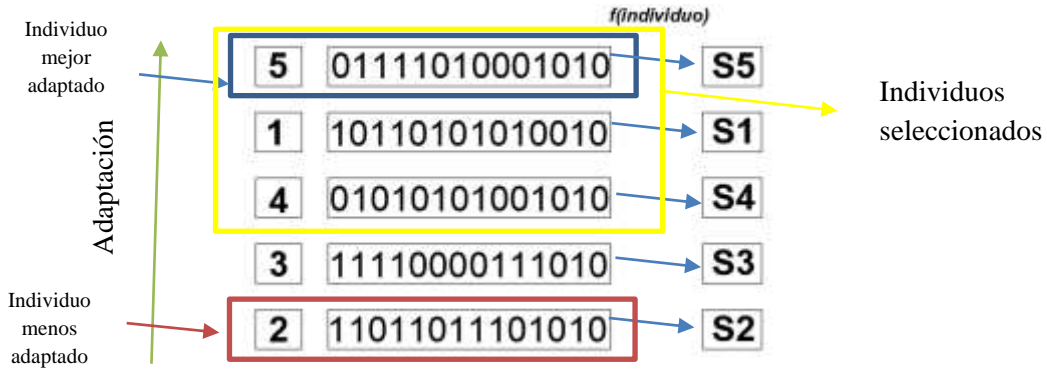


Figura 7 Selección directa

- **Ruleta:** Consiste en asignar a cada individuo una probabilidad de cruce proporcional a la calificación que obtiene de la función ajuste $f(x_i)$. Donde la probabilidad de selección se obtiene de la siguiente forma:

$$P(x) = \frac{f(x_i)}{\sum f(x_i)} \quad (2.21)$$

Con esta información se genera un arreglo que almacenará la suma acumulada de las probabilidades hasta llegar a 1. Posteriormente se crea un vector de números aleatorios entre 0 y 1 de longitud equivalente al número de individuos en la población, estos valores se comparan con los acumulados y se selecciona al individuo si el valor aleatorio se encuentra dentro del rango de este. Un solo individuo puede ser seleccionado más de una vez si así lo determina la ruleta.

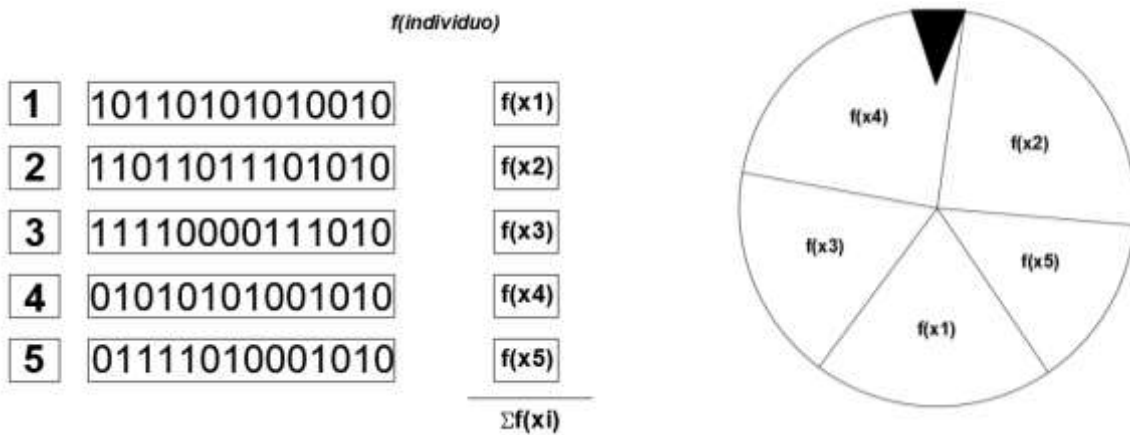


Figura 8 Selección por ruleta

- *Torneo*: Se elige un número determinado de individuos al azar de la población entre ellos y se ponen a competir entre ellos comparando su fitness. El individuo con el fitness más alto pasará a la siguiente generación. Este método tiende a presentar un grado de elitismo ya que proporciona mayor probabilidad de ser elegido a los más adaptados. No tiende a una convergencia prematura y es fácil de implementar computacionalmente.



Figura 9 Selección por torneo

2.4.2.5 Cruza

Una vez que los mejores individuos han sido seleccionados, la *Cruza* tiene como objetivo mezclar el material genético de los padres y transferirlo a sus descendientes, implícitamente hace evolucionar a la población. Al igual que en la naturaleza, en los algoritmos genéticos las reproducciones pueden darse de un modo sexual y asexual. Por lo general la reproducción asexual no es un método recomendable por sí mismo, pues transmite la información genética a los descendientes exactamente igual que la de los padres y no aporta diversidad alguna a la nueva población.

Por otra parte, la reproducción sexual combina el material genético de los progenitores para dar lugar a nuevos individuos, diferentes a sus padres pero conservando la misma tendencia. Con este método se garantiza la diversidad de la población, evolucionando en todas direcciones para obtener tanto individuos mejor dotados como de peor calidad. Existen distintos métodos para realizar la *Cruza* de los individuos entre los cuales por mencionar algunos están, cruce basado en n puntos, cruce uniforme, cruce baricéntrico, etc. Sin embargo, el más utilizado por su sencillez es el *cruce en un punto*.

El *cruce en un punto* consiste en escoger un punto al azar del cromosoma dando lugar a dos secciones que posteriormente nos servirán para hacer la combinación entre los padres.

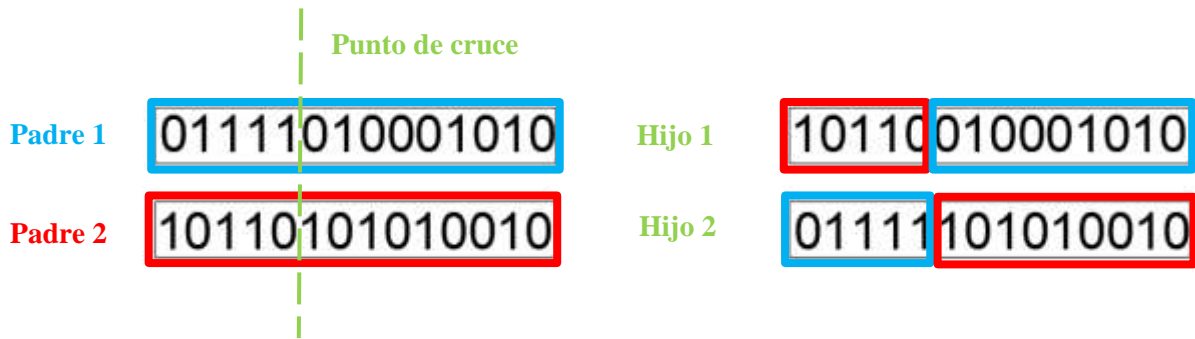


Figura 10 Cruza en un punto

2.4.2.6 Mutación

La mutación es un complemento a la función de selección para insertar cambios inesperados en los individuos y explorar nuevas alternativas para lograr individuos que produzcan nuevas soluciones, garantizando de esta manera que ningún zona del espacio de búsqueda tenga una probabilidad cero de ser explorada. Para lograr este objetivo se define una probabilidad de que un gen, pueda ser mutado. Por ejemplo, si se tiene una población de 10 individuos con una longitud de cromosoma de 10 bits entonces se tienen 100 genes que integran la población y si se considera una probabilidad de mutación del 0.05 se debe cambiar de forma aleatoria a 5 genes de la población, ya sea de 0 a 1 o viceversa.



Figura 11 Mutación de la población

El problema en particular que aqueja a la mutación es la decisión correcta de la probabilidad con la que se va a mutar la población. Un valor demasiado alto haría crecer de forma súbita la diversidad de la población. Convertiría la optimización en una búsqueda meramente aleatoria dificultando la convergencia a un valor final. Por otro lado, un valor demasiado bajo no complementaria a la función de selección para obtener una población diversificada.

Entonces no existe un valor correcto para la probabilidad de mutación aunque la bibliografía sugiere utilizar el valor inverso de la longitud del cromosoma, es un complemento de ambos factores que se puede determinar con base en la experiencia y en el tipo de problemas que se piensa solucionar con los algoritmos genéticos.

2.4.2.7 Criterio de parada del algoritmo genético

El criterio de parada consiste en determinar un umbral para decidir que el algoritmo ha llegado al valor óptimo y se pueda dar por finalizada la búsqueda. Existen distintos métodos para que el algoritmo se detenga el más eficiente es aquel que contempla la convergencia genética de la población y define un criterio de parada basado en su saturación genética. Algunas bibliografías mencionan que un gen a convergido cuando el 95% de su población comparte el mismo valor para un mismo gen. Sin embargo, este método no es del todo confiable pues no se debe perder de vista que el algoritmo es un método pseudoaleatorio de forma que en ocasiones la convergencia genética puede necesitar demasiadas generaciones.

El método más empleado por la facilidad en su programación es definir un máximo de generaciones a emplear. Las desventajas de este método es que si se asigna un número muy pequeño de generaciones este puede no llegar a converger y no alcanzar el objetivo. Por el contrario, si las generaciones son demasiadas el problema puede converger muy rápido y las demás generaciones pueden resultar innecesarias.

En este trabajo se considera un número determinado de generaciones como criterio de término del algoritmo genético, además, considera un número de generaciones lo suficientemente grandes para observar de forma clara la convergencia del algoritmo genético bajo el supuesto de que se contaba con el tiempo y los recursos computacionales suficientes para la ejecución de dicho algoritmo.

2.4.3 Esquema general de un algoritmo genético

- I. Generar la población inicial.
- II. Evaluar la población, a cada individuo se le asigna un valor de aptitud.
- III. Selección de los individuos basada en la aptitud de cada individuo.
- IV. Cruzar a los individuos seleccionados para obtener una pareja de descendientes por cada pareja de padres.
- V. Mutación con una probabilidad determinada de los nuevos individuos (hijos).
- VI. Los individuos forman una nueva población, sustituyendo por completo a la anterior.
- VII. Evaluar la nueva población obtenida.
- VIII. Se repiten los pasos del III al VII hasta alcanzar el número de generaciones predeterminado.

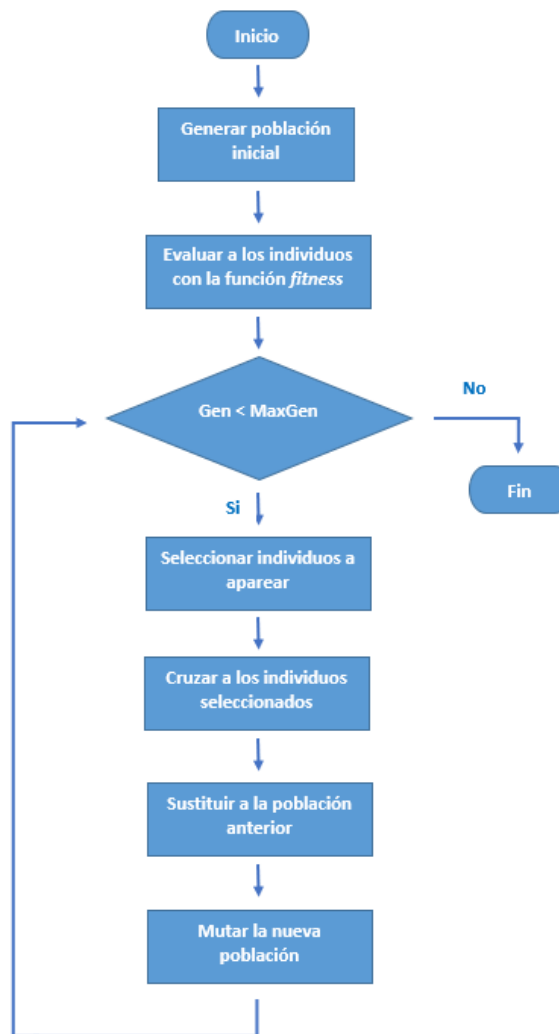


Figura 12 Estructura de ejecución de un algoritmo genético

2.4.4 Aplicación del algoritmo genético en hidrología

El empleo del algoritmo genético como se explicó anteriormente es un método muy complejo, que se puede realizar por una persona sobre papel y llegar a algún resultado después de un cierto periodo de tiempo, sin embargo, perdería su eficacia y no cumpliría con el objetivo de obtener resultados en el momento que se le precisa, es por esto que es necesario la utilización de la programación en alguno de los lenguajes existentes, donde todos estos cientos de procesos iterativos pueden llegar a realizarse en cuestión de segundos.

Un algoritmo genético puede desarrollarse en cualquier lenguaje de programación como lo es C, C++, Matlab, Java, Python, etc. Los factores que determinarán cuál de todos estos lenguajes será el anfitrión del programa serán el conocimiento y la comodidad que sienta el programador al trabajar en alguno de ellos.

En este trabajo de tesis se optó por desarrollar el algoritmo en Python, debido a que en comparación de otros lenguajes cuenta con muchas funciones dedicadas especialmente a cuestiones matemáticas que permiten escribir un código relativamente más pequeño. Su sintaxis es más amigable con el programador y cuenta con una estructura del código que obliga al programador tener un mayor orden en las instrucciones. Además, resulta ser un lenguaje más permisivo en cuanto a la tolerancia de resultados que pueden llegar a ser indeterminados, dicho en otras palabras, si el programa detecta que uno de los valores que arroja el algoritmo es indeterminado este no se detiene para alertar de dicho error, este continúa con los valores que funcionan para el modelo que se está analizando.

Específicamente, el programa se desarrolló en un paquete de desarrollo llamado Anaconda que incluye módulos orientados al procesamiento de datos. Dentro de este se encuentra el entorno de desarrollo Spyder que tiene un intérprete, que funciona como herramienta para escribir y correr programas en Python, brinda formato al código e incluye una interfaz gráfica donde se pueden revisar las instrucciones que están en el código y donde se tiene la posibilidad de visualizar el valor que toman las variables en cada iteración.

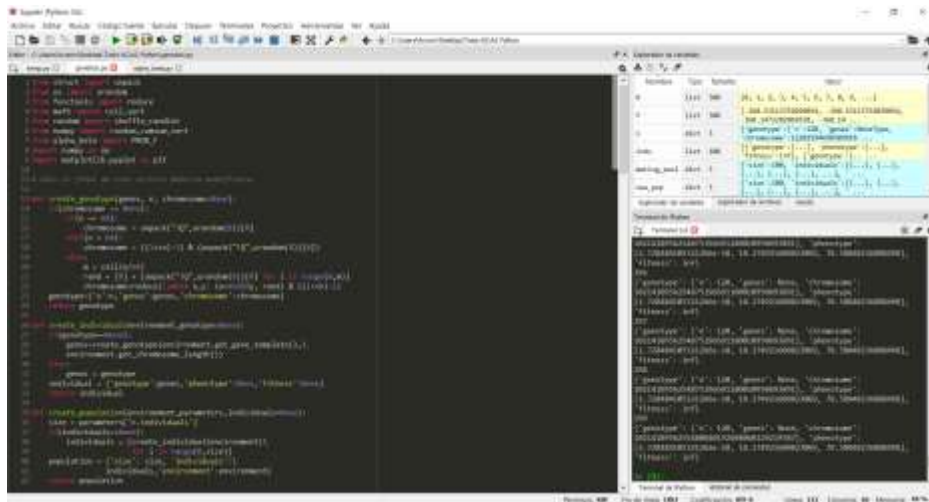


Ilustración 1 Interfaz Spyder Anaconda

En el Anexo 1 se encuentra el código principal del algoritmo genético en el cual se pueden hacer cambios a algunos parámetros de este como lo son el número de individuos, la probabilidad de mutación, el método de selección elegido, el número de generaciones a realizar y por último se incluye una interfaz gráfica interactiva para ver de manera clara cómo se comporta el valor de la función de adaptación en cada generación.

La función de adaptación del algoritmo genético será la función de máxima verosimilitud que se menciona en las ecuaciones 2.3 y 2.4, pues se busca optimizar estas expresiones de tal manera que se consiga un máximo.

$$FO_1 = \text{máx}(L) \quad (2.22)$$

$$FO_2 = \text{máx}[\ln(L)] \quad (2.23)$$

Por otra parte, se utiliza una función objetivo que fue planteada de la siguiente manera:

$$FO_3 = \text{mín} \left(\sum_{j=1}^{np} e_j^2 \right) \quad (2.24)$$

Donde np es el número de parámetros de la función de densidad de probabilidad, e_j son funciones obtenidas como las de la expresión 2.5 que representan el error que se puede cometerse al tratar de igualar a cero las derivadas del logaritmo natural de L .

El código que se muestra en el Anexo 2 contiene el apartado donde es insertada cada una de las funciones objetivo anteriores, así como los intervalos de búsqueda que se le permitirán a cada parámetro. Y en el Anexo 3 se encuentran las funciones objetivo correspondientes a cada función de distribución con la excepción de la función doble Gumbel que solo es posible evaluar su función máxima de verosimilitud y el logaritmo natural de la misma, debido a la complejidad al obtener sus derivadas parciales.

CAPÍTULO 3. DISTRIBUCIONES DE PROBABILIDAD PROPUESTAS

3.1 Normal

La distribución más utilizada para modelar experimentos aleatorios es la distribución normal. Su uso se centra cuando las variables aleatorias X_i son independientes y están idénticamente distribuidas con media μ y desviación estándar σ , representada por $N(\mu, \sigma)$.

Su función de densidad de probabilidad se define como:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (3.1)$$

Donde

$$\begin{aligned} -\infty < x < \infty \\ -\infty < \mu < \infty \\ \sigma > 0 \end{aligned}$$

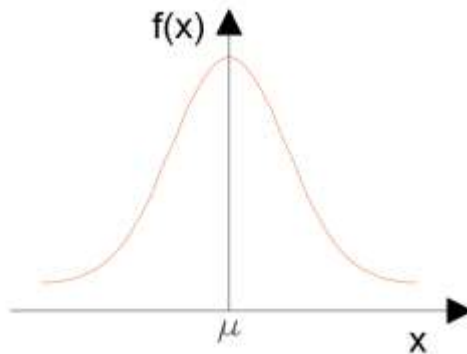


Figura 13 Función de distribución normal

μ y σ son los parámetros de ubicación y de escala respectivamente de la distribución, los cuales determinan la forma de la función $f(x)$ y su posición en el eje x .

Y su función de distribución de probabilidad normal es:

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx \quad (3.2)$$

Si se hace:

$$Z_i = \frac{x_i - \bar{x}}{S} \quad (3.3)$$

Se tendrá una distribución normal con media igual a cero y desviación estándar a uno $N(0,1)$. Teniendo lo siguiente:

$$F(x) = F(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \quad (3.4)$$

Es imposible evaluar la integral de una función de distribución normal sin recurrir a los métodos numéricos. Sin embargo existen métodos que permiten aproximarse a los valores buscados. Como nos lo muestra la aproximación de Abramowitz empleada en hidrología:

Para una probabilidad acumulada $0 < F(x) \leq 0.5$ se tiene que:

$$U_T \approx v - \frac{b_0 + b_1v + b_2v^2}{1 + b_3v + b_4v^2 + b_5v^3} \quad (3.5)$$

Donde

$$b_0 = 2.515517 \quad b_4 = 1.432788$$

$$b_2 = 0.802853 \quad b_5 = 0.189269$$

$$b_3 = 0.010328 \quad b_6 = 0.001308$$

$$v = \sqrt{\ln\left\{\frac{1}{[F(x)]^2}\right\}} \quad (3.6)$$

Para una probabilidad acumulada $0.5 < F(x) \leq 1$ se cambia $F(x)$ por $[1 - F(x)]$ en la expresión (1.14) y el signo al valor de U_T calculado con la ecuación (1.13), aquí $F(x) = 1/T$. (Escalante Sandoval & Reyes Chávez, 2008)

La solución a la integral también se puede obtener mediante tablas que proporciona la bibliografía, donde para cada valor de μ y σ existe un valor de z .

