



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE MAESTRÍA Y DOCTORADO EN INGENIERÍA
INGENIERÍA CIVIL – HIDRÁULICA

ESTIMACIÓN DE PARÁMETROS ÓPTIMOS MEDIANTE UN MODELO HÍBRIDO CON
APLICACIÓN A FUNCIONES EMPLEADAS EN HIDROLOGÍA

TESIS
QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN INGENIERÍA

PRESENTA:
CECILIA GONZÁLEZ CORREA

TUTORES PRINCIPALES:
DRA. MARITZA LILIANA ARGANIS JUÁREZ, INSTITUTO DE INGENIERÍA
DR. JOSÉ LUIS HERRERA ALANÍS, FACULTAD DE INGENIERÍA

MÉXICO, CIUDAD UNIVERSITARIA, CIUDAD DE MÉXICO, NOVIEMBRE DE 2017

JURADO ASIGNADO:

Presidente: Dr. Domínguez Mora Ramón

Secretario: Dr. Berezowsky Verduzco Moisés

Vocal: Dra. Arganis Juárez Maritza Liliana

1er. Suplente: Dr. Pozos Estrada Óscar

2do. Suplente: Dr. Herrera Alanis José Luis

Lugar donde se realizó la tesis:

INSTITUTO DE INGENIERÍA, UNAM.

TUTOR DE TESIS: Dra. Maritza Liliana Arganis Juárez y Dr. José Luis Herrera Alanis.

Agradecimientos

A dios,

A mis padres y hermana, los amo

A Víctor Uribe Martín por tu compañía, amor, paciencia y ser lo que siempre busqué

A mis amigos y compañeros por compartir el camino

A la Dra. Maritza L. Arganis Juárez por su dedicación y confianza

Al Dr. Ramón Domínguez Mora y todos mis sinodales y maestros por todas las enseñanzas,
oportunidades y dedicación

A la Universidad Nacional Autónoma de México

Al Instituto de Ingeniería, hidráulica

A CONACYT

Resumen

En el presente trabajo se realizó la adaptación de dos modelos de optimización, en un modelo híbrido; siendo el primero de ellos los Algoritmos Genéticos, que es un método de búsqueda estocástica; mientras que el segundo es un método de búsqueda geométrico. El primero converge a óptimos globales en tiempos computacionales relativamente elevados. El segundo método es Nelder-Mead y converge de manera más rápida a un óptimo. El resultado de la implementación de ambos métodos es el algoritmo híbrido y en su calibración se realizó un ejercicio de ajuste de parámetros de función exponencial. Las aplicaciones del método fueron dos, una de ellas es una función Gumbel por máxima verosimilitud para datos de precipitación de la estación Atenco, en el Estado de México, y el segundo caso fue una función doble Gumbel bivariada para los datos de escurrimiento de la Presa Infiernillo, Mich. Los resultados obtenidos muestran un ahorro de tiempo computacional y un mejoramiento en la eficiencia del ajuste de las funciones de densidad empleadas.

Contenido

Agradecimientos	3
Resumen.....	4
1. Introducción	7
1.1. Descripción	8
1.2. Objetivos	9
1.2.1. Objetivo general.....	9
1.2.2. Objetivos particulares	9
1.3. Planteamiento del problema	9
1.4. Estado del arte	10
1.4.1. A nivel mundial.....	10
1.4.2. Función de densidad doble Gumbel bivariada.....	12
2. Metodología.....	13
2.1. Descripción de un modelo matemático	13
2.2. Funciones de distribución de probabilidad.....	15
2.2.1. Función de distribución Gumbel	15
2.2.2. Función de distribución doble Gumbel	16
2.2.3. Análisis de frecuencias bivariado	17
2.3. Estimación de parámetros	25
2.3.1. Método de momentos	25
2.3.2. Método de máxima verosimilitud	26
2.3.3. Método de mínimos cuadrados	27
2.4. Análisis del error.....	28
2.4.1. Suma de error.....	28
2.4.2. Suma de los errores al cuadrado.....	28
2.4.3. Error medio cuadrático	29
2.4.4. Coeficiente de determinación (R^2)	29
2.5. Métodos de optimización	29
2.5.1. Algoritmos genéticos.....	32
2.5.2. Nelder-Mead	39
2.5.3. Algortimo Nelder-Mead	42
2.6. Método híbrido	43
2.6.1. Función objetivo.....	43

2.6.2.	Calibración de los modelos de optimización con la función exponencial.....	44
2.7.	Implementación del modelo híbrido.....	45
3.	Aplicaciones y resultados.....	47
3.1.	Caso de aplicación I. Gumbel	47
3.1.1.	Descripción del sitio de estudio	47
3.1.2.	Base de Datos	48
3.1.3.	Implementación del modelo híbrido.....	50
3.1.4.	Ajuste de la función de distribución Gumbel	50
3.2.	Caso de aplicación II. Función de distribución doble gumbel bivariada aplicada a la Presa Infiernillo.	53
3.2.1.	Descripción del sitio de estudio	53
3.2.2.	Sistema hidrológico	54
3.2.3.	Características de la presa Infiernillo	55
3.2.4.	Base de datos	56
3.2.5.	Análisis de gastos	60
3.2.6.	Ajuste de parámetros.....	67
3.2.7.	Implementación del modelo híbrido	68
3.2.8.	Ajuste de la función de distribución de densidad doble Gumbel bivariada	74
4.	Conclusiones y recomendaciones	79
5.	Referencias.....	82

1. Introducción

En varias regiones de México se presentan eventos hidroclimatológicos extremos, tales como tormentas severas, crecientes y sequías. La magnitud de un evento máximo anual está inversamente relacionada con su frecuencia de ocurrencia (Chow et al., 1994). Estos fenómenos tienen un comportamiento que se puede representar por medio de funciones de distribución mezcladas de dos o más poblaciones.

Estas funciones tienen varios parámetros, por lo que es importante contar con un método de optimización que puedan determinarlos. Por ello en este trabajo se plantea el uso conjunto de algoritmos genéticos y del método de Nelder-Mead en un método híbrido cuya finalidad es disminuir el tiempo de cálculo de los parámetros buscados con la optimización de una función objetivo.

Los Algoritmos Genéticos Holland (1975) consisten en la generación aleatoria de una población inicial de individuos cuyo desempeño se evalúa y con una serie de operadores definidos previamente se genera una nueva población, el proceso se repite hasta encontrar una solución. El método Nelder-Mead (1975) es un método de optimización simplex sin restricciones y requiere de soluciones iniciales para comenzar la optimización. El propósito del modelo híbrido es iniciar con el método de Algoritmos Genéticos, debido a que éste no requiere de puntos iniciales de búsqueda a diferencia del método Nelder-Mead, pero el costo computacional puede ser alto debido a que tiene que evaluar a todos los individuos. El método de Nelder-Mead se usa para finalizar la búsqueda debido a que converge más rápido.

Se aplicó el método a dos casos: para la función de distribución Gumbel aplicada a datos de precipitación en la estación "Atenco", Estado de México y para la función de distribución doble Gumbel bivariada aplicada a datos de escurrimiento en la presa "Infiernillo", Michoacán, considerando las estaciones La Caimanera, Los Pinzanes, Los Panches y La Pastoria.

La importancia del método bivariado radica en que permite estimar el gasto pico y el volumen de avenidas de diseño, para un período de retorno conjunto, a partir de considerar que las avenidas históricas registradas en un sitio tienen un comportamiento aleatorio representado por una función de distribución bivariada de gasto y volumen. En el caso de una función bivariada el período de retorno conjunto, se define como el lapso de tiempo en que un evento "x" sea igualado o excedido y, simultáneamente otro evento "y" sea igualado o excedido, Ramírez y Aldama (2000).

Para la función de distribución bivariada se pueden tener múltiples valores de gasto y volumen que satisfacen a la función de distribución bivariada, por lo que la selección de ellos dependerá del tipo y capacidad de las obras, buscando los escenarios más desfavorables para ser conservadores. La función de distribución doble Gumbel bivariada se calcula a través de sus funciones marginales (considerándolas por separado, para la variable gasto y para la variable volumen), las cuales pueden ser de distinto tipo, ya sea Gumbel, doble Gumbel o incluso una combinación de ellas.

Para encontrar el valor o la solución óptima, por lo general, en el campo de la hidrología se acostumbra emplear métodos tradicionales de ajuste para distintas funciones de probabilidad en

muestras de datos y calcular el error estándar; sin embargo, no existe garantía de dicho óptimo pueda localizarse de manera satisfactoria. Por ello es importante contar con métodos de optimización robustos y de eficiencia comprobable, con la finalidad de obtener los parámetros de la función de distribución que se esté empleando para un caso determinado que permitan la mejor representación del sistema en cuestión y con ellos tomar mejores elecciones, por ejemplo: del período de retorno para la ubicación de estaciones hidrométricas y/o de lluvia, entre otros.

Por otra parte, los métodos estocásticos, representan herramientas confiables, con tiempos computacionales eficientes, tanto en ejecución como en el mejoramiento de la implementación del método en cualquier programador, enfocadas a la optimización de funciones multivariables y multiobjetivo (Bonilla et al., 2005).

En el campo de la hidrología, los modelos de variables dependientes del tiempo que se emplean por lo general son de naturaleza estocástica (gasto, precipitación, etc.) ya que alguna o todas de sus variables se considera aleatoria (y se le puede ajustar una función de distribución de probabilidad). Por el contrario si todas las variables se consideran libres de la variación aleatoria tal que ninguna tiene una distribución probabilística, entonces el modelo es considerado como determinístico.

Debido a que la función doble Gumbel bivarida tiene once parámetros de ajuste, el método de optimización empleado puede conducir a un óptimo local. Por tal motivo, el empleo de métodos de optimización se hace necesario en este tipo de funciones, siendo de gran utilidad los métodos de optimización como son los algoritmos genéticos (AG) y el método Nelder-Mead, entre otros. El algoritmo híbrido corresponde a un método de optimización que utiliza lo mejor de los modelos de (AG) y Nelder-Mead para encontrar el óptimo con un menor tiempo computacional.

1.1. Descripción

Con la finalidad de cumplir con los objetivos de este trabajo, en el primer capítulo se describe, a manera de introducción una descripción general de los temas que se desarrollaron en la tesis, así como los métodos que se proponen en esta tesis, los objetivos generales y particulares, el planteamiento del problema y un esquema general de tema en el estado de arte.

En el capítulo dos, se describe de manera general, la teoría de las funciones de distribución, los modelos de optimización y la metodología empleados en la presente tesis.

El capítulo tres, corresponde a las aplicaciones y resultados. Se describen dos casos en los cuales se estiman los parámetros de ajuste de dos funciones distintas. La primera de ellas es la función de distribución Gumbel para datos de precipitación en la estación "Atenco", Estado de México y la segunda es la función de distribución doble Gumbel bivariada para datos de escurrimientos de la presa Infiernillo, Mich. Para ambos casos también se presentan los resultados.

Finalmente, en el capítulo cuatro se presentan los comentarios y recomendaciones realizadas para cada caso de aplicación y en el cinco las referencias consultadas.

1.2. Objetivos

1.2.1. Objetivo general

Aplicar los conceptos de la programación no lineal en la elaboración de un algoritmo híbrido con la finalidad de implementar un método computacional de optimización de bajo costo computacional empleado en la optimización de funciones matemáticas de forma general y demostrar que dicho algoritmo se puede aplicar de manera eficiente en el campo de la hidrología

1.2.2. Objetivos particulares

- Obtener los parámetros de funciones con aplicación general, a una función exponencial (para calibrar el algoritmo propuesto) y posteriormente a funciones aplicadas en la hidrología, tales como doble Gumbel y la función doble Gumbel bivariada; mediante algoritmos genéticos, midiendo el costo computacional empleado.
- Optimizar los parámetros mediante el método numérico llamado Nelder-Mead aplicado a diversas funciones en el campo de la hidrología, midiendo el costo computacional empleado.
- Proponer un método híbrido, producto del acoplamiento de los métodos descritos en los dos puntos anteriores con la finalidad de mejorar la optimización y efectividad de dichas técnicas.