La estimación de valores de la variable aleatoria puede realizarse mediante la siguiente expresión:

$$X_T = \mu + \sigma U_T \quad (3.7)$$

3.1.1 Estimadores por momentos y máxima verosimilitud

$$\mu = \bar{x} \quad (3.8)$$

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \quad (3.9)$$

3.1.2 Estimadores por momentos – L

$$\mu = l_1 \quad (3.10)$$

$$\sigma = 1.772 l_2 \quad (3.11)$$

Donde:

$$l_1 = \beta_0 \quad (3.12)$$

$$l_2 = 2\beta_1 - \beta_0 \quad (3.13)$$

Y a su vez β_0 y β_1 pueden aproximarse como:

$$\beta_0 = M_{(0)} = \frac{1}{n} \sum_{i=1}^n x_i \quad (3.14)$$

$$\beta_1 = M_{(1)} = \frac{1}{n(n-1)} \sum_{i=1}^n x_i(n-i) \quad (3.15)$$

3.1.3 Estimadores por Algoritmos Genéticos

$$L = \prod_{i=1}^n f(x_i; \mu, \sigma) = \frac{1}{\sigma^n (\sqrt{2\pi})^n} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2} \quad (3.16)$$

$$\ln L = -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 - \ln [\sigma^n (\sqrt{2\pi})^n] \quad (3.17)$$

Derivando parcialmente el $\ln L$ respecto a μ

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2} \left(\sum_{i=1}^n x_i - n\mu \right) \quad (3.18)$$

Igualando a cero se obtiene la función

$$e_1 = \sum_{i=1}^n x_i - n\mu \quad (3.19)$$

Derivando parcialmente el $\ln L$ respecto a σ

$$\frac{\partial \ln L}{\partial \sigma} = \frac{1}{\sigma^3} \left(\sum_{i=1}^n (x_i - \mu)^2 \right) - \frac{n}{\sigma} \quad (3.20)$$

Igualando a cero se obtiene la función

$$e_2 = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 - n \quad (3.21)$$

3.2 Log Normal 2 parámetros

En esta función de distribución los logaritmos naturales de la variable aleatoria se distribuyen normalmente:

La función de densidad está distribuida como:

$$f(x) = \frac{1}{x\sigma_y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln x - \mu_y}{\sigma_y}\right)^2} \quad (3.22)$$

Donde $x > 0$

μ_y y σ_y son la media y la desviación estándar del logaritmo natural de x . Una justificación teórica para el uso de esta distribución es que los factores para muchas de las variables hidrológicas actúan multiplicativamente mejor que aditivamente y es por esta razón que el logaritmo de esos factores satisfacen las condiciones básicas de la distribución normal.

En la *Figura 3.2* se observa una gráfica de la función densidad de probabilidad para diferentes valores de μ_y y σ_y . Como se observa, esta función no necesariamente es simétrica.

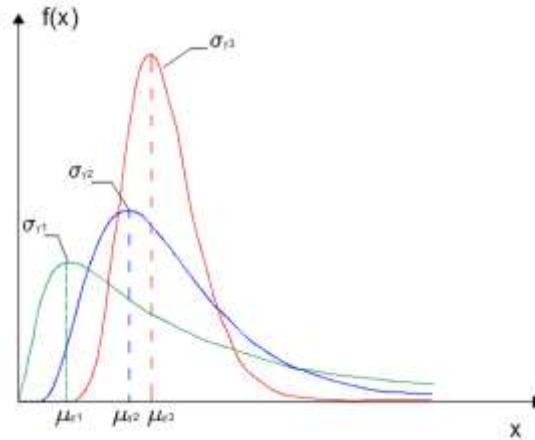


Figura 14 Función de distribución lognormal

La función de distribución de probabilidad es:

$$F(x) = \int_0^x \frac{1}{x\sigma_y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln x - \mu_y}{\sigma_y}\right)^2} dx \quad (3.23)$$

Los valores de la función de distribución de probabilidad se pueden obtener mediante tablas si la variable estandarizada se define como:

$$Z_i = \frac{\ln x - \mu_y}{\sigma_y} \quad (3.24)$$

La estimación de valores de la variable aleatoria puede realizarse mediante la siguiente expresión:

$$X_T = e^{\mu_y + U_T * \sigma_y} \quad (3.25)$$

3.2.1 Estimadores por momentos y máxima verosimilitud

$$\mu_y = \frac{1}{n} \sum_{i=1}^n \ln x_i \quad (3.26)$$

$$\sigma_y^2 = \frac{1}{n} \sum_{i=1}^n (\ln x_i - \mu_y)^2 \quad (3.27)$$

3.2.2 Estimadores por Algoritmos Genéticos

$$L = \prod_{i=1}^n f(x_i; \mu_y, \sigma_y) = \frac{1}{\sigma_y^n (\sqrt{2\pi})^n \prod_{i=1}^n x_i} e^{-\frac{1}{2\sigma_y^2} \sum_{i=1}^n (\ln x_i - \mu_y)^2} \quad (3.28)$$

$$\ln L = -\frac{1}{2\sigma_y^2} \sum_{i=1}^n (\ln x_i - \mu_y)^2 - \ln \left[\prod_{i=1}^n x_i \right] - \ln (\sigma_y^n (\sqrt{2\pi})^n) \quad (3.29)$$

Derivando parcialmente el $\ln L$ respecto a μ_y

$$\frac{\partial \ln L}{\partial \mu_y} = \frac{1}{\sigma_y^2} \left(\sum_{i=1}^n \ln x_i - n\mu_y \right) \quad (3.30)$$

Igualando a cero se obtiene la función

$$e_1 = \sum_{i=1}^n \ln x_i - n\mu_y \quad (3.31)$$

Derivando parcialmente el $\ln L$ respecto a σ_y

$$\frac{\partial \ln L}{\partial \sigma_y} = \frac{1}{\sigma_y^3} \sum_{i=1}^n (\ln x_i - \mu_y)^2 - \frac{n}{\sigma_y} \quad (3.32)$$

Igualando a cero se obtiene la función

$$e_2 = \frac{1}{\sigma_y^2} \sum_{i=1}^n (\ln x_i - \mu_y)^2 - n \quad (3.33)$$

3.3 Log Normal 3 parámetros

Tal como la distribución log normal representa la distribución normal de los logaritmos de la variable x , entonces la distribución log normal representa la distribución normal de los logaritmos de la variable reducida $(x - x_0)$ donde x_0 es el límite inferior. (Kite, 1978)

La función de densidad está distribuida como:

$$f(x) = \frac{1}{(x - x_0)\sigma_y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(x-x_0)-\mu_y}{\sigma_y}\right)^2} \quad (3.34)$$

Donde $x > x_0$

μ_y y σ_y^2 son la media y la varianza de los logaritmos de $(x - x_0)$ y a su vez los parámetros de forma y escala.

Si el límite inferior x_0 es conocido entonces la variable reducida $(x - x_0)$ puede ser utilizado para el procedimiento descrito en la distribución log normal de 2 parámetros, si x_0 es desconocido entonces la estimación de este parámetro precede la estimación de los parámetros μ_y y σ_y^2 .

La función de distribución de probabilidad es:

$$F(x) = \int_{x_0}^x \frac{1}{(x - x_0)\sigma_y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(x-x_0)-\mu_y}{\sigma_y}\right)^2} dx \quad (3.35)$$

La estimación de valores de la variable aleatoria puede realizarse mediante la siguiente expresión:

$$X_T = x_0 + e^{\mu_y + U_T \sigma_y} \quad (3.36)$$

3.3.1 Estimadores por momentos

$$x_0 = \bar{x} \left(1 - \frac{n_x}{n_z}\right) \quad (3.37)$$

Donde

$$n_x = \frac{S}{\bar{x}} \quad (3.38)$$

$$n_z = \frac{1 - W^{2/3}}{W^{1/3}} \quad (3.39)$$

$$W = \frac{(g^2 + 4)^{1/2} - g}{2} \quad (3.40)$$

Para el cálculo de g revise la expresión (1.6) para datos muestrales

$$\mu_y = \ln\left(\frac{S}{n_z}\right) - \frac{1}{2} \ln(n_z^2 + 1) \quad (3.41)$$

$$\sigma_y = \sqrt{\ln(n_z^2 + 1)} \quad (3.42)$$

3.3.2 Estimadores por máxima verosimilitud

$$\mu_y = \frac{1}{n} \sum_{i=1}^n \ln(x_i - x_0) \quad (3.43)$$

$$\sigma_y^2 = \frac{1}{n} \sum_{i=1}^n [\ln(x_i - x_0) - \mu_y]^2 \quad (3.44)$$

$$F(x_0) = \sum_{i=1}^n \frac{1}{(x_i - x_0)} (\mu_y - \sigma_y^2) - \sum_{i=1}^n \left\{ \left[\frac{1}{(x_i - x_0)} \right] \ln(x_i - x_0) \right\} \quad (3.45)$$

3.3.3 Estimadores por Algoritmos Genéticos

$$L = \prod_{i=1}^n f(x_i; \mu_y, \sigma_y, x_0) = \frac{1}{\sigma_y^n (\sqrt{2\pi})^n \prod_{i=1}^n (x_i - x_0)} e^{-\frac{1}{2\sigma_y^2} \sum_{i=1}^n [\ln(x_i - x_0) - \mu_y]^2} \quad (3.46)$$

$$\ln L = -\frac{1}{2\sigma_y^2} \sum_{i=1}^n [\ln(x_i - x_0) - \mu_y]^2 - \ln \left[\prod_{i=1}^n (x_i - x_0) \right] - \ln [\sigma_y^n (\sqrt{2\pi})^n] \quad (3.47)$$

Derivando parcialmente el $\ln L$ respecto a μ_y

$$\frac{\partial \ln L}{\partial \mu_y} = \frac{1}{\sigma_y^2} \left(\sum_{i=1}^n \ln(x_i - x_0) - n\mu_y \right) \quad (3.48)$$

Igualando a cero se obtiene la función

$$e_1 = \sum_{i=1}^n \ln(x_i - x_0) - n\mu_y \quad (3.49)$$

Derivando parcialmente el $\ln L$ respecto a σ_y

$$\frac{\partial \ln L}{\partial \sigma_y} = \frac{1}{\sigma_y^3} \sum_{i=1}^n (\ln(x_i - x_0) - \mu_y)^2 - \frac{n}{\sigma_y} \quad (3.50)$$

Igualando a cero se obtiene la función:

$$e_2 = \frac{1}{\sigma_y^2} \sum_{i=1}^n (\ln(x_i - x_0) - \mu_y)^2 - n \quad (3.51)$$

Derivando parcialmente el $\ln L$ respecto a x_0 e igualando a cero se obtiene la función

$$e_3 = \sum_{i=1}^n \frac{1}{(x_i - x_0)} (\mu_y - \sigma_y^2) - \sum_{i=1}^n \left\{ \left[\frac{1}{(x_i - x_0)} \right] \ln(x_i - x_0) \right\} \quad (3.52)$$

3.4 Gumbel

La distribución Gumbel se emplea para analizar valores extremos, ya sean máximos o mínimos. En el caso de la hidrología aplicada se utiliza para el análisis de gastos o precipitaciones máximas anuales. Su función de densidad de probabilidad está dada por:

$$f(x) = \alpha e^{-\alpha(x-\beta)} - e^{-\alpha(x-\beta)} \quad (3.53)$$

Donde

$$-\infty < x < \infty$$

$$\alpha > 0$$

α y β son parámetros de la función.

La función de distribución de probabilidad es:

$$F(x) = e^{-e^{-\left(\frac{x-\beta}{\alpha}\right)}} \quad (3.54)$$

La estimación de valores de la variable aleatoria puede realizarse mediante la siguiente expresión:

$$X_T = \beta - \alpha \ln \left(-\ln \left(1 - \frac{1}{T} \right) \right) \quad (3.55)$$

3.4.1 Estimador por momentos

$$\beta = \bar{x} - 0.45S \quad (3.56)$$

$$\alpha = 0.78S \quad (3.57)$$

3.4.2 Estimador por máxima verosimilitud

$$P = n - \sum_{i=1}^n e^{-y_i} \quad (3.58)$$

$$R = n - \sum_{i=1}^n y_i + \sum_{i=1}^n y_i e^{-y_i} \quad (3.59)$$

$$y_i = \frac{x_i - \hat{\nu}}{\hat{\alpha}} \quad (3.60)$$

El criterio de convergencia es:

$$\frac{P}{\hat{\alpha}} \approx 0, -\frac{R}{\hat{\alpha}} \approx 0 \quad (3.61)$$

Incrementos:

$$\delta_{\hat{\nu}_j} = (1.11P_j - 0.26R_j) \frac{\hat{\alpha}_j}{n} \quad (3.62)$$

$$\delta_{\hat{\alpha}_j} = (0.26P_j - 0.61R_j) \frac{\hat{\alpha}_j}{n} \quad (3.63)$$

nuevos valores::

$$\hat{\nu}_{j+1} = \hat{\nu}_j + \delta_{\hat{\nu}_j} \quad (3.64)$$

$$\hat{\alpha}_{j+1} = \hat{\alpha}_j + \delta_{\hat{\alpha}_j} \quad (3.65)$$

3.4.3 Estimador por momentos – L

$$\beta = \lambda_1 - 0.57216\alpha \quad (3.66)$$

$$\alpha = \frac{\lambda_2}{\ln(2)} \quad (3.67)$$

3.4.4 Estimador por algoritmos Genéticos

$$L = \prod_{i=1}^n f(x_i; \alpha, \beta) = \alpha^n e^{-\alpha \sum_{i=1}^n (x_i - \beta) - \sum_{i=1}^n e^{-\alpha(x_i - \beta)}} \quad (3.68)$$

$$\ln L = n \ln \alpha - \alpha \sum_{i=1}^n x_i + n\alpha\beta - \sum_{i=1}^n e^{-\alpha(x_i-\beta)} \quad (3.69)$$

Derivando parcialmente el $\ln L$ respecto a α

$$\frac{\partial \ln L}{\partial \alpha} = \frac{n}{\alpha} - \sum_{i=1}^n x_i - n\beta + \sum_{i=1}^n (x_i - \beta) e^{-\alpha(x_i-\beta)} \quad (3.70)$$

Igualando a cero se obtiene la función

$$e_1 = \sum_{i=1}^n x_i e^{-\alpha x_i} - \left(\bar{x} - \frac{1}{\alpha} \right) \sum_{i=1}^n e^{-\alpha x_i} \quad (3.71)$$

Derivando parcialmente el $\ln L$ respecto a β

$$\frac{\partial \ln L}{\partial \beta} = \alpha n - \alpha e^{\alpha\beta} \sum_{i=1}^n e^{-\alpha x_i} \quad (3.72)$$

Igualando a cero se obtiene la función

$$e_2 = \beta - \frac{1}{\alpha} \ln \left[\frac{n}{\sum_{i=1}^n e^{-\alpha x_i}} \right] \quad (3.73)$$

3.5 Doble Gumbel

Esta distribución es utilizada para realizar análisis hidrológicos en zonas donde se presentan dos poblaciones diferentes para una misma zona de estudio. Por ejemplo, en zonas costeras de México la primera población estaría referida a eventos meteorológicos no ciclónicos, mientras que la segunda población comprendería a los gastos producidos por las precipitaciones ciclónicas. La función de distribución de probabilidad está dada por la siguiente expresión:

Su función de densidad de probabilidad es:

$$f(x) = pe^{-\alpha_1(x-\beta_1)-e^{-\alpha_1(x-\beta_1)}} + (1-p)e^{-\alpha_2(x-\beta_2)-e^{-\alpha_2(x-\beta_2)}} \quad (3.74)$$

Para $x > 0$, $\alpha_i > 0$ y $0 < p < 1$

Donde:

- v_1 parámetro de ubicación de la primera población
- α_1 parámetro de escala de la primera población
- v_2 parámetro de ubicación de la segunda población
- α_2 parámetro de escala de la segunda población
- β_2 parámetro de forma de la segunda población
- p parámetro de asociación

La función de distribución de probabilidad está dada por:

$$F(x) = pe^{-e^{-\alpha_1(x-\beta_1)}} + (1-p)e^{-e^{-\alpha_2(x-\beta_2)}} \quad (3.75)$$

3.5.1 Estimador por máxima verosimilitud y algoritmos genéticos

$$L = \prod_{i=1}^n f(x_i; p, \alpha_1, \beta_1, \alpha_2, \beta_2) = \prod_{i=1}^n \left(p e^{-\alpha_1(x_i - \beta_1) - e^{-\alpha_1(x_i - \beta_1)}} + (1 - P) e^{-\alpha_2(x_i - \beta_2) - e^{-\alpha_2(x_i - \beta_2)}} \right) \quad (3.76)$$

$$\ln L = \ln \left[\prod_{i=1}^n f(x_i; p, \alpha_1, \beta_1, \alpha_2, \beta_2) \right] = \ln \left[\prod_{i=1}^n \left(p e^{-\alpha_1(x_i - \beta_1) - e^{-\alpha_1(x_i - \beta_1)}} + (1 - P) e^{-\alpha_2(x_i - \beta_2) - e^{-\alpha_2(x_i - \beta_2)}} \right) \right] \quad (3.77)$$

CAPÍTULO 4. APLICACIÓN A DISTINTOS TIPOS DE DATOS HIDROLÓGICOS

A continuación se aplicará la teoría de los algoritmos genéticos con las funciones de distribución empleadas en hidrología a dos conjuntos de datos que corresponden a los gastos máximos anuales de un día de las estaciones hidrométricas 28039 “Paso del toro” localizada en el río Cotaxtla y 28040 “El tejar” localizada en el río Jamapa, ambas ubicadas en el estado de Veracruz. Esta información fue proporcionada por el Instituto de Ingeniería de la UNAM. (Copca Maya, 2017)

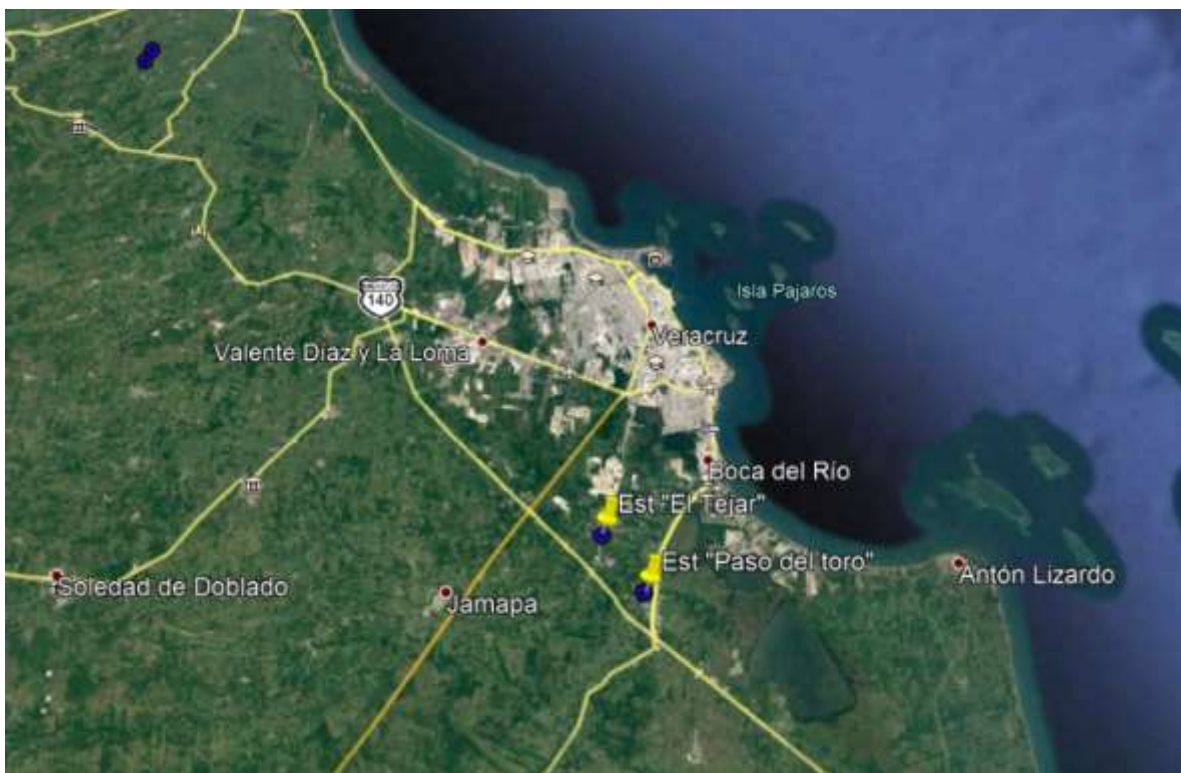


Ilustración 2 Ubicación de las estaciones hidrométricas, Google Earth

Se evaluarán ambos conjuntos de datos para las cinco funciones de distribución antes mencionadas, cada una maximizando las funciones L , $\ln(L)$, y la función que minimiza el error al igualar las derivadas parciales a cero con la excepción de la función doble Gumbel. Se evaluarán los parámetros que el algoritmo arroje y se determinará mediante el error cuadrático cuál de estos brinda un mejor ajuste a los datos observados.