1.3. Planteamiento del problema

Con frecuencia, en el campo de la ingeniería, necesitamos interpretar y correlacionar datos experimentales o medidos que se encuentran en forma de datos discretos; estos pueden ser aproximados mediante funciones, las cuales se pueden evaluar, integrar o diferenciar fácilmente (Quinatana et al., 2005). Por ejemplo, en el campo de la hidrología, los datos medidos por estaciones climatológicas o hidrométricas, necesitan ser estudiados mediante la construcción de funciones de los n datos medidos con la finalidad de estudiar las propiedades de dichas funciones construidas; entonces, se puede apoyar para tal finalidad mediante la estadística, herramienta la cual nos ayuda a predecir u obtener gastos de diseño de las diversas obras hidráulicas así como índices o directrices que nos indiquen la forma de medir o mejorar dichas mediciones, en función del tiempo y de las frecuencias de ocurrencia de dichos fenómenos.

El fenómeno de la lluvia va de la mano con los eventos climatológicos y dichos fenómenos presentan comportamientos caóticos, por lo que, en la actualidad, con los cambios estudiados en el clima, se hace necesario implementar metodologías nuevas a los modelos ya establecidos, tales como sistemas dinámicos y con variables aleatorias.

El objetivo es encontrar los parámetros de ajuste a una función de densidad, mediante la elección adecuada una función objetivo, la cual se pueda acoplar al problema planteado; ésta fungirá como una medida del grado de exactitud entre los parámetros calculados y los medidos encontrando de esta manera una función de distribución que mejor se ajuste a la nube de datos medidos.

Para lograr el objetivo mencionado en el párrafo anterior, fue necesario hacer un ejercicio de calibración con un ejercicio diseñado previamente y en este caso se empleó una función

exponencial, conocidos sus parámetros de ajuste y encontrando los parámetros que ajusten los datos a la función.

Para la función exponencial la función objetivo corresponde a la suma de la diferencia de los cuadrados entre los rangos de análisis, mientras que para los casos de aplicación (función de distribución Gumbel y la Función de Distribución doble Gumbel bivariada), la función objetivo fue de máxima verosimilitud.

1.4. Estado del arte

1.4.1. A nivel mundial

Los primeros métodos de optimización se remontan a los años 1600. Gracias a Isaac Newton y Gottfried Wilhelm Von Leibniz, se desarrollaron los métodos de cálculo diferencial; Johann Bernoulli, Leonhard Euler, Joseph-Louis Lagrange, Pierre Fermat y Karl Weierstrass contribuyeron con los fundamentos del cálculo de variaciones; también destacan aportaciones de Lagrange y Augustin-Louis Cauchy, Menchaca (2008).

La palabra optimización, en la actualidad se ha convertido en un término empleado ampliamente en el sector industrial, pero que ha ido tomando relevancias en varios aspectos. Las aplicaciones han sido diversas, tanto dentro de áreas tales como la física como en problemas técnicos y económicos. Sobre todo, en campos de la industria y de la ingeniería, el término es usado extensivamente.

En el caso de los métodos de optimización, se pueden identificar aquellos que involucran la programación lineal, representada por el método Simplex debido a Dantzig y a Kantorovich (1939), de manera simultánea con el desarrollo de las computadoras y la optimización matemática; posteriormente aparecen avances en los algoritmos de optimización incluyendo la programación dinámica de Bellman (1957); formas especiales de programación dinámica de Dantzing (1963) y redes de flujo como las analizadas por Ford y Fulkerson (1962). El desarrollo de estas técnicas se complementó con los avances tecnológicos de las computadoras.

Además de la programación lineal, existen otras técnicas clásicas de optimización, como por ejemplo la programación cuadrática, no lineal, geométrica o entera; cuya finalidad, es resolver una clase particular de problemas. Las técnicas de programación no lineal son métodos más generales, empleados cuando el modelo por optimizar involucra ecuaciones no lineales. *Ejemplos de estos métodos son el Gradiente Reducido Generalizado (GRG), de optimización no lineal de Lasdon y Waren (1989), programación dinámica diferencial de Jacobson y Mayne (1970), Arganis (2004).*

Existen métodos convencionales de optimización, entre los cuales destacan aquellos basados en gradiente, los cuales se dirigen hacia la solución más próxima siguiendo la dirección en la que el gradiente disminuye o aumenta, según la forma de abordar el problema (minimización o maximización), sin capacidad para discernir entre solución local y global. Nuevos métodos heurísticos se han desarrollado durante las dos últimas décadas, entre los cuales destacan el recocido simulado (RS), los algoritmos genéticos (AG), optimización con cúmulo de partículas (PSO) y más recientemente con el método de ondas gravitatorias. Estos métodos han alcanzado una gran

popularidad debido a su capacidad para acometer la resolución de problemas complejos de naturaleza muy diversa.

Los trabajos en computación evolutiva dieron comienzo a finales de los años 50 y fueron llevados a cabo por investigadores como Friedberg , Box y Bremermann entre otros Ratón (2002). El campo permaneció prácticamente en desconocimiento durante tres décadas debido en parte a la “ausencia de una plataforma computacional poderosa y defectos metodológicos de los primeros métodos”. En la década de 1960 nació la programación evolutiva con los trabajos de Fogel, mientras que las estrategias evolutivas fueron propuestas por Schwefel e Rechenberg y las estrategias evolutivas en 1975 por Holland, *quienes en conjunto serían los encargados de cambiar lentamente esta situación* Ratón (2002).

También existen métodos simplex, basados en una búsqueda directa, el cual pertenece a una clase de métodos que no requiere el uso de derivadas. El algoritmo simplex Nelder-Mead es uno de los más conocidos dentro de los paquetes de programación no lineal siendo el más conocido el método de Rosenbrock (2012).