El algoritmo genético tendrá como características los siguientes valores:

Longitud de cromosoma de cada individuo:	40
Número de individuos en la población inicial:	500
Probabilidad de mutación:	0.02
Método de selección:	Torneo
Método de cruce:	Un punto
Número de generaciones:	1000

Con ayuda del Programa AX (Jiménez, 1996) desarrollado por el CENAPRED se compararán los datos obtenidos con los que resultan de los otros métodos de estimación de parámetros presentados en este mismo trabajo que tienen como base, algoritmos determinísticos. Además el algoritmo genético se apoyará en dichos valores ante la incertidumbre de que el uso de la primera derivada pueda llevar a un resultado que corresponda a un valor alejado del máximo global. Para cada parámetro se utilizará un intervalo de búsqueda amplio y otro reducido.

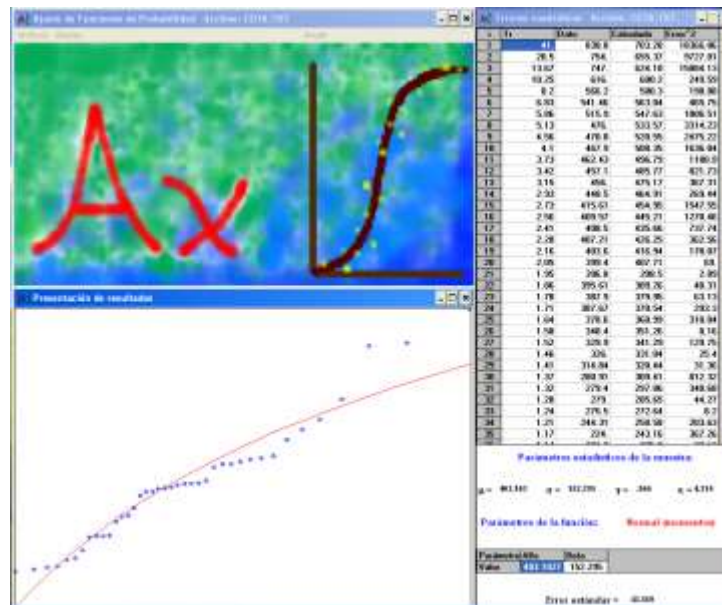


Ilustración 3 Programa AX (Jiménez, 1996)

Por último, se observara la diferencia que se presenta en el error cuadrático como resultado de la obtención del periodo de retorno como lo indican los autores Weibull y Gringorten, este último con los valores requeridos de la Tabla 1.

4.1 Estación hidrométrica del río Cotaxtla

En la Tabla 2 se muestran los gastos máximos anuales que corresponden a esta estación.

Año	Q (m ³ /s)	Año	Q (m ³ /s)	Año	Q (m ³ /s)	Año	Q (m ³ /s)
1952	456.00	1962	329.90	1972	747.00	1987	244.31
1953	279.40	1963	275.50	1973	396.80	1988	395.61
1954	467.90	1964	204.30	1974	838.80	1991	407.21
1955	566.20	1965	201.20	1975	754.00	1992	314.84
1956	348.40	1966	378.60	1976	476.00	1993	409.57
1957	326.00	1967	387.90	1977	224.00	1994	190.69
1958	470.80	1968	221.20	1978	279.00	1998	387.67
1959	399.40	1969	616.00	1979	408.50	2000	462.43
1960	457.10	1970	403.60	1980	515.90	2003	415.61
1961	448.50	1971	195.90	1986	280.91	2005	541.46

Tabla 2 Gastos máximos anuales de la estación "Paso del toro"

4.1.1 Distribución Normal

Intervalo	AX	Algoritmo genético						Rango	
		máx (L)		máx [ln(L)]		mín (e ₁ ² + e ₂ ²)		Amplio	Reducido
		Amplio	Reducido	Amplio	Reducido	Amplio	Reducido		
$\mu =$	403.102	403.103	403.103	403.103	403.103	403.103	403.103	0 - 4000	350 - 450
$s =$	152.295	150.379	150.379	150.379	150.379	150.379	150.379	0 - 4000	100 - 200
F.O. =		1.83E-112	1.83E-112	-257.43	-257.43	-1.87E-15	-1.87E-15		
EE Weibul =	43.569	43.853		43.853		43.853			
EE Gringorten=	42.877	42.806		42.806		42.806			

Tabla 3 Resultados obtenidos de la distribución Normal para estación "Paso del toro"

En la siguiente tabla se muestran los valores calculados con esta función de distribución con periodos de retorno calculados para Weibull y Gringorten.

Weibull			Gringorten		
Tr (años)	Q (m ³ /s)		Tr (años)	Q (m ³ /s)	
	AX	AG		AX	AG
1.025	102.927	107.016	1.016	74.983	79.101
1.051	150.832	154.146	1.042	137.351	140.686
1.079	182.023	184.834	1.070	172.957	175.845
1.108	206.004	208.431	1.099	199.138	201.698
1.139	225.904	228.016	1.130	220.379	222.672
1.171	243.164	245.010	1.162	238.557	240.622
1.206	258.580	260.192	1.197	254.650	256.513
1.242	272.637	274.040	1.234	269.235	270.915
1.281	285.654	286.872	1.273	282.684	284.196
1.323	297.858	298.907	1.314	295.253	296.606
1.367	309.412	310.307	1.359	307.125	308.329
1.414	320.440	321.195	1.406	318.438	319.501
1.464	331.040	331.668	1.457	329.300	330.227
1.519	341.291	341.803	1.512	339.796	340.591
1.577	351.257	351.663	1.571	349.997	350.664
1.640	360.992	361.304	1.635	359.962	360.503
1.708	370.544	370.772	1.704	369.740	370.159
1.783	379.955	380.109	1.779	379.379	379.677
1.864	389.261	389.354	1.861	388.918	389.096
1.952	398.499	398.540	1.952	398.395	398.455
2.050	407.707	407.666	2.051	407.810	407.751
2.158	416.945	416.852	2.161	417.288	417.110
2.278	426.251	426.097	2.284	426.827	426.529
2.412	435.662	435.434	2.421	436.465	436.047
2.563	445.214	444.902	2.576	446.244	445.703
2.733	454.949	454.543	2.752	456.208	455.542
2.929	464.915	464.403	2.954	466.409	465.615
3.154	475.166	474.538	3.188	476.905	475.979
3.417	485.766	485.011	3.462	487.767	486.705
3.727	496.794	495.899	3.788	499.081	497.877
4.100	508.348	507.299	4.182	510.953	509.600
4.556	520.552	519.334	4.667	523.521	522.010
5.125	533.570	532.166	5.279	536.970	535.291
5.857	547.626	546.014	6.075	551.555	549.693
6.833	563.042	561.196	7.156	567.648	565.584
8.200	580.303	578.190	8.703	585.827	583.534
10.250	600.202	597.775	11.103	607.067	604.508
13.667	624.183	621.372	15.333	633.249	630.361
20.500	655.375	652.060	24.769	668.854	665.520
41.000	703.279	699.190	64.400	731.222	727.105

Tabla 4 Valores calculados. Función Normal – Weibull y Gringorten

En la interfaz de Spyder es posible ver de manera gráfica el comportamiento del algoritmo genético. En el eje de las abscisas se tiene el número de la generación, mientras que en el eje de las ordenadas se tiene el valor de la función objetivo. Sin embargo, al manejar valores muy cercanos a cero la interfaz los interpreta como tal y no es posible visualizar todos los casos. El único caso que es muy evidente a simple vista es el logaritmo natural de la función de máxima verosimilitud como se muestra en la siguiente ilustración:

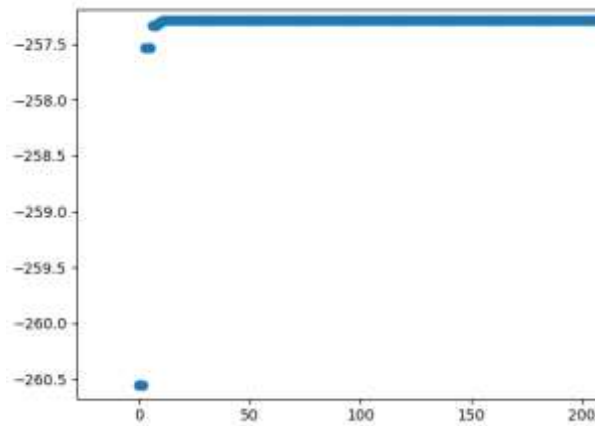


Ilustración 4 Gráfica del $\ln(L)$ en cada generación para la función de distribución Normal

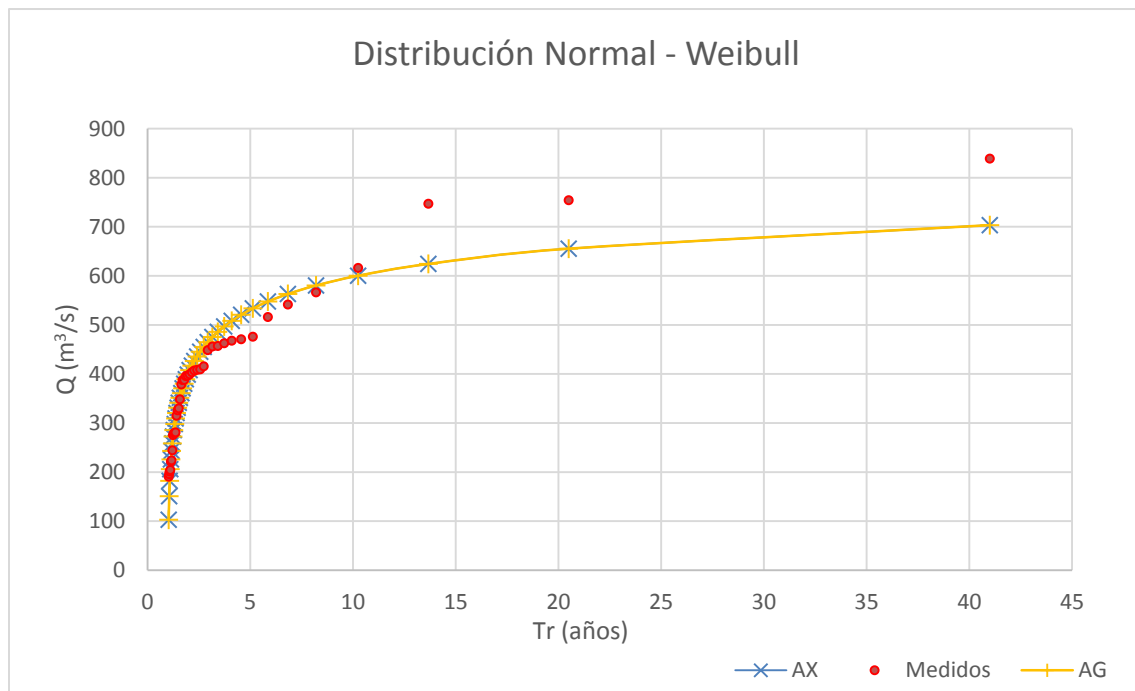


Figura 15 Comparación entre valores históricos y calculados. Función Normal – Weibull

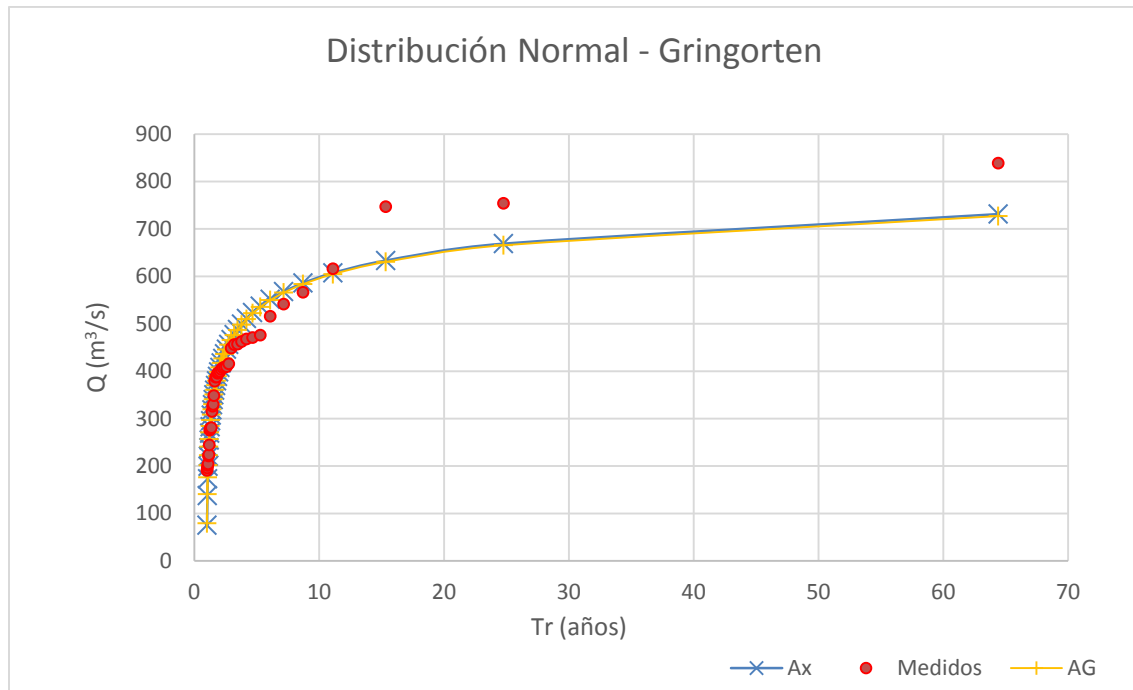


Figura 16 Comparación entre valores históricos y calculados. Función Normal – Gringorten

4.1.2 Distribución Log Normal 2 Parámetros

Intervalo	AX	Algoritmo genético						Rango	
		máx (L)		máx [ln(L)]		mín ($e_1^2 + e_2^2$)		Amplio	Reducido
		Amplio	Reducido	Amplio	Reducido	Amplio	Reducido		
$\mu_y =$	5.932	5.932	5.932	5.932	5.932	5.932	5.932	0 - 4000	0 - 10
$s_y =$	0.3678	0.3678	0.3678	0.3678	0.3678	0.3678	0.3678	0 - 4000	0 - 1
F.O. =		4.72E-111	4.72E-111	-254.036	-254.036	-7.28E-16	-7.28E-16		
EE Weibul =	29.868	29.868		29.868		29.868			
EE Gringorten =	25.762	25.762		25.762		25.762			

Tabla 5 Resultados obtenidos de la distribución Log Normal 2P para estación “Paso del toro”

Weibull			Gringorten		
Tr (años)	Q (m3/s)		Tr (años)	Q (m3/s)	
	AX	AG		AX	AG
1.025	182.696		1.016	170.639	
1.051	205.018		1.042	198.378	
1.079	220.998		1.070	216.192	
1.108	234.128		1.099	230.304	
1.139	245.616		1.130	242.426	
1.171	256.040		1.162	253.307	
1.206	265.726		1.197	263.346	
1.242	274.881		1.234	272.788	
1.281	283.644		1.273	281.793	
1.323	292.117		1.314	290.478	
1.367	300.377		1.359	298.927	
1.414	308.484		1.406	307.208	
1.464	316.487		1.457	315.373	
1.519	324.430		1.512	323.470	
1.577	332.350		1.571	331.538	
1.640	340.280		1.635	339.614	
1.708	348.252		1.704	347.730	
1.783	356.296		1.779	355.919	
1.864	364.443		1.861	364.214	
1.952	372.725		1.952	372.647	
2.050	381.138		2.051	381.217	
2.158	389.798		2.161	390.044	
2.278	398.712		2.284	399.134	
2.412	407.922		2.421	408.534	
2.563	417.478		2.576	418.297	
2.733	427.439		2.752	428.485	
2.929	437.873		2.954	439.173	
3.154	448.863		3.188	450.448	
3.417	460.509		3.462	462.421	
3.727	472.937		3.788	475.230	
4.100	486.310		4.182	489.053	
4.556	500.837		4.667	504.126	
5.125	516.804		5.279	520.769	
5.857	534.609		6.075	539.440	
6.833	554.833		7.156	560.819	
8.200	578.380		8.703	585.990	
10.250	606.760		11.103	616.835	
13.667	642.809		15.333	657.097	
20.500	692.913		24.769	716.103	
41.000	777.571		64.400	832.514	

Tabla 6 Valores calculados. Función Log Normal 2P – Weibull y Gringorten

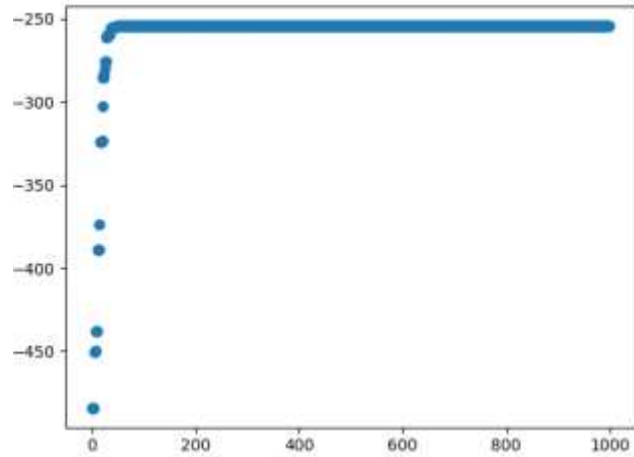


Ilustración 5 Gráfica del $\ln(L)$ en cada generación para la función de distribución Log Normal 2P

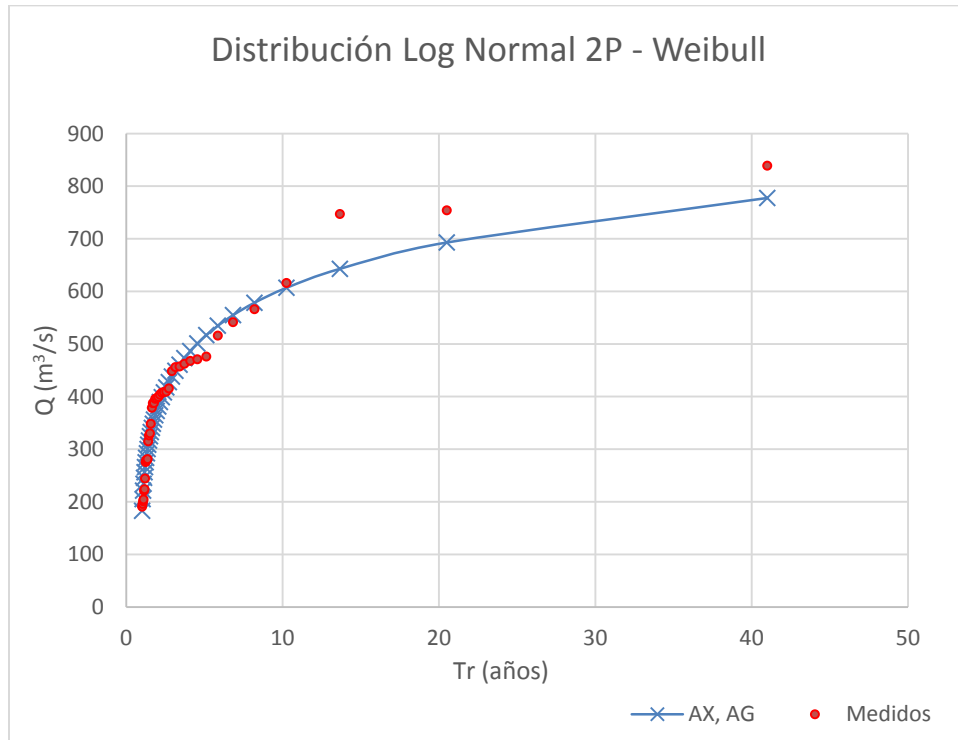


Figura 17 Comparación entre valores históricos y calculados. Función Log Normal 2P - Weibull