Como bien menciona Menchaca (2008), *el método simplex de búsqueda directa fue propuesto por Spendley, Hertz y Himworth en 1962, y posteriormente mejorado por Nelder y Mead en 1965,*

En los últimos años, en un intento por mejorar la eficiencia de las técnicas existentes, para resolver problemas de optimización, se han propuesto varios algoritmos híbridos. Se entiende por algoritmo híbrido a un conjunto de pasos o procedimientos, producto a su vez, de diversos algoritmos de distinta naturaleza. Estos algoritmos han combinado algunas técnicas de optimización como: algoritmos genéticos (AGs) y el método de Nelder-Mead, Chelouah y Patrick Siarry (2003), cúmulo de partículas (PSO) y el método de Nelder-Mead (2005), evolución diferencial (ED) y el método de Nelder-Mead por Luo y Yu (2007), cúmulo de partículas y evolución diferencial por Zhang y Xie (2003), evolución diferencial y el método de actualización de multiplicadores por Lin, Kao-Shing Hwang y Wang (2002), algoritmo cultural y DPSO (híbrido de evolución diferencial y cúmulo de partículas) por Liu, Wang y You (2007) , Mathews y Fink (2000), por mencionar algunos.

Durante las dos últimas décadas, el incremento en el empleo de algoritmos genéticos (AG) se ha llevado a cabo en el campo de la planeación, ya que representan una herramienta poderosa y flexible para resolver problemas de planeación de los recursos hídricos, tal como lo plantean Nicklow y otros (2010), también se han aplicado a la optimización de la función de máxima verosimilitud para el caso las funciones Gumbel y Gumbel de dos poblaciones (Fuentes et al., 2016). En cuanto al tema de evaluación de suelos y agua Zhang y colaboradores (2009) han empleado la combinación de los algoritmos genéticos y el modelo Bayesiano para realizar simultáneamente el modelo a los datos observados, el análisis de la incertidumbre y proporcionar un intervalo de incertidumbre, Jain y Srinivasulu (2004) proponen un nuevo enfoque empleando AG codificados en forma real, con ayuda de redes neuronales, para formar modelos de lluvia- escurrimiento, usando datos diarios de precipitación y caudales de la cuenca del río Kentucky. Wang (1991) introdujo un AG para la optimización de funciones aplicadas a la calibración de un modelo conceptual de lluvia- escurrimiento.

Los métodos de optimización Nelder-Mead y los de cúmulo de partículas han demostrado ser altamente eficaces en la aplicación de problemas de optimización con restricciones. (Zahara & Kao, 2009), presentan una aplicación en el diseño de ingeniería. En el campo de la estadística, con aplicaciones en la optimización (Fan, Liang, & Zahara, 2006) proponen un método híbrido de búsqueda con el método simplex Nelder-Mead, con AG y optimización de cúmulo de partículas. En cuanto a la estimación de parámetros para funciones no lineales aplicados al flujo de ríos (Barati, 2011) propuso un algoritmo Nelder-Mead (N-M) simplex para dicha tarea. En la exploración de mínimos globales (Chelouah & Siarry, 2003) propone un algoritmo híbrido llamado "híbrido continuo que realiza la exploración de en un dominio de búsqueda para encontrar área óptimas, empleando AG y Nelder-Mead. También (Durand & Alliot, 1999) emplean el modelo híbrido de algoritmo genéticos con Nelder Mead para garantizar una búsqueda global para la solución de la función de Griewanks.

Dominique y Fuamba (2013) hacen una discusión sobre el uso del método Nelder-Mead para la estimación de parámetros empleado por Barati (2014) quien lleva a cabo un estudio estocástico referente a la distribución Beta Gumbel, empleando el método de máxima verosimilitud para estimar los parámetros del modelo, empleando un método híbrido con los métodos de optimización "quasi-Newton" y "Nelder-Mead". Addulla y Lettenmainer (1996), describen una metodología para el desarrollo de ecuaciones de estimación de parámetros regionales, diseñado para su aplicación a cuencas hidrográficas de escala continental. En cuanto a métodos nuevos de optimización Immerzeel y Droogers (2008) con base en el método Nelder-Mead hacen una calibración de modelos hidrológicos y enfatizan la falta de datos confiables, la incertidumbre en la representación de las características fisiográficas de una cuenca fluvial y presentan un enfoque innovador que incorpora la evapotranspiración con apoyo de las imágenes satelitales e imágenes tipo ráster.

La bibliografía relativa a métodos de optimización empleado en funciones para el manejo de recursos hidráulicos es extensa, con aplicación en la hidrología se tienen autores como Brownz (1972) y su contribución sobre el primer enfoque de separación de muestras, sobre el segundo enfoque de usar función de distribución combinadas: *Morris (1982) y Campos (1987), para el análisis probabilístico de avenidas máximas con poblaciones mezcladas, Campos Aranda (1987) y (1987), López (2005), Saavedra (2012).*

En cuanto a los métodos heurísticos se han desarrollado trabajos de investigación mediante algoritmos genéticos para la optimización de los parámetros de la función de distribución doble Gumbel aplicados a una serie de gastos máximos anuales, Domínguez, Mariles, Arganis Juárez (2004), entre otros.

En la mayoría de los casos los métodos descritos anteriormente no se han empleado, éstos no se han empleado de modo sencillo, debido a la falta y consistencia en la medición de datos hidrológicos de entrada para alimentar los programas, a la carencia misma de software disponible o lo la falta de adaptabilidad en problemas particulares, entre otros motivos, Arganis (2004).

1.4.2. Función de densidad doble Gumbel bivariada

El enfoque conjunto en el diseño hidrológico es un campo que se está planteando una nueva forma del análisis de las avenidas abriendo paso a un gran campo de aplicación. Este enfoque va

encaminándose a cambiar la apreciación que se tiene en cuanto al riesgo aceptado en función de un llamado período de retorno conjunto.

Es por eso que con este nuevo enfoque, se está buscando y probando el uso de otras funciones de distribución multivariadas así como un desarrollo en los métodos de ajuste de las mismas, ya que éstas están incorporando efectos del cambio climático y en el análisis de frecuencias, Ramírez (2010).