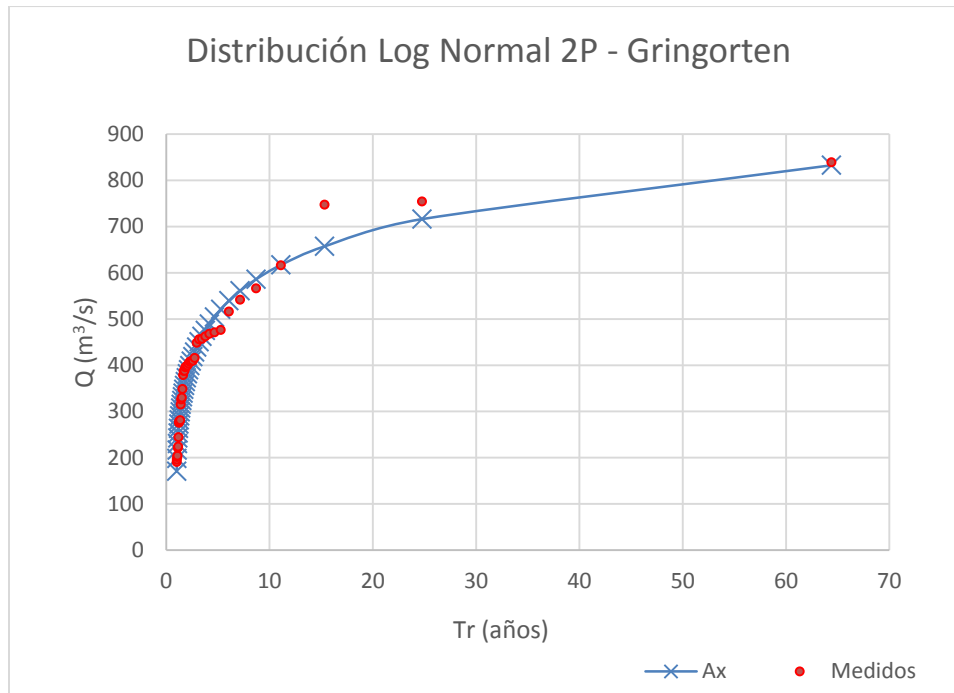


Figura 18 Comparación entre valores históricos y calculados. Función Log Normal 2P – Gringorten

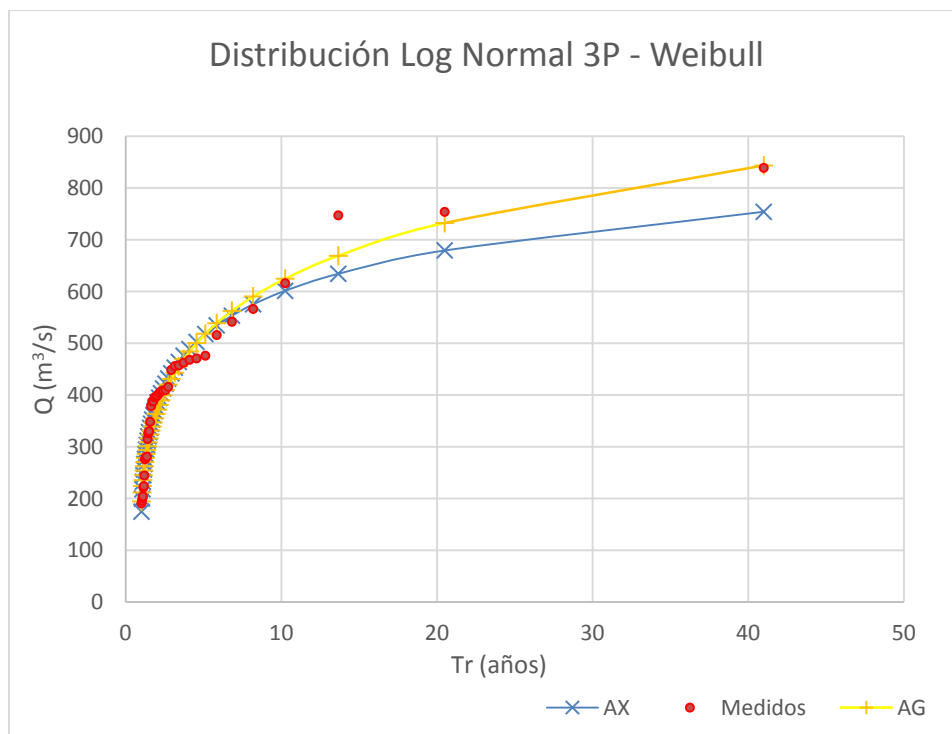
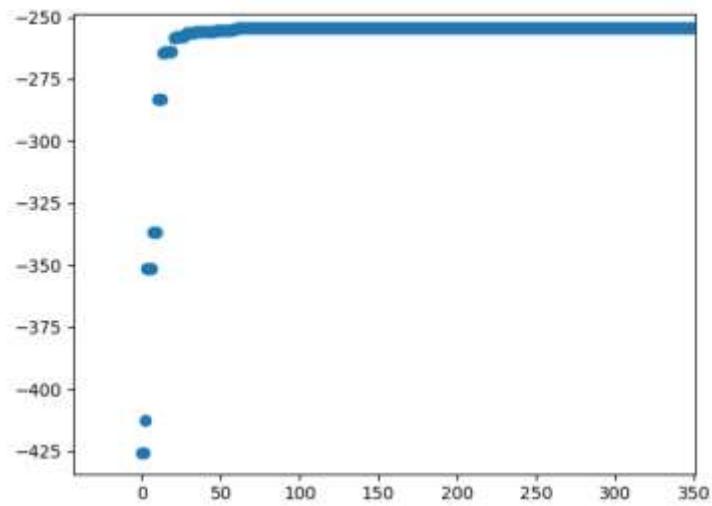
4.1.3 Distribución Log Normal 3 Parámetros

Intervalo	AX	Algoritmo genético						Rango	
		máx (L)		máx [ln(L)]		mín ($e_1^2 + e_2^2 + e_3^2$)		Amplio	Reducido
		Amplio	Reducido	Amplio	Reducido	Amplio	Reducido		
$\mu_y =$	6.1457	5.608	6.111	5.608	6.111	5.84	6.1342	0 - 1000	0 - 10
$s_y =$	0.2979	0.513	0.3081	0.513	0.3081	0.4033	0.3013	0 - 1000	0 - 1
$x_0 =$	-85.02	95	-70	95	-70	30.739	-79.918	-190 - 190	-100 - -70
F.O. =		4.40E-111	4.30E-111	-254.08	-254.13	-3.89E-08	-3.41E-08		
EE Weibul =	30.025	27.225	32.143	27.225	32.143	28.815	32.349		
EE Gringorten =	27.896	29.641	27.581	29.641	27.581	25.863	27.776		

Tabla 7 Resultados obtenidos de la distribución Log normal 3P para estación “Paso del toro”

Weibull			Gringorten		
Tr (años)	Q (m3/s)		Tr (años)	Q (m3/s)	
	AX	AG		AX	AG
1.025	174.583	194.279	1.016	160.617	174.919
1.051	199.988	211.596	1.042	192.489	200.812
1.079	217.852	224.464	1.070	212.506	217.629
1.108	232.346	235.316	1.099	228.141	231.046
1.139	244.901	245.011	1.130	241.427	242.637
1.171	256.197	253.965	1.162	253.244	253.087
1.206	266.615	262.415	1.197	264.062	262.768
1.242	276.395	270.514	1.234	274.165	271.906
1.281	285.7	278.367	1.273	283.74	280.65
1.323	294.644	286.052	1.314	292.918	289.108
1.367	303.316	293.629	1.359	301.798	297.36
1.414	311.784	301.145	1.406	310.454	305.469
1.464	320.102	308.643	1.457	318.947	313.486
1.519	328.318	316.159	1.512	327.327	321.456
1.577	336.471	323.725	1.571	335.637	329.416
1.640	344.598	331.372	1.635	343.917	337.403
1.708	352.732	339.132	1.704	352.201	345.448
1.783	360.904	347.033	1.779	360.523	353.584
1.864	369.146	355.108	1.861	368.914	361.844
1.952	377.486	363.388	1.952	377.408	370.259
2.050	385.924	371.875	2.051	386.004	378.831
2.158	394.573	380.689	2.161	394.818	387.678
2.278	403.436	389.842	2.284	403.855	396.81
2.412	412.555	399.385	2.421	413.16	406.274
2.563	421.976	409.376	2.576	422.781	416.126
2.733	431.752	419.888	2.752	432.776	426.431
2.929	441.946	431.002	2.954	443.212	437.266
3.154	452.632	442.822	3.188	454.17	448.724
3.417	463.903	455.473	3.462	465.749	460.923
3.727	475.872	469.115	3.788	478.074	474.006
4.100	488.683	483.951	4.182	491.303	488.164
4.556	502.525	500.252	4.667	505.648	503.645
5.125	517.651	518.384	5.279	521.394	520.791
5.857	534.414	538.867	6.075	538.944	540.09
6.833	553.327	562.461	7.156	558.899	562.267
8.200	575.182	590.361	8.703	582.209	588.481
10.250	601.3	624.588	11.103	610.516	620.753
13.667	634.144	668.982	15.333	647.065	663.113
20.500	679.221	732.33	24.769	699.872	725.643
41.000	754.01	843.495	64.400	801.715	850.438

Tabla 8 Valores calculados. Función Log Normal 3P – Weibull y Gringorten



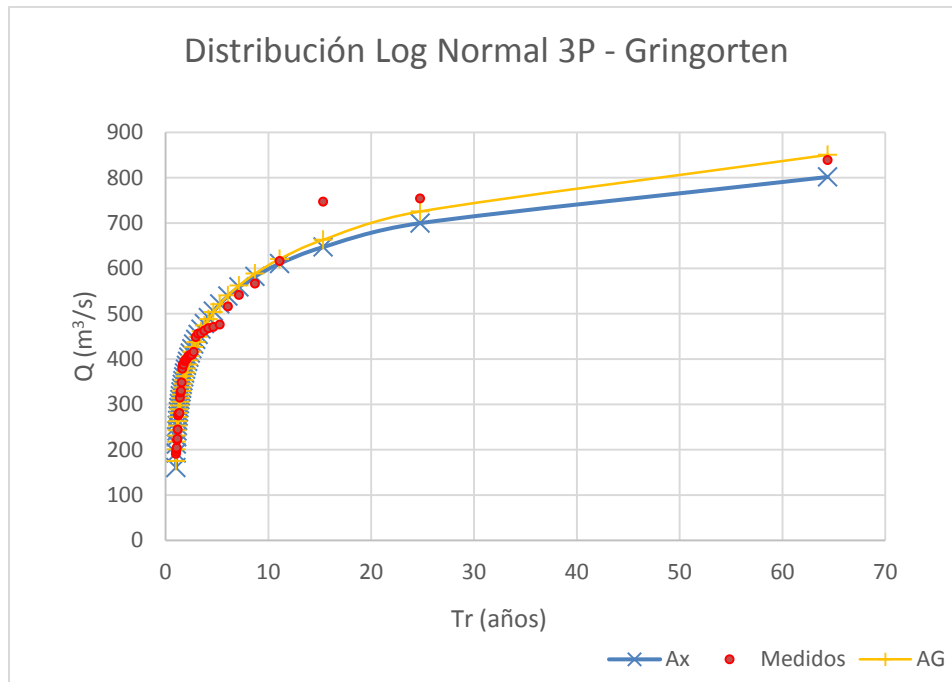


Figura 20 Comparación entre valores históricos y calculados. Función Log Normal 3P – Gringorten

4.1.4 Distribución Gumbel

Intervalo	AX	Algoritmo genético						Rango	
		máx (L)		máx [ln(L)]		mín ($e_1^2 + e_2^2$)		Amplio	Reducido
		Amplio	Reducido	Amplio	Reducido	Amplio	Reducido		
a =	119.0447	121.96	121.96	121.96	121.96	121.96	121.96	0 - 4000	100 - 150
b =	333.897	335.23	335.23	335.23	335.23	335.23	335.23	0 - 4000	300 - 400
F.O. =		4.11E-111	4.11E-111	-254.17	-254.17	-3.97E-20	-3.97E-20		
EE Weibul =	30.026	28.18							
EE Gringorten =	25.535	25.441							

Tabla 9 Resultados obtenidos de la distribución Gumbel para estación “Paso del toro”

Weibull			Gringorten		
Tr (años)	Q (m3/s)		Tr (años)	Q (m3/s)	
	AX	AG		AX	AG
1.025	177.711	175.219	1.014	161.043	158.143
1.051	202.305	200.416	1.040	193.687	191.587
1.079	219.465	217.996	1.068	213.39	211.772
1.108	233.34	232.211	1.097	228.585	227.339
1.139	245.339	244.503	1.128	241.42	240.489
1.171	256.129	255.557	1.161	252.802	252.149
1.206	266.082	265.755	1.195	263.206	262.808
1.242	275.433	275.334	1.232	272.919	272.759
1.281	284.338	284.458	1.271	282.127	282.192
1.323	292.911	293.24	1.313	290.961	291.243
1.367	301.237	301.77	1.357	299.519	300.01
1.414	309.382	310.115	1.405	307.875	308.57
1.464	317.402	318.331	1.456	316.089	316.986
1.519	325.341	326.465	1.511	324.213	325.309
1.577	333.241	334.558	1.570	332.289	333.583
1.640	341.138	342.649	1.634	340.358	341.849
1.708	349.066	350.771	1.703	348.456	350.145
1.783	357.058	358.958	1.778	356.617	358.507
1.864	365.145	367.243	1.861	364.877	366.969
1.952	373.361	375.661	1.951	373.27	375.568
2.050	381.741	384.245	2.051	381.834	384.341
2.158	390.32	393.035	2.162	390.606	393.328
2.278	399.139	402.07	2.285	399.63	402.573
2.412	408.243	411.397	2.423	408.955	412.126
2.563	417.682	421.067	2.578	418.634	422.042
2.733	427.515	431.141	2.755	428.731	432.387
2.929	437.812	441.69	2.959	439.322	443.237
3.154	448.655	452.798	3.194	450.498	454.686
3.417	460.146	464.571	3.471	462.37	466.849
3.727	472.412	477.137	3.799	475.079	479.87
4.100	485.613	490.662	4.197	488.808	493.934
4.556	499.963	505.363	4.687	503.795	509.289
5.125	515.745	521.531	5.307	520.372	526.271
5.857	533.358	539.576	6.116	539.004	545.36
6.833	553.385	560.093	7.216	560.393	567.273
8.200	576.726	584.005	8.798	585.657	593.155
10.250	604.889	612.858	11.270	616.749	625.009
13.667	640.7	649.546	15.672	657.589	666.849
20.500	690.498	700.564	25.718	718.105	728.847
41.000	774.511	786.635	71.643	841.584	855.35

Tabla 10 Valores calculados. Función Gumbel – Weibull y Gringorten

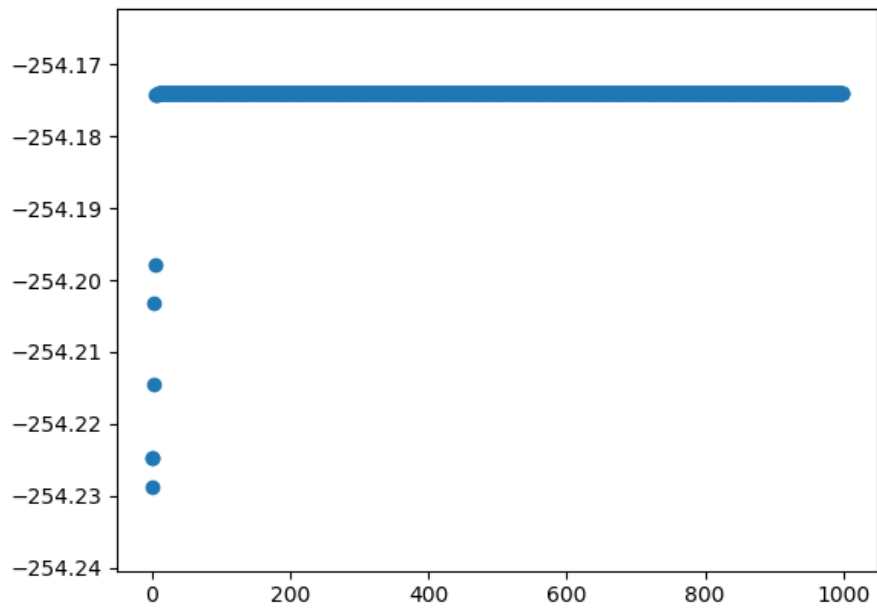


Ilustración 7 Gráfica del $\ln(L)$ en cada generación para la función de distribución Gumbel

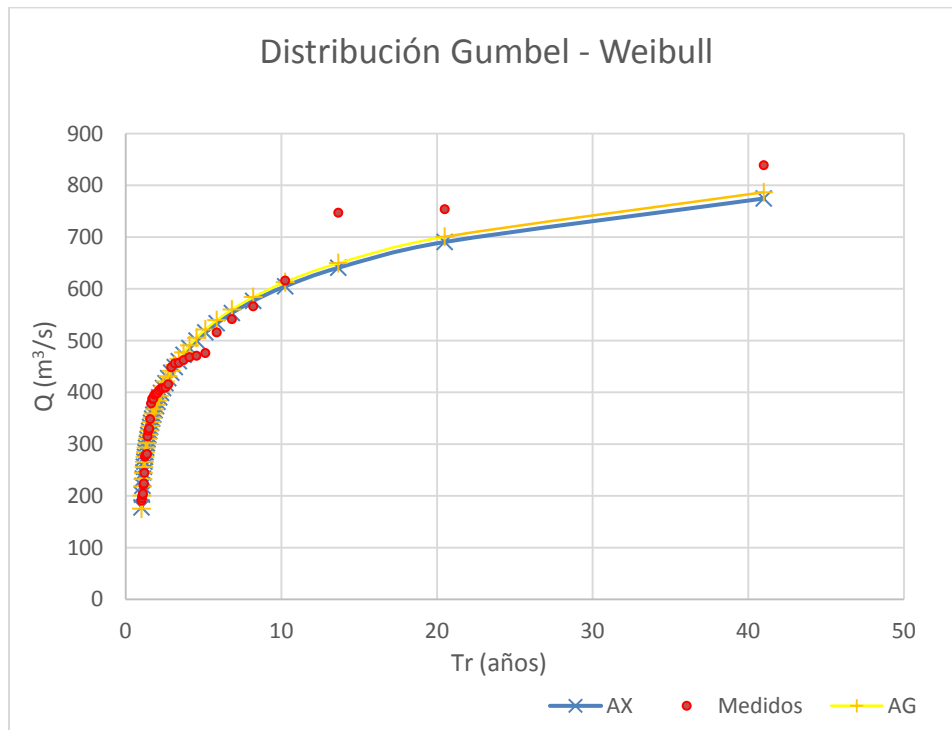


Figura 21 Comparación entre valores históricos y calculados. Función Gumbel – Weibull

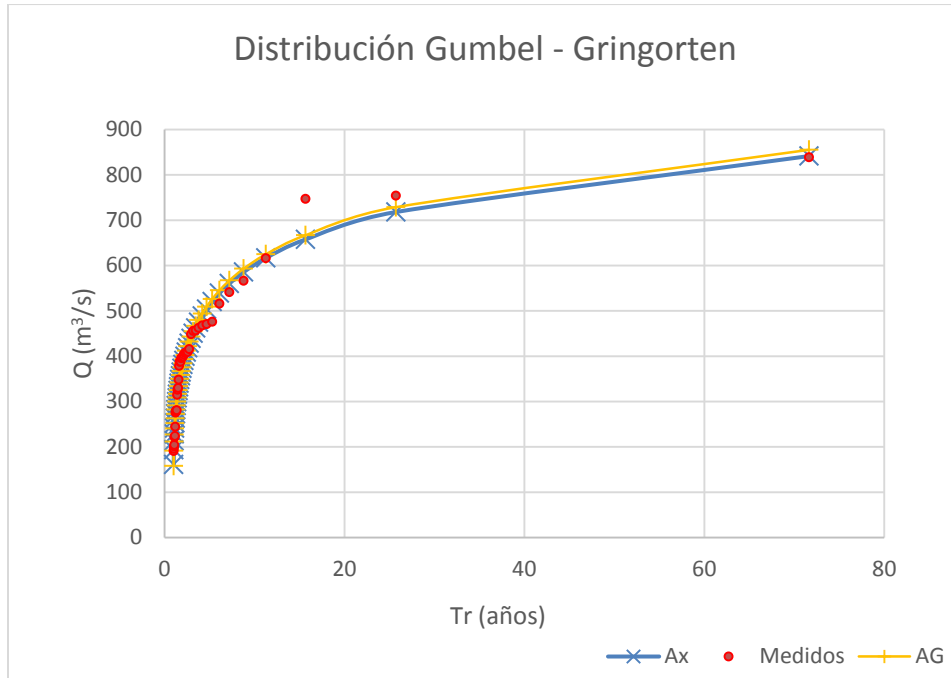


Figura 22 Comparación entre valores históricos y calculados. Función Gumbel – Gringorten

4.1.5 Distribución Doble Gumbel

Intervalo	AX	Algoritmo genético				Rango	
		máx (L)		máx [ln(L)]		Amplio	Reducido
		Amplio	Reducido	Amplio	Reducido		
p =	0.82	0.825	0.825	0.825	0.825	0 - 1	0.7 - 0.9
a1 =	85.89589	90.57971	90.57971	90.57971	90.57971	0 - 4000	50 - 150
b1 =	308.0548	305.802	305.802	305.802	305.802	0 - 4000	250 - 350
a2 =	141.3228	92.9368	92.9368	92.9368	92.9368	0 - 4000	50 - 150
b2 =	574.4003	598.646	598.646	598.646	598.646	0 - 4000	500 - 600
F.O. =		8.82E-33	8.82E-33	-71.06	-71.06		
EE Weibul =	30.099	36.0503					
EE Gringorten =	35.14	40.23					

Tabla 11 Resultados obtenidos de la distribución Doble Gumbel para estación “Paso del toro”

Weibull			Gringorten		
Tr (años)	Q (m3/s)		Tr (años)	Q (m3/s)	
	AX	AG		AX	AG
1.025	200.06	191.8182	1.014	199.140346	186.766664
1.051	218.93	211.7187	1.040	219.108404	211.840584
1.079	232.25	225.6692	1.068	232.345206	227.985366
1.108	243.14	237.1118	1.097	243.727025	239.378473
1.139	252.65	247.1145	1.128	253.143446	249.90094
1.171	261.29	256.1957	1.161	261.682788	258.811678
1.206	269.34	264.6378	1.195	269.72298	267.076602
1.242	276.97	272.6622	1.232	277.528977	274.695066
1.281	284.32	280.4166	1.271	284.824314	282.489136
1.323	291.47	287.894	1.313	292.07105	289.824605
1.367	298.49	295.2653	1.357	299.201902	297.01925
1.414	305.44	302.5628	1.405	306.269509	304.128629
1.464	312.36	309.8392	1.456	313.321722	311.230781
1.519	319.3	317.1416	1.511	320.403458	318.315208
1.577	326.31	324.5146	1.570	327.557791	325.454377
1.640	333.41	332.0022	1.634	334.952239	332.980906
1.708	340.66	339.6493	1.703	342.149789	339.834615
1.783	348.07	347.504	1.778	349.599449	347.502665
1.864	355.71	355.6191	1.861	357.71431	355.153396
1.952	363.62	364.0546	1.951	365.875958	363.51151
2.050	371.85	372.8804	2.051	374.043182	371.594624
2.158	380.45	382.1801	2.162	383.057018	380.456426
2.278	389.49	392.0553	2.285	392.574253	389.817121
2.412	399.04	402.6318	2.423	402.629105	399.716147
2.563	409.2	414.1094	2.578	413.250559	410.098624
2.733	420.06	426.5534	2.755	424.415396	421.307929
2.929	431.74	440.2872	2.959	436.787494	433.012126
3.154	444.39	455.514	3.194	449.991386	446.48011
3.417	458.15	472.3374	3.471	464.307295	460.729881
3.727	473.24	490.6594	3.799	479.866679	476.235129
4.100	489.86	510.1477	4.197	496.804858	493.218812
4.556	508.29	530.3816	4.687	516.513568	513.156598
5.125	528.88	551.1074	5.307	537.619533	534.837945
5.857	552.06	572.5139	6.116	561.148907	559.201389
6.833	578.48	594.9579	7.216	587.609886	586.951148
8.200	609.11	619.2518	8.798	617.603212	620.627447
10.250	645.63	646.6472	11.270	655.303537	660.664146
13.667	691.25	679.4019	15.672	700.334878	711.355349
20.500	753.32	722.3936	25.718	762.618548	786.666097
41.000	855.52	791.9321	71.643	861.171637	933.372723

Tabla 12 Valores calculados. Función Doble Gumbel – Weibull y Gringorten

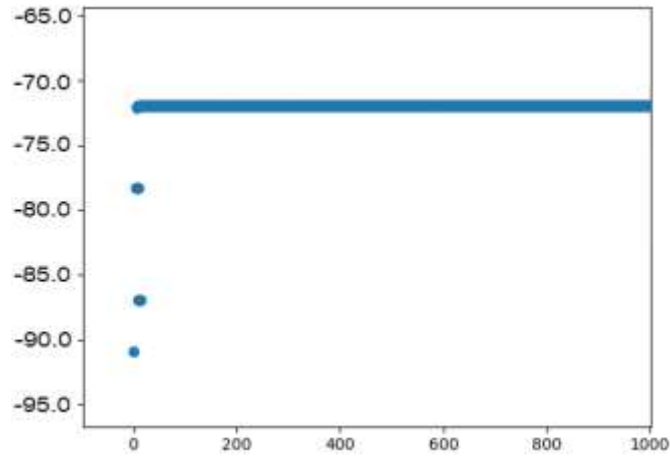


Ilustración 8 Gráfica del $\ln(L)$ en cada generación para la función de distribución Doble Gumbel

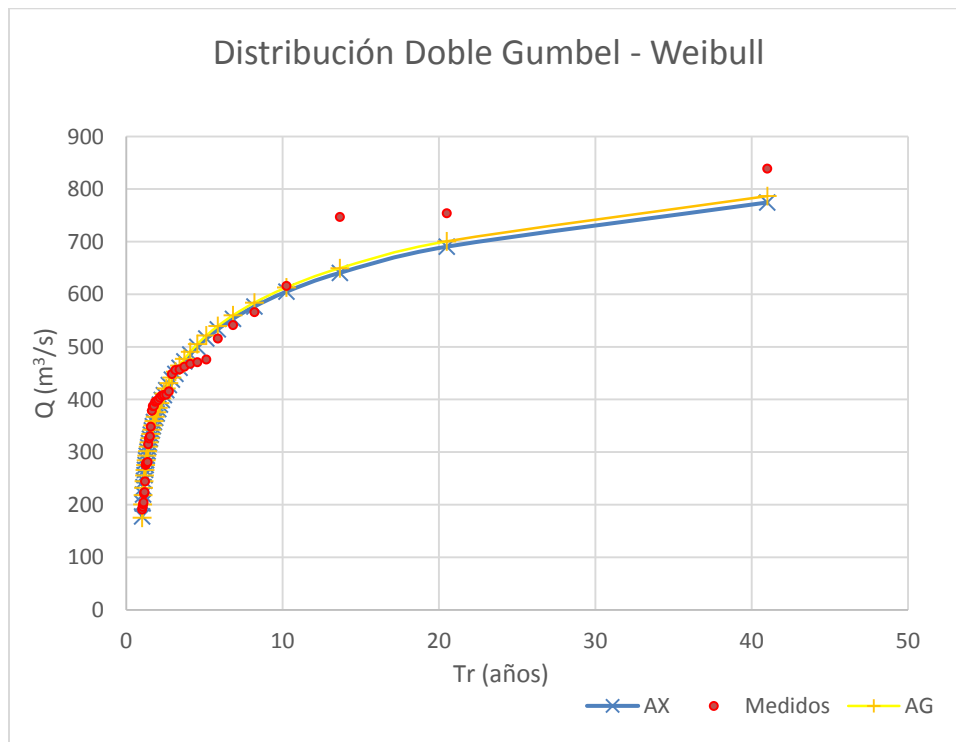


Figura 23 Comparación entre valores históricos y calculados. Función Doble Gumbel – Weibull

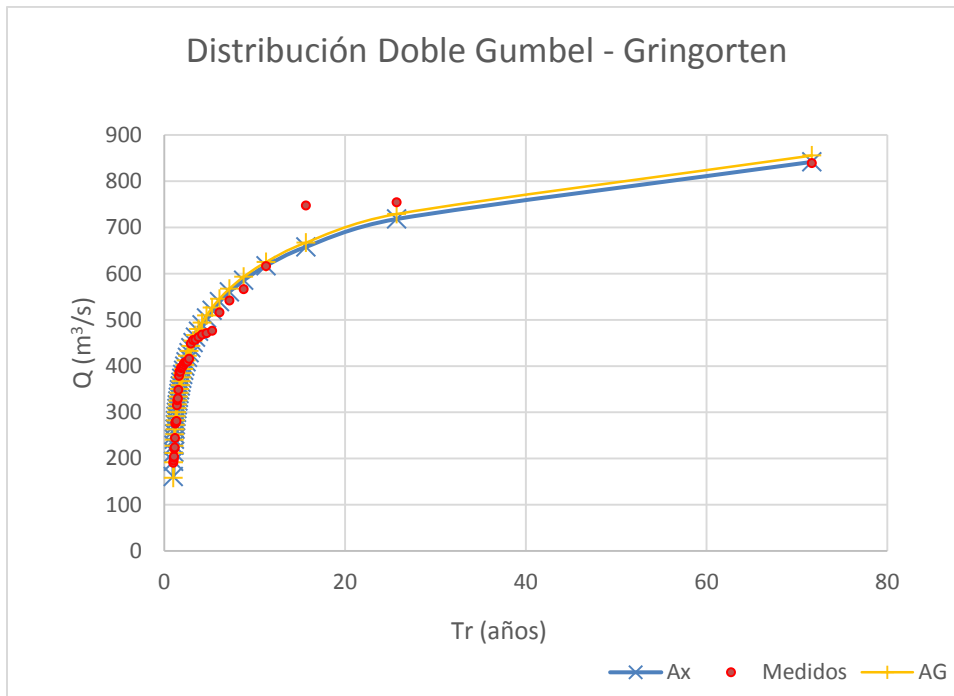


Figura 24 Comparación entre valores históricos y calculados. Función Doble Gumbel – Gringorten

Resumen de errores estándar Archivo analizado: COTA.TXT

Función	Momentos		Máxima Verosimilitud	
	2 parámetros	3 parámetros	2 parámetros	3 parámetros
Normal	43.569	-----	43.569	-----
Lognormal	30.373	30.902	29.868	30.025
Gumbel	29.854	-----	30.026	-----
Exponencial	38.635	-----	210.026	-----
Gamma	32.157	30.740	34.525	28.920
Doble Gumbel	37.749			

Mínimo error estándar: 28.920

Calculado por la función: **Gamma (m x. ver.) 3 p.**

Ilustración 9 Resultados globales del programa AX

4.2 Estación hidrométrica del río Jamapa

En la Tabla 13 se muestran los gastos máximos anuales que corresponden a esta estación.

Año	Q (m ³ /s)	Año	Q (m ³ /s)	Año	Q (m ³ /s)	Año	Q (m ³ /s)
1952	347.14	1963	147.84	1974	540.28	1987	199.95
1953	109.78	1964	126.02	1975	428.68	1992	501.22
1954	264.74	1965	159.92	1976	442.71	1994	83.41
1955	549.29	1966	205.15	1977	142.31	1995	100.85
1956	324.69	1967	430.62	1978	197.4	1997	479.88
1957	89.63	1968	231.96	1979	158.1	2002	103.61
1958	307.09	1969	406.55	1980	345.27	2005	182.37
1959	178.06	1970	192.92	1981	461.88	2010	375.26
1960	347.79	1971	121.39	1982	66.32		
1961	455.12	1972	591.37	1983	382.11		
1962	211.16	1973	214.32	1984	157.47		

Tabla 13 Gastos máximos anuales de la estación "El Tejar"

4.2.1 Distribución Normal

Intervalo	AX	Algoritmo genético						Rango	
		máx (L)		máx [ln(L)]		mín (e ₁ ² + e ₂ ²)		Amplio	Reducido
		Amplio	Reducido	Amplio	Reducido	Amplio	Reducido		
$\mu =$	277.113	277.113	277.113	277.113	277.113	277.113	277.113	0 - 4000	200 - 300
$s =$	151.049	149.190	149.190	149.190	149.190	149.190	149.190	0 - 4000	100 - 200
F.O. =		4.08E-115	4.08E-115	-263.39	-263.39	-2.60E-18	-2.60E-18		
EE Weibul =	36.810	36.998		36.998		36.998			
EE Gringorten =	38.207	38.184		38.184		38.184			

Tabla 14 Resultados obtenidos de la distribución Normal para estación "El Tejar"

Weibull			Gringorten		
Tr (años)	Q (m3/s)		Tr (años)	Q (m3/s)	
	AX	AG		AX	AG
1.024	-21.839	-18.16	1.015	-49.803	-45.78
1.050	25.296	28.395	1.041	11.809	15.074
1.077	55.961	58.682	1.068	46.947	49.780
1.105	79.521	81.953	1.096	72.763	75.278
1.135	99.061	101.252	1.126	93.691	95.948
1.167	116.002	117.985	1.158	111.588	113.625
1.200	131.127	132.924	1.191	127.419	129.262
1.235	144.913	146.540	1.227	141.756	143.422
1.273	157.676	159.146	1.264	154.966	156.469
1.313	169.637	170.960	1.304	167.301	168.652
1.355	180.959	182.142	1.347	178.942	180.150
1.400	191.762	192.812	1.392	190.025	191.097
1.448	202.144	203.066	1.441	200.657	201.597
1.500	212.180	212.979	1.493	210.920	211.734
1.556	221.935	222.614	1.549	220.883	221.575
1.615	231.462	232.024	1.610	230.604	231.176
1.680	240.807	241.254	1.675	240.133	240.588
1.750	250.011	250.345	1.746	249.512	249.852
1.826	259.111	259.332	1.823	258.782	259.007
1.909	268.141	268.251	1.908	267.977	268.089
2.000	277.132	277.132	2.000	277.132	277.132
2.100	286.085	285.975	2.102	286.249	286.137
2.211	295.115	294.893	2.215	295.444	295.219
2.333	304.215	303.881	2.340	304.714	304.374
2.471	313.419	312.972	2.481	314.093	313.638
2.625	322.764	322.202	2.640	323.622	323.049
2.800	332.291	331.612	2.821	333.343	332.651
3.000	342.046	341.247	3.028	343.307	342.492
3.231	352.082	351.160	3.267	353.569	352.628
3.500	362.464	361.413	3.548	364.201	363.129
3.818	373.267	372.084	3.882	375.284	374.076
4.200	384.589	383.266	4.286	386.925	385.574
4.667	396.550	395.080	4.783	399.260	397.756
5.250	409.313	407.686	5.410	412.470	410.804
6.000	423.099	421.302	6.226	426.807	424.964
7.000	438.224	436.241	7.333	442.638	440.601
8.400	455.165	452.974	8.919	460.535	458.278
10.500	474.705	472.273	11.379	481.463	478.948
14.000	498.265	495.543	15.714	507.279	504.446
21.000	528.929	525.830	25.385	542.417	539.152
42.000	576.065	572.386	66.000	604.030	600.006

Tabla 15 Valores calculados. Función Normal – Weibull y Gringorten

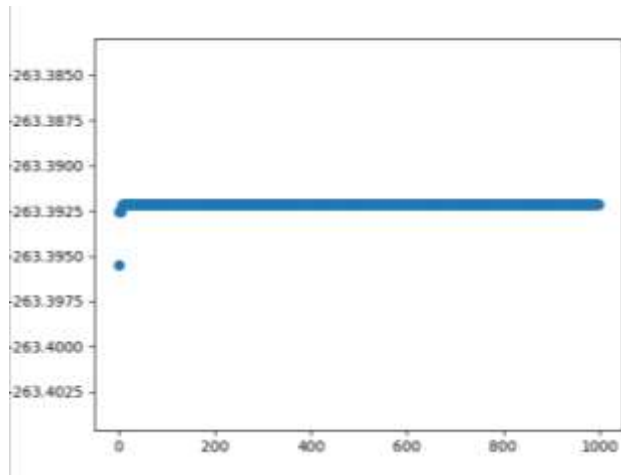


Ilustración 10 Gráfica del $\ln(L)$ en cada generación para la función de distribución Normal

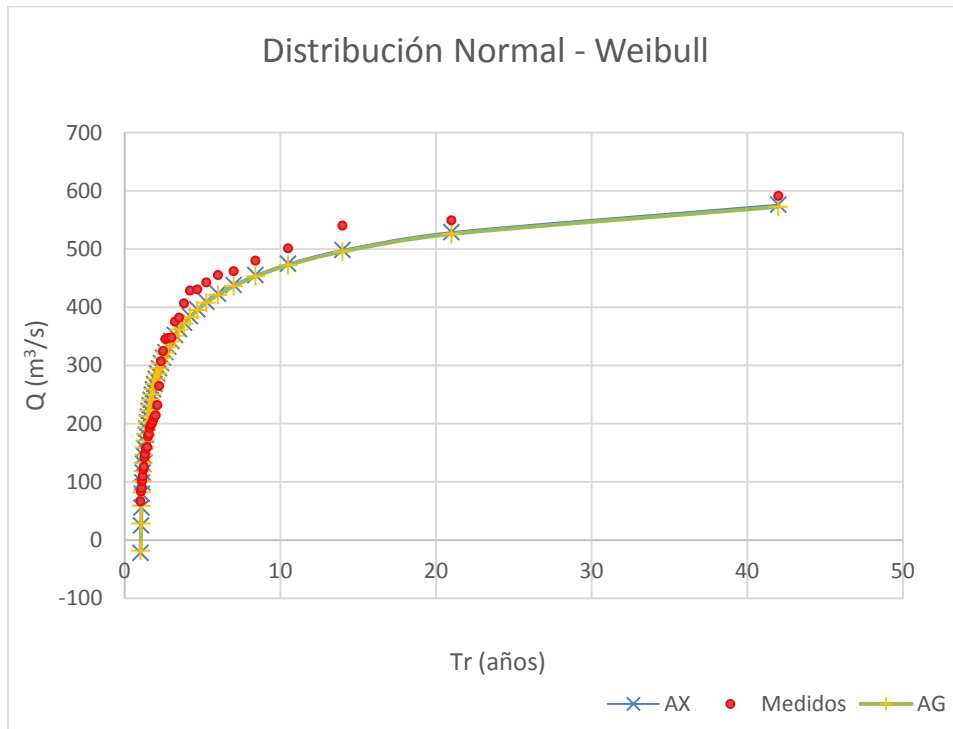


Figura 25 Comparación entre valores históricos y calculados. Función Normal – Weibull