Este enfoque se está aplicando en las funciones de distribución multivariadas, por ejemplo:

Se desarrollaron inicialmente en la estimación de avenidas basado en el uso de distribución bivariadas de valores extremos de los gastos máximos y los volúmenes de escurrimientos, Aldama y Ramírez (2005); también se han realizado aplicaciones para la caracterización probabilística bivariada de avenidas de diseño, como es el caso de la tesis de “Simulación continua del comportamiento de dos vasos que trabajan en paralelo” en el caso del sistema de presas Temascal - Cerro de Oro, Rivas (2016) en el análisis, control y predicción de avenidas hidrológicas con un enfoque bivariado, a través del período de retorno, aplicando a la seguridad de presas, tomando en cuenta que este es seleccionado bajo recomendaciones emitidas por entes gubernamentales, datos estadísticos o empíricos, en el artículo “La Seguridad de presas desde la perspectiva hidrológica”, aplicándolo para el cálculo de las avenidas de diseño de las presas “El Infiernillo”, en Michoacán y Luis Donald Colosio “Huites”, en el artículo de Arganis et al (2014); en el cálculo de hidrogramas para el análisis de inundaciones prolongadas, Fuentes (2016); así como en aplicaciones de estimación de la función de distribución bivariada de avenidas históricas máximas anuales en presas de gran capacidad de regulación, en los cuales se debe considerar tanto el gasto pico como el volumen y la forma de la avenida del hidrograma en cuestión como un análisis conjunto apoyado de algoritmos genéticos para la determinación de los parámetros de la función y, a su vez son necesarios los algoritmos de cómputo evolutivo para estimar la forma de la función de distribución bivariada en la presa Luis Donald Colosio (Huites) en Sinaloa y en la obtención de avenidas de diseño en el proyecto hidroeléctrico “Las Cruces”, en el estado de Nayarit como parte del desarrollo de tesis para el grado de maestría de Restrepo (2016), entre otras aplicaciones.

2. Metodología

2.1. Descripción de un modelo matemático

Un modelo matemático es aquel que busca representar partes de la naturaleza o comportamientos, basados, generalmente en analogías, las cuales se construyen contextualizando la parte de la naturaleza que se desea describir o estudiar, con un objetivo específico, Chamizo (2010).

Gracias a los avances de la tecnología computacional de las últimas décadas y debido también a que las ramas del conocimiento que usan la modelación matemática también se han multiplicado, el empleo de diversos modelos se ha incrementado, así como sus correspondientes observaciones, mediciones y estudio de la naturaleza, economía, estudios poblacionales, medicina, entre otras disciplinas.

La modelación implica un proceso de imaginación y se entiende como una representación, por medio de un objeto, imagen, símbolo o concepto, de otro objeto o de un proceso físico, biológico, económico, etc.

En la Figura 2-1 se puede observar de manera breve el método científico.

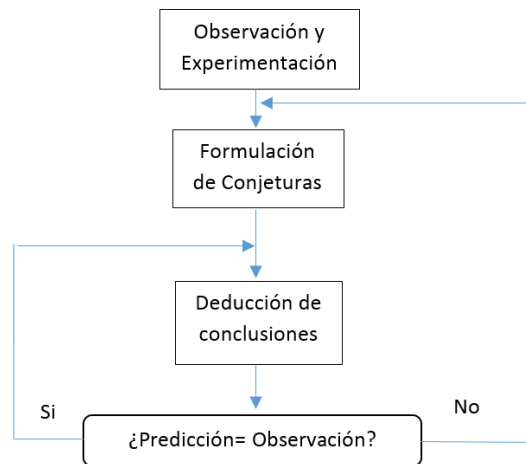


Figura 2-1. Diagrama estructural del método científico, Saavedra (2012).

Un modelo matemático es un conjunto de formalismos matemáticos, cuya finalidad de manifestar relaciones, proposiciones sustantivas de hechos, variables, parámetros, entidades y relaciones entre variables para estudiar el comportamiento de sistemas o fenómenos naturales y se entiende por modelar al hecho de buscar que un conjunto de datos se ajusten a un modelo, Real Academia Española (2017)

Un buen modelo es aquel que es capaz de representar de mejor manera los fenómenos que éste desea describir.

Establecer un modelo que describa de manera realista cualquier fenómeno que con él se quiera estudiar es un proceso que debe ser considerado dinámico, ya que los fenómenos de la naturaleza o de otros campos suelen estar en función del tiempo. Por ello, la elaboración de un modelo es una tarea con un grado de complejidad alto, ya que para cada momento y para cada fenómeno se necesitaría renovarlo. Existen modelos de diversas índoles, como por ejemplo:

- *Análogos, los cuales imitan al objeto de estudio modificando su escala, como la maqueta de una casa o de un sistema solar.*
- *Dragmáticos, que usan imágenes, dibujos o diagramas para describir el objeto de estudio, y*
- *Modelos conceptuales, que recurren a ideas para representar a los objetos de estudio, como los modelos matemáticos (Saavedra Barrera P. , 2012)*

De esta manera se puede enunciar que, “un modelo matemático es una representación abstracta, que expresa, en lenguaje matemático, un proceso, fenómeno o sistema físico, biológico, económico, social, etc.

Ahora bien, para iniciar el planteamiento de un modelo matemático, el primer paso es realizar suposiciones o hipótesis, por ejemplo, al tratar de modelar la trayectoria de una bala, se supone, por ejemplo, que el día es claro y sin viento, lo cual nos permite suponer que la bala se moverá en un plano y también se puede suponer que la fricción del aire no es significativa, Saavedra (2012).