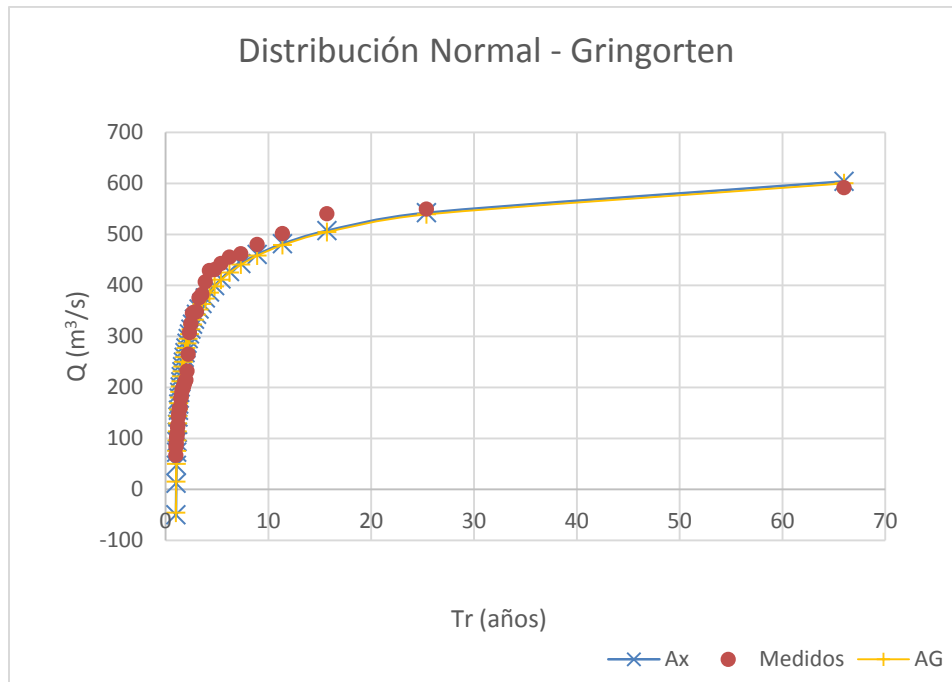


Figura 26 Comparación entre valores históricos y calculados. Función Normal – Gringorten

4.2.2 Distribución Log Normal 2 Parámetros

Intervalo	AX	Algoritmo genético						Rango	
		máx (L)		máx [ln(L)]		mín ($e_1^2 + e_2^2$)		Amplio	Reducido
		Amplio	Reducido	Amplio	Reducido	Amplio	Reducido		
$\mu_y =$	5.462	5.462	5.462	5.462	5.462	5.462	5.462	0 - 4000	0 - 10
$s_y =$	0.5921	0.5921	0.5921	0.5921	0.5921	0.5921	0.5921	0 - 4000	0 - 1
F.O. =		6.47E-114	6.47E-114	-260.67	-260.67	-2.04E-14	-2.04E-14		
EE Weibul =	44.404	44.404		44.404		44.404			
EE Gringorten =	54.801	54.801		54.801		54.801			

Tabla 16 Resultados obtenidos de la distribución Log Normal 2P para estación “El Tejar”

Weibull			Gringorten		
Tr (años)	Q (m3/s)		Tr (años)	Q (m3/s)	
	AX	AG		AX	AG
1.024	72.976		1.015	65.399	
1.050	87.786		1.041	83.265	
1.077	98.998		1.068	95.561	
1.105	108.576		1.096	105.738	
1.135	117.22		1.126	114.778	
1.167	125.268		1.158	123.119	
1.200	132.92		1.191	131.002	
1.235	140.3		1.227	138.575	
1.273	147.498		1.264	145.94	
1.313	154.579		1.304	153.169	
1.355	161.593		1.347	160.321	
1.400	168.583		1.392	167.44	
1.448	175.585		1.441	174.565	
1.500	182.631		1.493	181.731	
1.556	189.75		1.549	188.969	
1.615	196.97		1.610	196.309	
1.680	204.319		1.675	203.78	
1.750	211.826		1.746	211.411	
1.826	219.518		1.823	219.235	
1.909	227.427		1.908	227.281	
2.000	235.586		2.000	235.586	
2.100	244.001		2.102	244.157	
2.211	252.792		2.215	253.119	
2.333	261.972		2.340	262.485	
2.471	271.596		2.481	272.315	
2.625	281.73		2.640	282.679	
2.800	292.45		2.821	293.659	
3.000	303.849		3.028	305.354	
3.231	316.042		3.267	317.889	
3.500	329.168		3.548	331.417	
3.818	343.407		3.882	346.133	
4.200	358.99		4.286	362.293	
4.667	376.224		4.783	380.241	
5.250	395.525		5.410	400.449	
6.000	417.487		6.226	423.599	
7.000	442.987		7.333	450.72	
8.400	473.405		8.919	483.476	
10.500	511.09		11.379	524.81	
14.000	560.539		15.714	580.698	
21.000	632.133		25.385	666.454	
42.000	760.418		66.000	848.514	

Tabla 17 Valores calculados. Función Log Normal 2P – Weibull y Gringorten

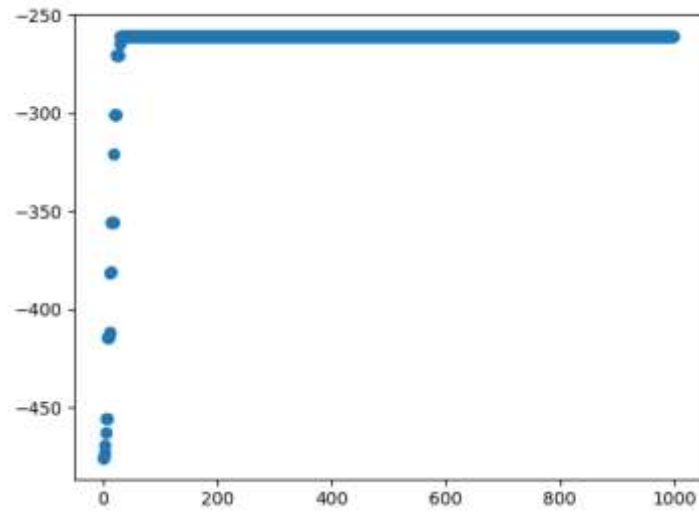


Ilustración 11 Gráfica del $\ln(L)$ en cada generación para la función de distribución Log Normal 2P

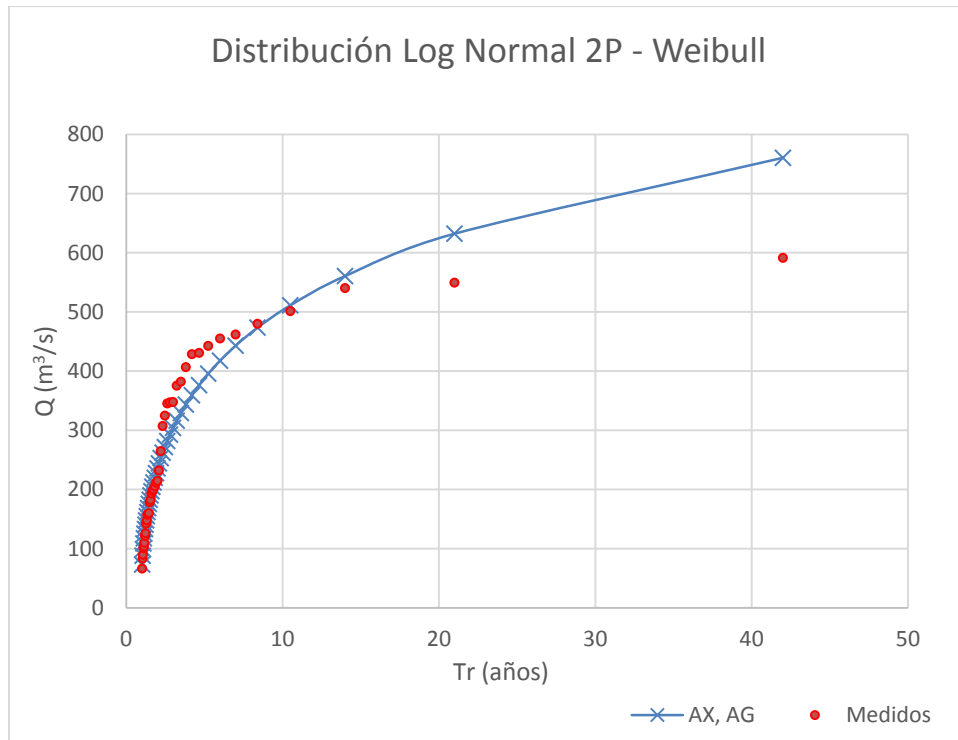


Figura 27 Comparación entre valores históricos y calculados. Función Log Normal 2P – Weibull

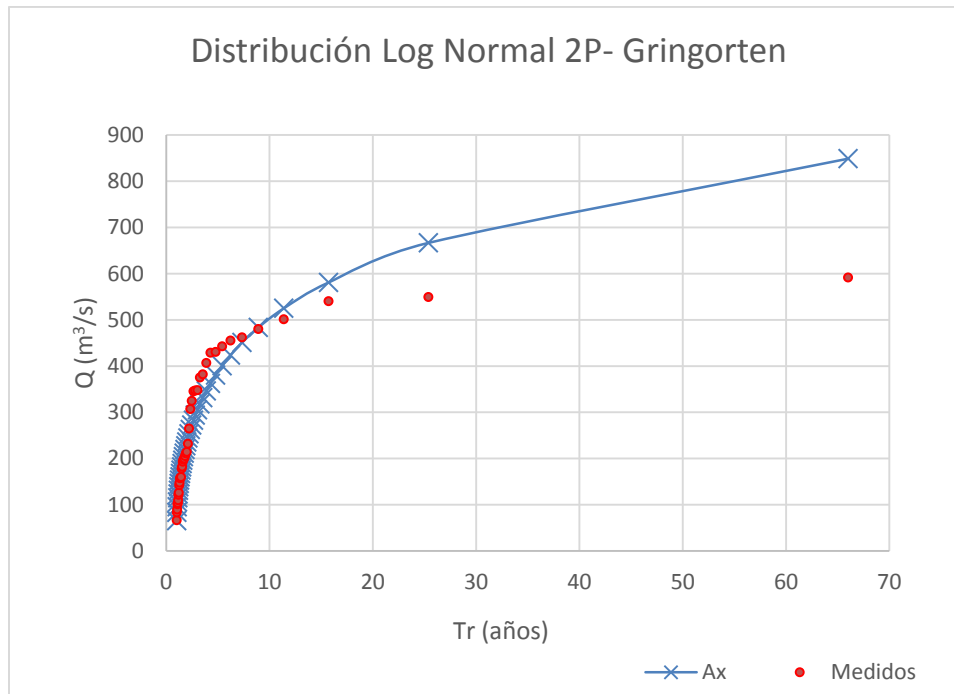


Figura 28 Comparación entre valores históricos y calculados. Función Log Normal 2P – Gringorten

4.2.3 Distribución Log Normal 3 Parámetros

Intervalo	AX	Algoritmo genético						Rango	
		máx (L)		máx [ln(L)]		mín ($e_1^2 + e_2^2 + e_3^2$)		Amplio	Reducido
		Amplio	Reducido	Amplio	Reducido	Amplio	Reducido		
$\mu_y =$	5.4699	5.557	5.498	5.557	5.498	5.531	5.457	0 - 1000	0 - 10
$s_y =$	0.5891	0.5384	0.5709	0.5384	0.5713	0.5527	0.5952	0 - 1000	0 - 1
$x_o =$	-1.561	-20	-7.304	-20	-7.1874	-14.21	1.044	-100 - 60	-10 - 10
F.O. =		6.51E-114	4.86E-114	-260.56	-260.62	-9.65E-08	-4.18E-07		
EE Weibul =	44.235	40.409	42.703	40.4409	42.734	41.378	44.625		
EE Gringorten =	55.364	49.394	52.958	49.394	53.052	50.944	55.917		

Tabla 18 Resultados obtenidos de la distribución Log normal 3P para estación “El Tejar”

Weibull			Gringorten		
Tr (años)	Q (m ³ /s)		Tr (años)	Q (m ³ /s)	
	AX	AG		AX	AG
1.024	72.432	69.247	1.015	64.787	60.781
1.050	87.365	85.575	1.041	82.808	80.62
1.077	98.662	97.769	1.068	95.2	94.045
1.105	108.307	108.086	1.096	105.449	105.038
1.135	117.007	117.325	1.126	114.549	114.721
1.167	125.106	125.873	1.158	122.943	123.596
1.200	132.802	133.953	1.191	130.873	131.932
1.235	140.224	141.707	1.227	138.489	139.898
1.273	147.46	149.233	1.264	145.894	147.607
1.313	154.577	156.605	1.304	153.16	155.14
1.355	161.625	163.877	1.347	160.347	162.56
1.400	168.648	171.096	1.392	167.499	169.916
1.448	175.681	178.3	1.441	174.656	177.251
1.500	182.756	185.522	1.493	181.852	184.601
1.556	189.904	192.794	1.549	189.119	191.997
1.615	197.151	200.144	1.610	196.488	199.472
1.680	204.527	207.6	1.675	203.986	207.054
1.750	212.059	215.191	1.746	211.644	214.773
1.826	219.777	222.945	1.823	219.493	222.659
1.909	227.711	230.891	1.908	227.564	230.745
2.000	235.893	239.062	2.000	235.893	239.062
2.100	244.331	247.463	2.102	244.488	247.619
2.211	253.145	256.211	2.215	253.472	256.536
2.333	262.347	265.317	2.340	262.861	265.825
2.471	271.992	274.833	2.481	272.712	275.543
2.625	282.146	284.819	2.640	283.097	285.753
2.800	292.886	295.348	2.821	294.096	296.533
3.000	304.304	306.506	3.028	305.811	307.976
3.231	316.513	318.398	3.267	318.363	320.196
3.500	329.656	331.154	3.548	331.907	333.335
3.818	343.91	344.94	3.882	346.638	347.573
4.200	359.505	359.968	4.286	362.81	363.146
4.667	376.748	376.518	4.783	380.767	380.367
5.250	396.056	394.974	5.410	400.981	399.669
6.000	418.019	415.874	6.226	424.13	421.673
7.000	443.514	440.017	7.333	451.243	447.313
8.400	473.914	468.652	8.919	483.977	478.095
10.500	511.565	503.899	11.379	525.269	516.672
14.000	560.948	549.795	15.714	581.073	568.398
21.000	632.407	615.604	25.385	666.648	646.908
42.000	760.351	731.887	66.000	848.148	810.696

Tabla 19 Valores calculados. Función Log Normal 3P – Weibull y Gringorten

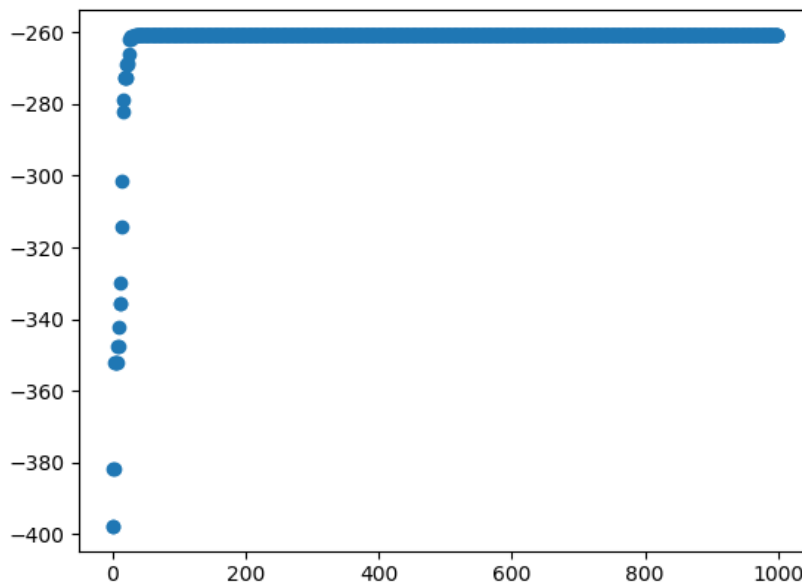


Ilustración 12 Gráfica del Ln (L) en cada generación para la función de distribución Log Normal 3P

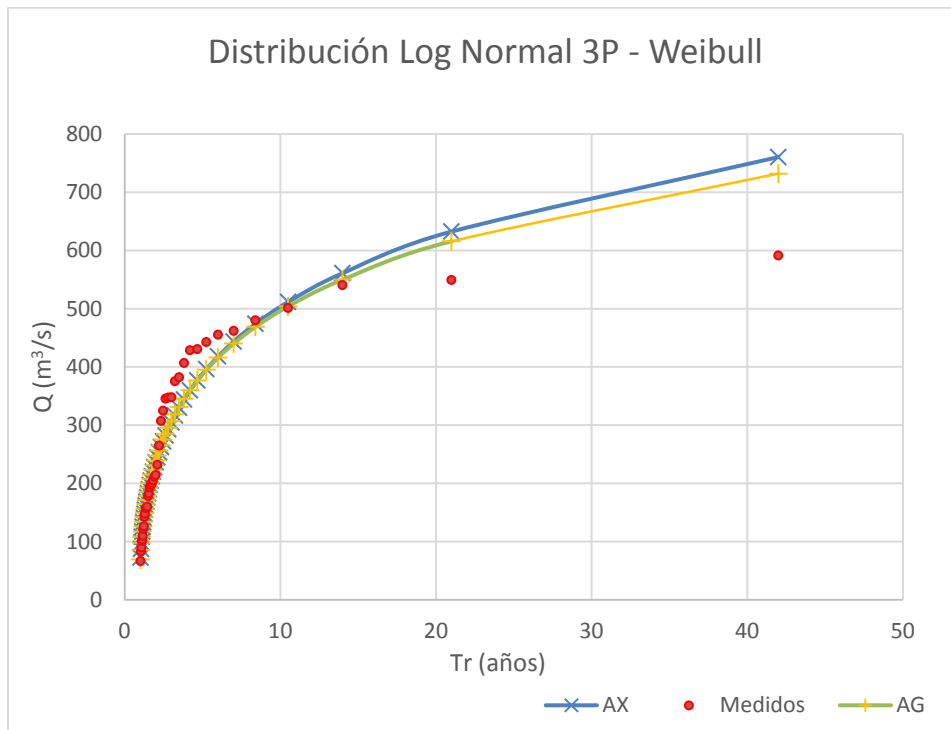


Figura 29 Comparación entre valores históricos y calculados. Función Log Normal 3P – Weibull

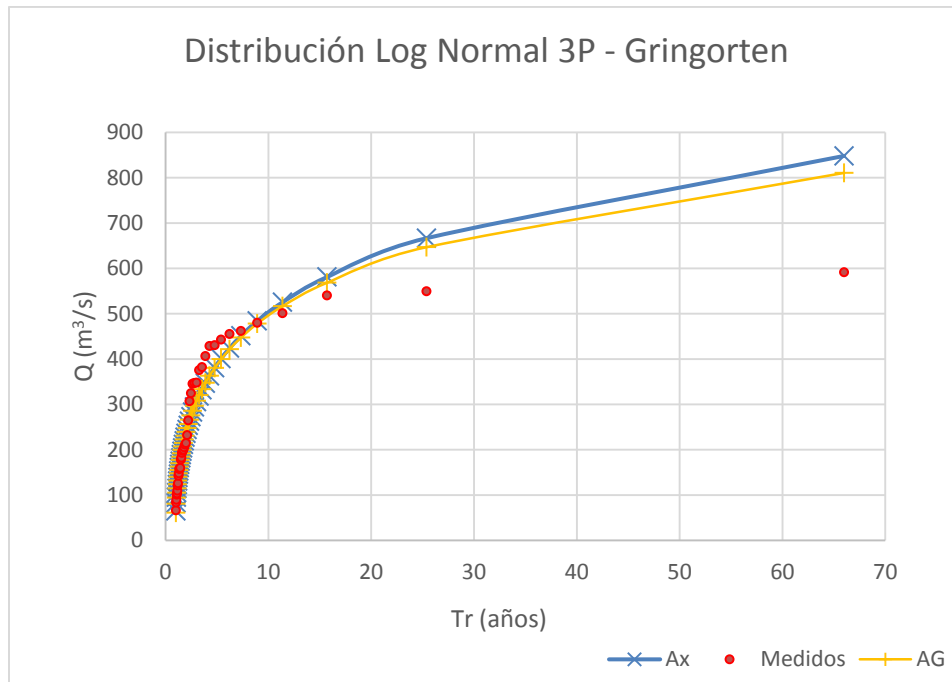


Figura 30 Comparación entre valores históricos y calculados. Función Log Normal 3P – Gringorten