Según Escalante y Reyes (2008), los diversos tipos de modelos son:

- ✓ Determinístico es aquel que se obtiene a través de relaciones físicas y no requiere de datos experimentales para su aplicación.
- ✓ Paramétrico puede ser considerado como un determinístico en el sentido de que una vez que se estiman los parámetros del modelo, éste siempre genera la misma salida a partir de la información de entrada. Por lo tanto un modelo paramétrico es estocástico en el sentido de que los parámetros estimados dependen de los datos observados y aquellos cambiarán cuando los datos observados también lo hagan.
- ✓ Estocástico es aquel que las salidas son predecibles solamente en un sentido estadístico. En un modelo estocástico, el empleo repetido de un grupo dado de entradas del modelo genera salidas que no son las mismas pero siguen cierto modelo estadístico.

Antes de hacer inferencias de cualquier modelo es importante la estimación de sus parámetros.

2.2. Funciones de distribución de probabilidad

Se mencionan brevemente las funciones de distribución de probabilidad (FDP) más empleadas en el campo de la hidrología.

2.2.1. Función de distribución Gumbel

La distribución de valores extremos tipo I o distribución Gumbel nace con Fisher y Tippett, en los años veinte con la teoría de valores extremos, Escalante y Reyes (2002). En los años treinta, Gumbel realiza aplicaciones prácticas usando los estadísticos de valores extremos de distribuciones del tiempo de duración de la vida humana y en 1941 publica algunos artículos con respecto a sus aplicaciones en el análisis de frecuencias de gastos máximos y mínimos. Jenkinson (1955) demostró que esta distribución es un caso especial de la distribución general de valores extremos, Chow *et al.* (1994).

La función de distribución acumulada está dada por la siguiente ecuación:

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

2.2-1

Con:

$$-\infty < x < \infty, -\infty < \beta < \infty \text{ y } \alpha > 0$$

Y su función de densidad de probabilidad es:

$$f(x) = -\alpha \left\{ \left[e^{-\alpha(x-\beta)} \right] \left[e^{-e^{-\alpha(x-\beta)}} \right] \right\}$$

2.2-2

Donde:

α = parámetro de escala

β = parámetro de ubicación

Los parámetros de la Función de Distribución de Probabilidad (FDP) se pueden estimar mediante varios métodos como son el método de momentos, máxima verosimilitud, máxima entropía, etc.

Mientras que para estimar los eventos correspondientes de una probabilidad de excedencia $\{1 - [F(x)]\}$ o para un período de retorno Tr , es necesario primero calcular los parámetros de ajustes y posteriormente se despeja la variable x de la función de distribución acumulada (*fda*), Gómez *et al* (2010), tal y como se presenta en la Ecuación 2.2-3.

$$x = \beta - \alpha \ln\{-\ln[F(x)]\}$$

2.2-3

2.2.2. Función de distribución doble Gumbel

Debido a que muchas regiones de nuestro país son afectadas esporádicamente por ciclones, los registros hidrométricos de gastos máximos anuales están conformados por eventos meteorológicos de dos tipos; uno de ellos se debe a las precipitaciones características de la región y otro de ellos, pertenece a las lluvias de origen ciclónico, siendo éstas últimas las que provocan las avenidas más grandes. Por ello, se hace necesario el ajuste de una función de distribución de probabilidad de dos poblaciones, ya que, por ejemplo la función de distribución tipo Gumbel, no resulta del todo satisfactoria.

Como Campos (1989) recalca, es necesario hacer análisis de frecuencias considerando que los gastos máximos anuales tienen dos tipos de origen, mismos que dan lugar a una función de distribución mezclada o de dos poblaciones.

Para realizar el análisis de frecuencias, es posible aplicar la función de distribución doble Gumbel para dos poblaciones considerando que los grupos son mutuamente excluyentes, es decir el valor de la variable se debe a un evento ciclónico o no. La *fda* (función de distribución acumulada) para una muestra que presente la presencia de dos poblaciones está definida como la Ecuación 2.2-4, descrita por Hann (1977), de la siguiente manera.

$$F(x) = P \exp\{-\exp(-\alpha_1(x - \beta_1))\} + (1 - P) \exp\{-\exp(-\alpha_2(x - \beta_2))\}$$

2.2-4

Donde:

P es la probabilidad de tener eventos no ciclónicos, adimensional.

x es la variable aleatoria para la cual se estima la probabilidad de no excedencia

α_1 es el parámetro de escala de la población no ciclónica

β_1 es el parámetro de ubicación de la población no ciclónica

α_2 es el parámetro de escala de la población ciclónica

β_2 es el parámetro de ubicación de la población ciclónica

La función de densidad de probabilidad, descrita por Gómez *et al.* (2010), se representa como:

$$f(x) = \frac{dF(x)}{dx} = \alpha_1 P * \exp\{-\alpha_1(x - \beta_1) - \exp(-\alpha_1(x - \beta_1))\} \\ + \alpha_2(1 - P)\exp\{-\alpha_2(x - \beta_2) - \exp[-\alpha_2(x - \beta_2)]\}$$

2.2-5

De la misma manera, esta función de distribución se puede resolver por diversos métodos; mientras que el cálculo del evento correspondiente a una probabilidad de excedencia $(1 - F(x))$ o un período de retorno Tr se realiza por iteraciones o por algún método numérico como puede ser Newton-Raphson, debido a que la variable x no se puede despejar de la *fd*. Por ejemplo para una $F(x) = 0.999$, equivalente a un período de retorno de 1000 años ($Tr = 1000$ años), se sustituyen parámetros de ajuste y se proponen valores de x hasta que se cumpla la igualdad representada en la Ecuación 2.2-6:

$$0.999 = P \exp\{-\exp[-\alpha_1(x - \beta_1)]\} + (1 - P) \exp\{-\exp[-\alpha_2(x - \beta_2)]\}$$

2.2-6

2.2.3. Análisis de frecuencias bivariado

En cuanto a la estimación de avenidas, ya sea en el diseño o revisión de obras hidráulicas, en el análisis de la capacidad de control y las obras de excedencias de vasos, por lo general, consiste en un método univariado, que considera la variable aleatoria del gasto pico de la avenida máxima anual registrada; sin embargo, en dichos análisis de frecuencias es necesario conocer la forma completa de la avenida, ya que la respuesta de los vasos de almacenamiento es sensible tanto al gasto pico como al volumen de escurrimiento de las avenidas, Gómez *et al.*, (2010).