4.2.4 Distribución Gumbel

Intervalo	AX	Algoritmo genético						Rango	
		máx (L)		máx [ln(L)]		mín ($e_1^2 + e_2^2$)		Amplio	Reducido
		Amplio	Reducido	Amplio	Reducido	Amplio	Reducido		
a =	119.0476	125.36	125.36	125.36	125.36	125.36	125.36	0 - 4000	100 - 150
b =	205.9725	208.145	208.145	208.145	208.145	208.145	208.145	0 - 4000	200 - 250
F.O. =		2.98E-114	2.98E-114	261.40	261.40	-1.39E-20	-1.39E-20		
EE Weibul =	33.689	31.768							
EE Gringorten =	37.298	38.335							

Tabla 20 Resultados obtenidos de la distribución Gumbel para estación “El tejár”

Weibull			Gringorten		
Tr (años)	Q (m3/s)		Tr (años)	Q (m3/s)	
	AX	AG		AX	AG
1.024	49.013	42.863	1.014	32.43	25.4
1.050	73.432	68.576	1.039	64.86	59.55
1.077	90.446	86.493	1.066	84.403	80.129
1.105	104.187	100.962	1.095	99.454	95.979
1.135	116.057	113.462	1.125	112.153	109.351
1.167	126.719	124.689	1.156	123.402	121.196
1.200	136.544	135.035	1.190	133.673	132.011
1.235	145.764	144.744	1.225	143.25	142.097
1.273	154.535	153.981	1.263	152.319	151.647
1.313	162.97	162.862	1.303	161.01	160.798
1.355	171.151	171.477	1.346	169.418	169.652
1.400	179.145	179.895	1.391	177.617	178.286
1.448	187.005	188.172	1.440	185.667	186.763
1.500	194.776	196.355	1.492	193.616	195.134
1.556	202.498	204.486	1.548	201.508	203.443
1.615	210.204	212.601	1.609	209.379	211.732
1.680	217.927	220.734	1.674	217.265	220.036
1.750	225.699	228.918	1.745	225.198	228.39
1.826	233.549	237.184	1.823	233.211	236.828
1.909	241.507	245.564	1.907	241.335	245.383
2.000	249.605	254.091	2.000	249.605	254.091
2.100	257.876	262.8	2.102	258.055	262.989
2.211	266.354	271.728	2.216	266.722	272.116
2.333	275.08	280.917	2.342	275.65	281.517
2.471	284.097	290.412	2.483	284.884	291.241
2.625	293.456	300.267	2.643	294.48	301.345
2.800	303.213	310.542	2.824	304.499	311.895
3.000	313.439	321.31	3.032	315.016	322.971
3.231	324.216	332.658	3.274	326.123	334.666
3.500	335.644	344.692	3.557	337.93	347.099
3.818	347.85	357.546	3.894	350.578	360.418
4.200	360.996	371.388	4.301	364.249	374.814
4.667	375.292	386.443	4.804	379.182	390.539
5.250	391.024	403.009	5.439	395.706	407.939
6.000	408.59	421.506	6.268	414.29	427.508
7.000	428.571	442.546	7.396	435.632	449.982
8.400	451.868	467.079	9.018	460.851	476.538
10.500	479.99	496.693	11.551	491.901	509.234
14.000	515.762	534.361	16.063	532.701	552.198
21.000	565.523	586.761	26.359	593.179	615.883
42.000	649.501	675.192	73.429	716.623	745.872

Tabla 21 Valores calculados. Función Gumbel – Weibull y Gringorten

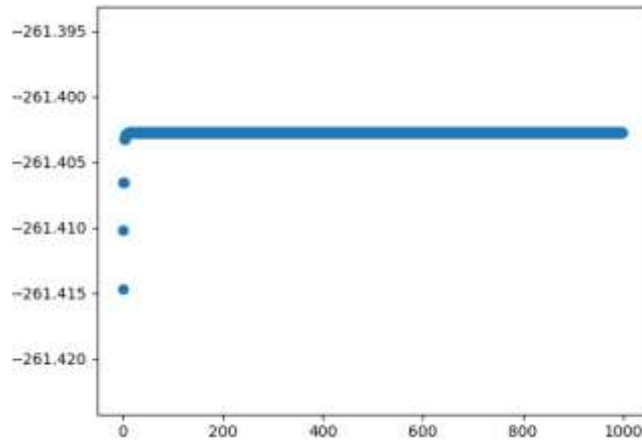


Ilustración 13 Gráfica del $\ln(L)$ en cada generación para la función de distribución Gumbel

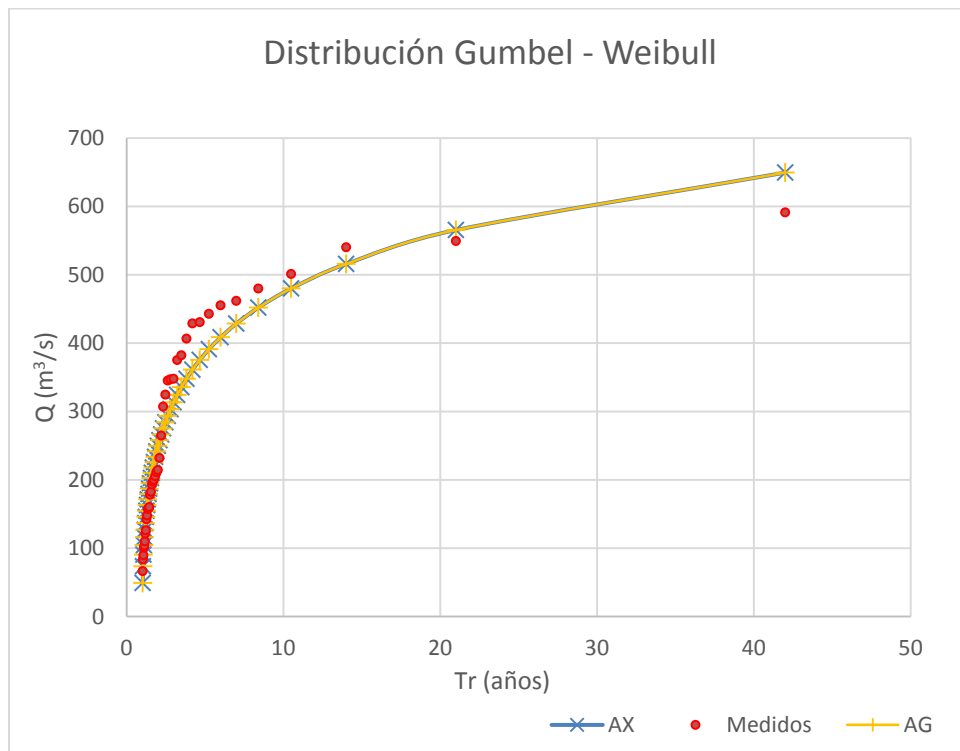


Figura 31 Comparación entre valores históricos y calculados. Función Gumbel – Weibull

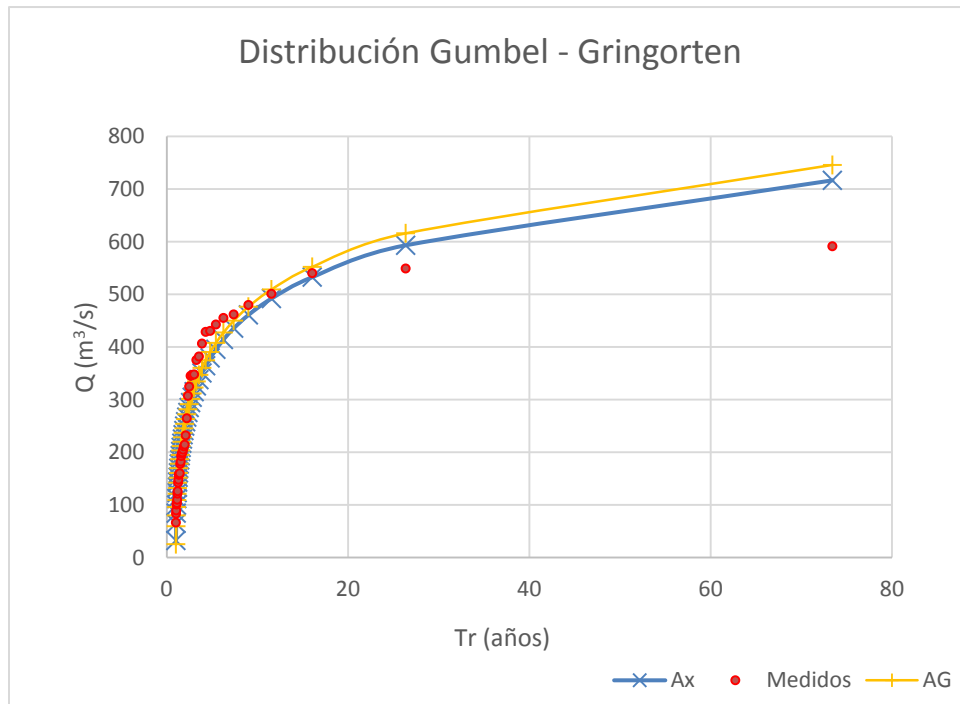


Figura 32 Comparación entre valores históricos y calculados. Función Gumbel – Gringorten

4.2.4 Distribución Doble Gumbel

Intervalo	AX	Algoritmo genético				Rango	
		máx (L)		máx [ln(L)]		Amplio	Reducido
		Amplio	Reducido	Amplio	Reducido		
p =	0.8	0.82	0.82	0.82	0.82	0 - 1	0.7 - 0.9
a1 =	98.73618	94.33962	94.33962	94.33962	94.33962	0 - 4000	50 - 150
b1 =	169.47	168.53	168.53	168.53	168.53	0 - 4000	100 - 200
a2 =	50.82076	0.02012	0.02012	0.02012	0.02012	0 - 4000	0 - 100
b2 =	469.43	467.237	467.237	467.237	467.237	0 - 4000	400 - 500
F.O. =		7.88E-33	7.88E-33	-73.92	-73.92		
EE Weibul =	21.859	23.549					
EE Gringorten =	24.36	29.68					

Tabla 22 Resultados obtenidos de la distribución Doble Gumbel para estación “El Tejar”

Weibull			Gringorten		
Tr (años)	Q (m ³ /s)		Tr (años)	Q (m ³ /s)	
	AX	AG		AX	AG
1.024	45.38	49.29	1.014	60.38	64.29
1.050	67.07	69.86	1.039	82.07	84.86
1.077	82.39	84.36	1.066	97.39	99.36
1.105	94.9	96.19	1.095	109.9	111.19
1.135	105.84	106.51	1.125	120.84	121.51
1.167	115.77	115.87	1.156	130.77	130.87
1.200	125.03	124.59	1.190	140.03	139.59
1.235	133.81	132.84	1.225	148.81	147.84
1.273	142.27	140.78	1.263	157.27	155.78
1.313	150.5	148.49	1.303	165.5	163.49
1.355	158.58	156.06	1.346	173.58	171.06
1.400	166.59	163.54	1.391	181.59	178.54
1.448	174.59	170.99	1.440	189.59	185.99
1.500	182.61	178.45	1.492	197.61	193.45
1.556	190.71	185.98	1.548	205.71	200.98
1.615	198.95	193.6	1.609	213.95	208.6
1.680	207.36	201.37	1.674	222.36	216.37
1.750	216.02	209.34	1.745	231.02	224.34
1.826	224.96	217.55	1.823	249.96	232.55
1.909	234.27	226.05	1.907	259.27	241.05
2.000	244.02	234.93	2.000	269.02	259.93
2.100	254.31	244.24	2.102	279.31	269.24
2.211	265.26	254.1	2.216	290.26	279.1
2.333	277.02	264.61	2.342	302.02	289.61
2.471	289.79	275.92	2.483	314.79	300.92
2.625	303.84	288.24	2.643	328.84	313.24
2.800	319.56	301.84	2.824	344.56	326.84
3.000	337.52	317.11	3.032	362.52	342.11
3.231	358.6	334.63	3.274	383.6	359.63
3.500	383.41	355.34	3.557	408.41	380.34
3.818	407.48	380.25	3.894	432.48	405.25
4.200	426.15	405.65	4.301	451.15	430.65
4.667	441.29	425.52	4.804	466.29	450.52
5.250	454.85	441.56	5.439	474.85	466.56
6.000	467.9	455.99	6.268	487.9	480.99
7.000	481.2	470.08	7.396	501.2	495.08
8.400	495.48	484.79	9.018	515.48	509.79
10.500	511.67	501.12	11.551	531.67	526.12
14.000	531.35	520.67	16.063	551.35	545.67
21.000	557.98	546.81	26.359	577.98	571.81
42.000	602.93	590.39	73.429	622.93	615.39

Tabla 23 Valores calculados. Función Doble Gumbel – Weibull y Gringorten

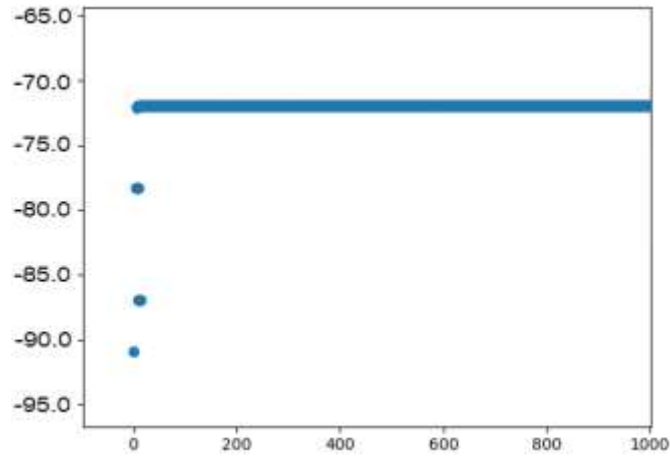


Ilustración 14 Gráfica del $\ln(L)$ en cada generación para la función de distribución Doble Gumbel

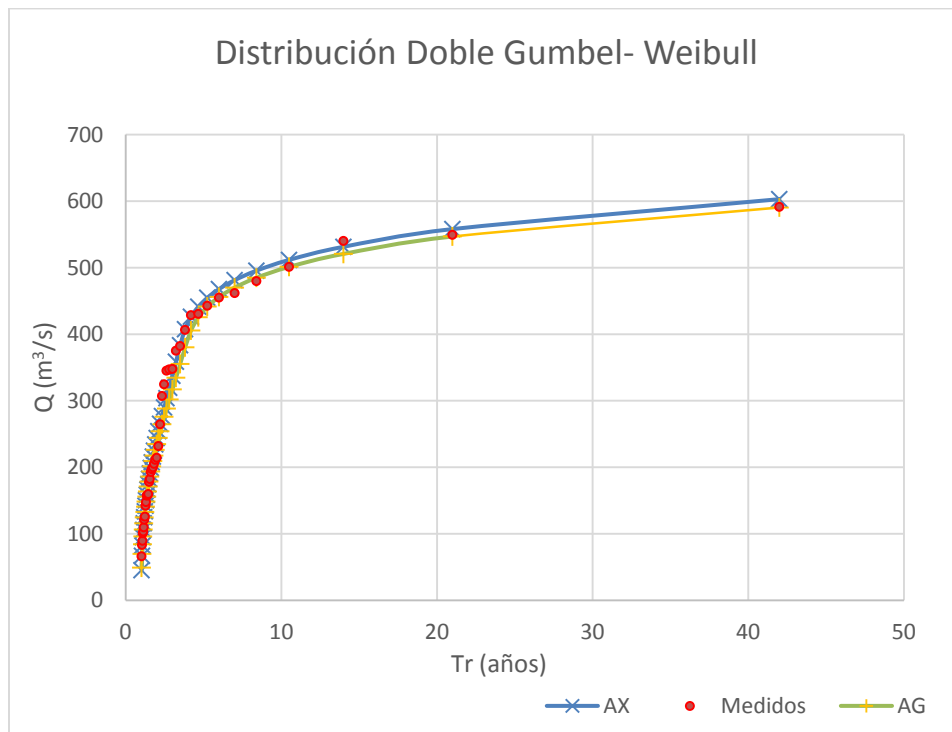


Figura 33 Comparación entre valores históricos y calculados. Función Doble Gumbel – Weibull

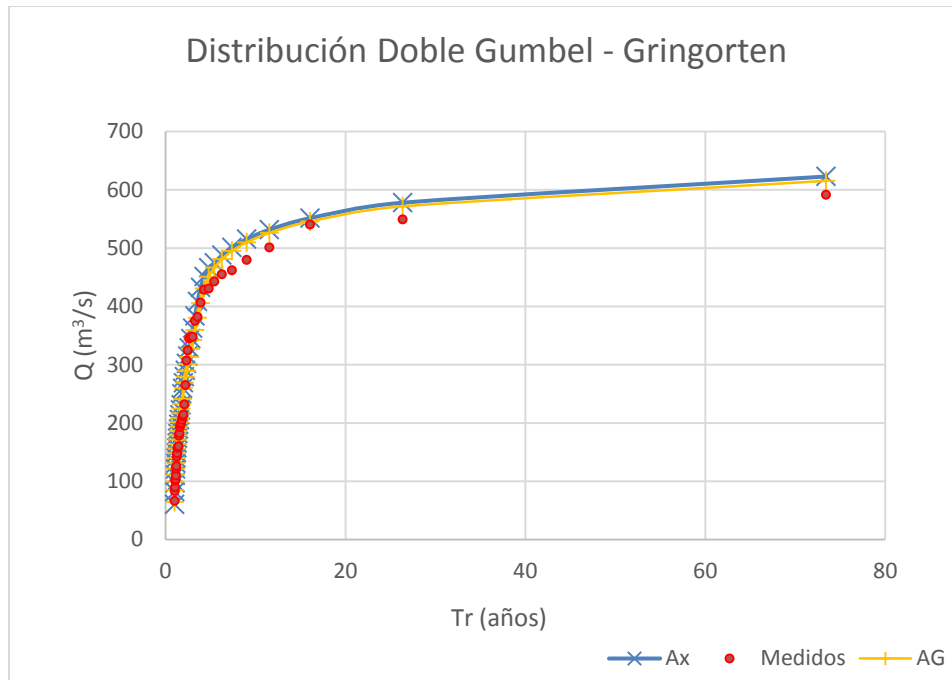


Figura 34 Comparación entre valores históricos y calculados. Función Doble Gumbel – Gringorten

Resumen de errores estándar Archivo analizado: JAMAPA-1....

Función	Momentos		Máxima Verosimilitud	
	2 parámetros	3 parámetros	2 parámetros	3 parámetros
Normal	36.810	-----	36.810	-----
Lognormal	41.554	31.686	44.404	44.235
Gumbel	33.927	-----	33.689	-----
Exponencial	47.807	-----	110.362	-----
Gamma	31.920	31.271	31.304	37.333
Doble Gumbel	21.859			

Mínimo error estándar: 21.859

Calculado por la función: Doble Gumbel

Ilustración 15 Resultados globales del programa AX

4.3 Análisis de resultados

Con los resultados obtenidos se puede observar que el algoritmo genético no tuvo éxito en todas las funciones de distribución. En la función de distribución doble Gumbel, a pesar de que se encontraron valores diferentes para cada uno de los parámetros, estos no reducían el error estándar como se muestra en las tablas 11 y 22.

En la función log normal de 2 parámetros se encontraron los mismos valores que calcula el programa AX. A pesar de que en la función de distribución normal se puede observar un caso similar, es importante mencionar que el parámetro de la desviación estándar al que llega el algoritmo es un parámetro de la población, en contra parte, el programa AX obtiene este parámetro únicamente para la muestra como se aprecia en las tablas 3 y 14. La razón por la cual se llega a los mismos valores que el programa AX es porque estas dos distribuciones en particular tienen funciones muy simples de evaluar como resultado de las derivadas parciales de la función de máxima verosimilitud. El espacio de búsqueda no es tan complejo y se puede converger a un resultado casi de manera inmediata.

El algoritmo genético cumplió con su objetivo en las funciones de distribución log normal de 3 parámetros y Gumbel, pues se encontraron valores en sus parámetros distintos a los del programa AX que reducen el error estándar. Se esperaba que el algoritmo encontrara valores diferentes, pues estas funciones muy complejas que incluso pueden llegar a estar una en función de otra. En las tablas 7, 9, 18 y 20 se muestra el valor de los errores estándar. Además, se puede verificar de manera gráfica como el algoritmo genético brinda un mejor ajuste a las funciones de distribución con los datos observados como es evidente en la figura 19.

El objetivo principal de utilizar intervalos amplios y reducidos es con la intención de encontrar los máximos globales y verificar los resultados que arroja el programa AX. En cada una de las tablas se especifica el valor de la función objetivo utilizada y como se esperaba en los casos que se lograba maximizar la función de verosimilitud se obtenían parámetros que ajustaban mejor a los datos.

Para cada función de distribución se adjuntó una gráfica que representa el comportamiento del algoritmo genético, son muy similares entre ellas. Es importante mencionar que el comportamiento de estas gráficas explica que el algoritmo genético no requiere de tantas generaciones, pues converge a los resultados relativamente rápido, esto a su vez porque se cuenta con poblaciones muy grandes.

En el cálculo del error estándar de las funciones de distribución se empleó el uso la expresión de Gringorten para agregar una variante que pudiera dar otra perspectiva de un ajuste. En general el periodo de retorno es mayor comparado con el autor Weibull. Se presentan excepciones como la distribución normal para el conjunto de datos "El Tejar". Tal y como se

presenta en la tabla 15 el valor calculado para el menor de los gastos resulta ser un número negativo, valor que para aplicaciones en hidrología nos resulta inútil. Aunque en otros casos el error estándar que se produce al utilizar la teoría de Gringorten es menor que con la de Weibull. Estas diferencias son debidas a los tipos de datos con los que se estén trabajando. La ausencia de datos para varios años de cada estación es un claro ejemplo de que el comportamiento de los datos es distinto.

CAPÍTULO 5. CONCLUSIONES

La aplicación de los algoritmos genéticos es muy extensa, en este trabajo se abordaron casos específicos de estos mismos. Y resulta satisfactorio que hayan cumplido con el objetivo principal. Se demuestra entonces que un método no determinístico tiene una ventaja sobre los determinísticos al tener la capacidad de explorar espacios de búsqueda a los que los métodos convencionales no pueden acceder.

Se redujo el error estándar de dos funciones de distribución que permiten darle a un especialista más elementos para formar un criterio de la utilización de esta información. Reducir esta diferencia con los datos observados puede beneficiar al ingeniero civil si se piensa por ejemplo que una presa se proyecta con un periodo de retorno de 10 000 años, al reducir el error se precisa hablando estrictamente en términos cuantitativos en los valores extrapolados para dicho periodo. Dando la posibilidad de que se puedan reducir costos.

Como trabajo de investigación los resultados son significativos pues se sabe ahora que para algunas funciones de distribución existen máximos globales, que son expresiones tan complicadas de visualizar incluso de manera gráfica debido a que se manejan números muy pequeños y algunos softwares especializados en el análisis de frecuencias los pueden considerar como un cero absoluto, hablando de los valores de la función de máxima verosimilitud.

Los algoritmos genéticos son una herramienta cuya aplicación no se limita solo al campo de la hidrología, si se cuenta con el programa realizado basta con hacerle ligeras modificaciones. Sería necesario contar con una función objetivo o adaptar una como se hizo con los errores cuadráticos. De igual forma, el número de parámetros a encontrar no sería un problema, ya que la mayoría de los lenguajes de programación tiene la capacidad para darle al cromosoma una longitud tan extensa como se deseé.

Los métodos convencionales que se utilizan para la búsqueda de soluciones resultan insuficientes para encontrar los valores óptimos de las mismas. Siempre existirá la posibilidad de aplicar nuevas teorías a los modelos ya existentes para beneficio de la materia en que se aplique.

Anexo 1

```

from struct import unpack
from os import urandom
from functools import reduce
from math import ceil,sqrt
from random import shuffle,randint
from numpy import random,cumsum,sort
from alpha_beta import PROB_F
import numpy as np
import matplotlib.pyplot as plt

def create_genotype(genes, n, chromosome=None):
    if(chromosome == None):
        if(n == 64):
            chromosome = unpack("!Q",urandom(8))[0]
        elif(n < 64):
            chromosome = ((1<<n)-1) & (unpack("!Q",urandom(8))[0])
        else:
            m = ceil(n/64)
            rand = [0] + [unpack("!Q",urandom(8))[0] for i in range(0,m)]
            chromosome=reduce(lambda x,y: (x<<64)|y, rand) & ((1<<n)-1)
        genotype={'n':n,'genes':genes,'chromosome':chromosome}
    return genotype

def create_individual(environment,genotype=None):
    if(genotype==None):
        genos=create_genotype(environment.get_gene_template(),\
            environment.get_chromosome_length())
    else:
        genos = genotype
    individual = {'genotype':genos,'phenotype':None,'fitness':None}
    return individual

def create_population(environment,parameters,individuals=None):
    size = parameters['n.individuals']
    if(individuals==None):
        individuals = [create_individual(environment)\
            for i in range(0,size)]
    population = {'size': size, 'individuals':\
        individuals,'environment':environment}

```

```

return population

def tournament(population,environment,parameters):
    k=parameters['tournament.size']
    pop=create_population(environment, parameters,\
    [max(enumerate([random.choice(population['individuals'])\
    for i in range(0,k)), key=lambda x: x[1]['fitness'])[1]\
    for i in range(0,population['size'])])
    return pop

def roulettewheel(population,environment,parameters):
    pointers = sort(random.uniform(size=population['size']))
    total_fitness = sum(population['individuals'][i]['fitness'] \
    for i in range(population['size'])) # falta desfase positivo
    probabilities = list(map(lambda x: x['fitness']/total_fitness,\
    population['individuals']))
    selected=[]
    delimiters = cumsum(probabilities)
    prueba=0
    i_p=0
    i_d=0
    while i_d<population.size:
        while pointers[i_p]<delimiters[i_d]:
            prueba+=1
            selected += [population['individuals'][i_d]]
            if(i_p+1<population['size']):
                i_p+=1
            else:
                break
        if(i_d+1<population['size']):
            i_d+=1
        else:
            break
    shuffle(selected)
    selected=selected[0:population['size']]
    return create_population(environment,parameters,selected)

def ranking(population,environment,parameters):
    n=population['size']
    pointers= random.uniform(1,(n*(n+1))>>1,n)
    population['individuals'].sort(key=lambda x:x['fitness'])
    selected= [population['individuals'][int(ceil((-1 + sqrt(1+8*p))/2)-1)]\
    for p in pointers]
    return create_population(environment,parameters,selected)

def crossover_individuals(individual1, individual2,environment):

```

```

genotype1=individual1['genotype']
genotype2=individual2['genotype']
n=genotype1['n']
crossop = randint(1,n-1)
mask1=((1<<(n-crossop))-1)<<crossop
mask2=(1<<crossop)-1
ch1= mask1 & genotype1['chromosome'] | mask2 & genotype2['chromosome']
ch2= mask1 & genotype2['chromosome'] | mask2 & genotype1['chromosome']
child1 = create_genotype(genotype1['genes'],genotype1['n'],ch1)
child2 = create_genotype(genotype1['genes'],genotype1['n'],ch2)
return [create_individual(environment,child1),\
        create_individual(environment,child2)]

```

```

def crossover(population,environment,parameters):
    new_pop = []
    for i in range(0,population['size']>>1):
        individual1 = population['individuals'][i]
        individual2 = population['individuals'][population['size']-i-1]
        if(random.uniform() < parameters['p.crossover']):
            children =\
                crossover_individuals(individual1,individual2,environment)
        else:
            children = [create_individual(environment,genotype=\
                individual1['genotype']),create_individual(environment,genotype=\
                individual2['genotype'])]
        new_pop = new_pop + children
    return create_population(environment,parameters,new_pop)

```

```

def mutate(population,environment,parameters):
    lenght = environment.get_chromosome_length()

    mutations =\
        random.binomial(lenght*population['size'],parameters['p.mutation'])

    to_mutate = [(randint(0,population['size']-1), randint(0,lenght-1)) \
        for i in range(0,mutations)]

    for ind,pos in to_mutate:
        population['individuals'][ind]['genotype']['chromosome']= \
            population['individuals'][ind]['genotype']['chromosome']^(1<<pos)

```

Lo que está sobre esta lienea no debería modificarse

```

parameters={\
    'n.individuals':100,\
    'p.mutation':.02,\
    'p.crossover':0.9,\
}

```

```

'tournament.size':2,\
'n.generations':300,\
'elitism':True,\
'selection.type':'tournament.selection'}

# Los parámetros determinan (en orden):
# número de individuos
# probabilidad de mutación
# probabilidad de cruza
# tamaño del torneo (se usa en caso de seleccionar 'torneo' como tipo de selección)
# número de generaciones
# elitismo True o elitismo = False
# tipo de selección

selections={"tournament.selection": tournament,
            "roulettwheel.selection": roulettwheel,
            "ranking.selection": ranking}

env = PROB_F()
population = create_population(env,parameters)
env.decode(population)
env.evaluate(population)
X=[]
Y=[]
for i in range(0,parameters['n.generations']):
    print(i)
    mating_pool = selections[parameters['selection.type']](\
        population,env,parameters)
    new_pop = crossover(population,env,parameters)
    mutate(new_pop,env,parameters)
    env.decode(new_pop)
    env.evaluate(new_pop)
    if('elitism' in parameters):
        inds = sorted(population['individuals']+new_pop['individuals'],\
            key=lambda x:x['fitness'], reverse=True)[:population['size']]
        print(inds[0])
        X.append(i)
        Y.append(inds[0]['fitness'])
    else:
        inds = new_pop['individuals']
        population = create_population(env,parameters,inds)

resultados=""
for i in population['individuals']:
    resultados+=str(i['phenotype'])+' '+str(i['fitness'])+"\n"
# El nombre del siguiente archivo es libre
with open("./resultados_cota_1ac.txt","w",encoding="utf-8") as file:

```

```

file.write(resultados)

plt.scatter(np.array(X), np.array(Y))
plt.show()

# Raúl Hernández en colaboración con Víctor Villegas (2018)

```

Anexo 2

```

from environment import Environment
from decoding_tools import DecodingTools as dt
import numpy as np

class PROB_F(Environment):

    # 'parameters' determina el número de parámetros a usar y 'precision' el
    # número de bits a utilizar para cada parámetro
    # E.G. si se quiere buscar 5 parámetros se debe cambiar el '2' por '5'
    def __init__(self, parameters=3, precision=40, file_name='./datos_cota.txt'):
        self.parameters=parameters
        self.precision=precision
        # Se espera que el archivo de texto tenga un dato por renglón
        file = open(file_name, 'r', encoding='utf-8')
        self.x = np.array([float(x.strip()) for x in file.readlines()])
        self.n = len(self.x)

    def evaluate(self, population):
        for individual in population['individuals']:
            # En este mismo archivo, en el método 'decode' se asignó al
            # diccionario 'individual['phenotype']' una lista ordenada con los
            # parámetros decodificados, si se tienen 5 parámetros se deben
            # asignar a variables distintas cada uno de los 5 parámetros desde
            # 'individual['phenotype'][0]' hasta 'individual['phenotype'][5]'

            sigma=individual['phenotype'][0] #sigma
            mu=individual['phenotype'][1] #mu
            x0=individual['phenotype'][2] #mu

            n = len(self.x)
            x = self.x
            bandera = True

```

```

# Aquí pongo algunas restricciones en el cálculo del 'fitness' de
# cada individuo para evitar divisiones entre 0, raíces de números
# negativos, logaritmos de números negativos, etc.
for xi in x:
    if xi<=x0: #Esta condición no permite xi=x0 para alguna i
        #if xi<x0: # Para lnL no parece necesario garantizar que sean distintos
            bandera = False
    if sigma!=0 and bandera:
        #e1 = sum(np.log(x-x0))-(n*mu)
        #e2 = ((1/(sigma**2))*(sum((np.log(x-x0)-mu)**2)))-n
        #e3 = (mu-sigma**2)*sum(1/(x-x0))-sum((1/(x-x0))*(np.log(x-x0)))
        #L = ( 1/(((sigma*np.sqrt(2*np.pi))**n)*np.product(x-x0)) ) * np.exp(-
(1/(2*(sigma**2)))*sum(((np.log(x-x0))-mu)**2))
        lnL = (-1/(2*sigma**2))*sum((np.log(x-x0)-mu)**2)-(np.log(np.product(x-x0)))-
np.log((sigma*np.sqrt(2*np.pi))**n)
    else:
        lnL=-1000
        #e1=1000
        #e2=1000
        #e3=1000
# Aquí se hace la asignación del fitness usando cálculos previos.
individual['fitness'] = lnL #-((e1**2)+(e2**2)+(e3**2))

def decode(self,population):
    for individual in population['individuals']:
        # La decodificación recibe lo siguiente (en el orden dado):
        # cromosoma, posición, tamaño e intervalo
        # Cromosoma: almacenado en 'individual['genotype']['chromosome']'
        # Posición: número que indica la posición inicial a partir de la
        # cual se va a tomar un intervalo
        # Tamaño: dado por la 'precisión'
        # Intervalo: Intervalo cerrado a considerar.
        # Ejemplo para 5 parámetros buscados en el intervalo [100,200]
        sigma =
dt.decode_real(individual['genotype']['chromosome'],0,self.precision,(0,100))
        mu =
dt.decode_real(individual['genotype']['chromosome'],self.precision,self.precision,(0,100))
        x0 =
dt.decode_real(individual['genotype']['chromosome'],self.precision*2,self.precision,(-
100,100))

        individual['phenotype'] = [sigma,mu,x0]
        # observación: 'p1' se obtiene con 'individual['phenotype'][0]'

def get_chromosome_length(self):
    return self.precision*self.parameters

```

Raúl Hernández, en colaboración con Víctor Villegas (2018)

Anexo 3

DISTRIBUCIÓN NORMAL

```
e1 = sum(self.x)-(len(self.x))*mu
e2 = ((1/(sigma**2))*(sum((self.x-mu)**2)))-len(self.x)
L = (1/(sigma*np.sqrt(2*np.pi))**self.n)*np.exp(-(1/(2*sigma**2))*sum((self.x-
mu)**2))
LnL = (-1/(2*sigma**2))*sum((self.x-mu)**2)-
(self.n*np.log(sigma*np.sqrt(2*np.pi)))
```

DISTRIBUCION LOG NORMAL 2P

```
e1 = sum(np.log(x))-(len(self.x))*mu
e2 = ((1/(sigma**2))*(sum((np.log(x)-mu)**2)))-len(self.x)
L = ( 1/(((sigma*np.sqrt(2*np.pi))**n)*np.product(x)) ) * np.exp(-
(1/(2*sigma**2))*sum((np.log(x)-mu)**2))
lnL = (-1/(2*sigma**2))*sum((np.log(x)-mu)**2)-(np.log(np.product(x)))-
np.log((sigma*np.sqrt(2*np.pi))**n)
```

DISTRIBUCION LOG NORMAL 3P

```
e1 = sum(np.log(x-x0))-(n*mu)
e2 = ((1/(sigma**2))*(sum((np.log(x-x0)-mu)**2)))-n
e3 = (mu-sigma**2)*sum(1/(x-x0))-sum((1/(x-x0))*(np.log(x-x0)))
L = ( 1/(((sigma*np.sqrt(2*np.pi))**n)*np.product(x-x0)) ) * np.exp(-
(1/(2*(sigma**2)))*sum(((np.log(x-x0))-mu)**2))
lnL = (-1/(2*sigma**2))*sum((np.log(x-x0)-mu)**2)-(np.log(np.product(x-x0)))-
np.log((sigma*np.sqrt(2*np.pi))**n)
```

DISTRIBUCION GUMBEL

```
e1 = sum(self.x*np.exp(-1*alpha*self.x))-((self.x.mean()-
(1/alpha))*sum(np.exp(-1*alpha*self.x)))
e2 = beta - ((1/alpha)*(np.log(len(self.x)/(sum(np.exp(-1*alpha*self.x))))))
L = (alpha**(n))*np.exp((-alpha*sum(self.x-beta))-sum(np.exp(-alpha*(self.x-
beta))))
lnL = self.n*np.log(alpha)-alpha*sum(self.x)+self.n*alpha*beta-sum(np.exp(-
1*alpha*(self.x-beta)))
```

DISTRIBUCION DOOBLE GUMBEL

$$\ln L = \ln L * \ln(((p * np.exp(-(1/alf1)*(xi-beta1)) - np.exp(-(1/alf1)*(xi-beta1)))) + ((1-p) * np.exp(-(1/alf2)*(xi-beta2)) - np.exp(-(1/alf2)*(xi-beta2))))))$$
$$\ln L = \ln L * \log(((p * np.exp(-(1/alf1)*(xi-beta1)) - np.exp(-(1/alf1)*(xi-beta1)))) + ((1-p) * np.exp(-(1/alf2)*(xi-beta2)) - np.exp(-(1/alf2)*(xi-beta2))))))$$

Bibliografía

- Aparicio Mijares, F. J. (2016). *Fundamentos de hidrología de superficie*. México: Limusa.
- Araujo, L., & Cervigón, C. (2009). *Algoritmos Evolutivos Un enfoque práctico*. México: Alfaomega.
- Campos Aranda, D. F. (1998). *Procesos del Ciclo Hidrológico*. San Luis Potosí: Facultad de ingeniería, Universidad Autónoma de San Luis Potosí.
- Campos Aranda, D. F. (2010). *Introducción a la Hidrología Urbana*. San Luis Potosí.
- Copca Maya, E. O. (2017). *COMPARACIÓN DE DOS MÉTODOS PARA OBTENCIÓN DE HIDROGRAMAS DE DISEÑO QUE RELACIONAN AL VOLUMEN CON EL CAUDAL MÁXIMO*. CDMX: UNAM.
- Escalante Sandoval, C. A., & Reyes Chávez, L. (2008). *Técnicas estadísticas en hidrología*. México: Facultad de Ingeniería, UNAM.
- Goldberg, E. D. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. USA: Addison-Wesley.
- Gringorten, I. (1963). "A plotting rule of extreme probability paper".
- Guzdial, M. (2013). *INTRODUCCION A LA COMPUTACION Y PROGRAMACION CON PYTHON*. Estado de México: Pearson.
- Hines, W. W., & Montgomery, D. C. (1995). *Probabilidad y estadística para ingeniería y administración*. México: Compañía editorial continental S.A. de C.V.
- Kite, G. W. (1978). *Frequency and risk analyses in hydrology*. Colorado : LithoCrafters.
- Montgomery, D. C. (1996). *Probabilidad y estadística aplicadas a la ingeniería*. México: Mc Graw Hill.
- Murray, R., & Spiegel, P. (1977). *Probabilidad y estadística*. México: Mc. Graw Hill.
- Ponce Soria, M. (2017). *Regionalización de precipitaciones para obtener lluvias de diseño en el estado de Veracruz*. CDMX: UNAM.

Van Rossum, G. (2014). *El tutorial de Python*. Argentina: PyAr.