Al emplear el análisis de frecuencias univariado, se determina el gasto para un período de retorno y para el cálculo del volumen correspondiente, por lo general se realiza de forma arbitraria, por lo que se puede hablar de un período de retorno del gasto pico, mas no del período de retorno de la avenida completa. Es por este motivo que la definición de período de retorno conjunto toma importancia, ya que este término determina el período de retorno del hidrograma completo.

Para fines de diseño, los modelos existentes de lluvia-escurrimiento no funcionan bien en condiciones extremas, que son las que precisamente resultan de interés para fines de diseño. Esto se debe en parte a que los parámetros de los modelos se calibran para condiciones medias, puesto que los datos para condiciones extremas no están disponibles en la mayoría de los casos. Cuando los valores de dichos parámetros no se identifican correctamente, los errores se pueden presentar en magnitudes considerables en la estimación del escurrimiento, Ramírez y Aldama (2010). Adicionalmente los conceptos empleados en la construcción de modelos simples que trabajan razonablemente bien en condiciones ideales pueden ser inválidos para condiciones extremas. Sólo en cuencas pequeñas se satisfacen razonablemente las hipótesis de la mayoría de los modelos.

A las dificultades descritas, se añade el hecho de que resulta prácticamente imposible estimar un periodo de retorno asociado con una avenida estimada con base en datos de lluvia transformados en escurrimiento. Lo anterior obedece a que el proceso lluvia-escurrimiento es altamente no lineal.

2.2.3.1. Función de distribución bivariada

La ecuación general de la función de la función bivariada para valores extremos, dada por Escalante (2007) se representa en la Ecuación 2.2-7.

$$F(x, y) = \exp \left\{ - \left[(-\ln F(x))^m + (-\ln F(y))^m \right]^{1/m} \right\}$$

2.2-7

Donde $F(x, y)$ es la función de distribución de probabilidad simultánea para los parámetros X y Y , como $X \leq x$ y $Y \leq y$; $F(x)$ y $F(y)$ son las distribuciones marginales y m es el parámetro de asociación que depende del grado de interdependencia entre la variable aleatorias X y Y ; cuando $m = 1$ se representa el caso de independencia estocástica y cuando $m \rightarrow \infty$ se presenta el caso de dependencia completa, Dueñas, (2016).

Las funciones marginales pueden estar dadas por una función de distribución ya sea Gumbel o Doble Gumbel Domínguez (2009).

2.2.3.1.1. Función de densidad bivariada

Ramírez y Aldama (2000) y Aldama y Ramírez (2002), proponen:

$$\tilde{Q}_p = \frac{Q_p - u_q}{\alpha_q}$$

$$\tilde{V} = \frac{V - u_v}{\alpha_v}$$

Donde \tilde{Q}_p y \tilde{V} son los valores normalizados del gasto pico y del volumen pico, respectivamente y u_q, α_q, u_v y α_v son los parámetros de ubicación y escala para la distribución de mejor ajuste de cada variable; de tal forma que $\tilde{F}_{qv}(\tilde{Q}_p, \tilde{V})$ representa la distribución bivariada adimensional de la descarga de pico y de volumen, cuyas distribuciones marginales se denotan como $\tilde{F}_q(\tilde{Q}_p)$ y $\tilde{F}_v(\tilde{V})$.

De acuerdo con Ramírez y Aldama, una función de distribución multivariada de valores extremos debe satisfacer las restricciones de Fréchet y las condiciones impuestas por el postulado de estabilidad, Ramírez y Aldama (2000); Aldama y Ramírez (2002); por lo que el modelo logístico toma la forma de:

$$\tilde{F}_{qv}(\tilde{Q}_p, \tilde{V}) = \exp\left[-\{-\ln\tilde{F}_q(\tilde{Q}_p)\}^m + \{-\ln\tilde{F}_v(\tilde{V})\}^m\right]^{1/m}$$

Donde $m \in [1, \infty)$, representa el parámetro de asociación; cuando $m = 1$ se presenta el caso de independencia estocástica y cuando $m \rightarrow \infty$ es el caso de dependencia completa. Para satisfacer el postulado de estabilidad de Fréchet descrito por Ramírez y Aldama (2000); Aldama y Ramírez (2002); las distribuciones marginales univariadas de valores extremos $\tilde{F}_q(\tilde{Q}_p)$ y $\tilde{F}_v(\tilde{V})$, pueden ser uno de los tres tipos siguientes: Gumbel, Fréchet y Weibull.

El parámetro de asociación m presentado en la ecuación anterior puede calcularse en términos del coeficiente de correlación entre \tilde{Q}_p y \tilde{V} dados en las ecuaciones 3 y 4, expuestas anteriormente

En forma general:

La función de densidad $f(x, y)$ de la función de distribución bivariada $F(x, y)$, se obtiene como:

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial y \partial x}$$

2.2-8

Si la función $g(x, y)$ se define como;

$$g(x, y) = -[\{-\ln F(x)\}^m + \{-\ln F(y)\}^m]^{1/m}$$

2.2-9

Considerando las derivadas parciales de la función anterior, la función de densidad puede ser expresada como:

$$f(x, y) = e^{g(x, y)} \left[\frac{\partial g(x, y)}{\partial x} \frac{\partial g(x, y)}{\partial y} + \frac{\partial^2 g(x, y)}{\partial y \partial x} \right]$$

2.2-10

La derivada parcial de $g(x, y)$ con respecto a x es:

$$\frac{\partial g(x, y)}{\partial x} = -\left[(-\ln F(x))^m + (-\ln F(y))^m\right]^{\left(\frac{1-m}{m}\right)} \left[\frac{(-\ln F(x))^{m-1}}{F(x)} \frac{\partial F(x)}{\partial x} \right]$$

2.2-11

De manera similar, para la derivada parcial de $g(x, y)$ con respecto a y :

$$\frac{\partial g(x, y)}{\partial y} = - \left[(-\ln F(x))^m + (-\ln F(y))^m \right]^{\left(\frac{1-m}{m}\right)} \left[\frac{(-\ln F(y))^{m-1} \partial F(y)}{F(y)} \frac{\partial F(y)}{\partial y} \right]$$

2.2-12

La derivada parcial simultánea con respecto a y y a x de $g(x, y)$ se obtiene por las derivadas parciales de la Ecuación 2-12 con respecto a y :

$$\begin{aligned} \frac{\partial^2 g(x, y)}{\partial y \partial x} = & - \left\{ \left[(-\ln F(x))^m + (-\ln F(y))^m \right]^{\left(\frac{1-m}{m}\right)} \left[\frac{(-\ln F(x))^{m-1} \partial F(x)}{F(x)} \frac{\partial F(x)}{\partial x} \right] \right. \\ & \left. + \left[(-\ln F(x))^m + (-\ln F(y))^m \right]^{\left(\frac{1-m}{m}\right)} \left[\frac{(-\ln F(y))^{m-1} \partial F(y)}{F(y)} \frac{\partial F(y)}{\partial y} \right] \right\} \end{aligned}$$

2.2-13

Simplificando:

$$\begin{aligned} \frac{\partial^2 g(x, y)}{\partial y \partial x} = & (m \\ & - 1) \left[(-\ln F(x))^m \right. \\ & \left. + (-\ln F(y))^m \right]^{\left(\frac{1-2m}{m}\right)} \left[\frac{(-\ln F(x))^{m-1} \partial F(x)}{F(x)} \frac{\partial F(x)}{\partial x} \right] \left[\frac{(-\ln F(y))^{m-1} \partial F(y)}{F(y)} \frac{\partial F(y)}{\partial y} \right] \end{aligned}$$

2.2-14

Con aproximaciones diferenciales y definiendo la función $v(x, y)$ como:

$$v(x, y) = (-\ln F(x))^m + (-\ln F(y))^m$$

2.2-15

El valor de $f(x, y)$ puede ser obtenida analíticamente como la Ecuación 2.2-16, Domínguez (2009).

$$f(x, y) = \frac{m}{F(x)F(y)} \frac{\partial F(x)}{\partial x} \frac{\partial F(y)}{\partial y} \left[(-\ln F(x))(-\ln F(y)) \right]^{m-1} e^{-v^{m-1}} v^{\frac{1}{m}-2} \left[\left(1 - \frac{1}{m}\right) + \frac{1}{m} v^{\frac{1}{m}} \right]$$

Ecuación 2.2-16

Donde $F(x)$ y $F(y)$ son funciones marginales para los parámetros de segregación, gasto y volumen respectivamente.

Finalmente la ecuación que se de sea optimizar puede ser reasentada como:

$$\text{máx}_{(Q_p, V)} \tilde{F}_{qv}(\tilde{Q}_p, \tilde{V}) = \tilde{F}_{qv}(\tilde{Q}_p, \tilde{V})(Q_p, V)$$

2.2-17

Período de retorno conjunto

La Ecuación 2.2-17 se encuentra sujeta a un periodo de retorno conjunto, el cual se describe a continuación:

Para una variable aleatoria X , el período de retorno T , en años, correspondiente a un evento de magnitud x , se define como el inverso de la probabilidad de que en un año cualquiera, el evento x sea igualado o excedido, esto es:

$$T = \frac{1}{P(X \geq x)}$$

Es decir; el inverso de dicho período de retorno, en términos de la probabilidad de no excedencia es:

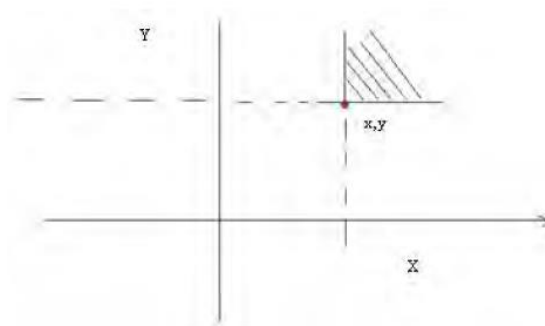
$$\frac{1}{T} = P(X \geq x) = 1 - F(x)$$

Para el caso de una función bivariada, Ramírez y Aldama consideran que el evento correspondiente al gasto máximo anual x y el volumen máximo anual y son excedidos simultáneamente, de tal manera que el período de retorno es:

$$T = \frac{1}{P(X \geq x, Y \geq y)}$$

Como se observa en la siguiente figura, el inverso de dicho período de retorno, en términos de la función de distribución bivariada y de las funciones marginales, se obtiene como:

$$\frac{1}{T} = P(X \geq x, Y \geq y) = 1 - F(x) - F(y) + F(x, y)$$



Finalmente, se dice que la Ec. 2.2.17 está sujeta a

$$T_{Q_p V} = \frac{1}{1 - F_q(Q_p) - F_V(V) + F_{q,V}(Q_p, V)}$$

2.2-18

2.2.3.2. Función Doble Gumbel bivariada

La función de distribución doble Gumbel bivariada (FDDGB) establece una relación entre dos variables aleatorias, cada una de estas variables está constituida por dos poblaciones, con datos de magnitud mayores, mismas que se identifican como condiciones extremas, Domínguez *et al.*, 2009.

El modelo bivariado que se utiliza tiene la forma de la Ecuación 2.2-8, usando funciones marginales doble Gumbel.

Las funciones marginales de una función doble Gumbel puede ser expresada como:

$$F(x) = \exp\left\{-\exp\left(-\frac{x+a_1}{c_1}\right)\right\} p_x + \exp\left\{-\exp\left(-\frac{x+a_2}{c_2}\right)\right\} (1-p_x)$$

2.2-19

$$F(y) = \exp\left\{-\exp\left(-\frac{y+a_3}{c_3}\right)\right\} p_y + \exp\left\{-\exp\left(-\frac{y+a_4}{c_4}\right)\right\} (1-p_y)$$

2.2-20

Donde a_1, a_2, a_3 y a_4 son parámetros de ubicación; c_1, c_2, c_3 y c_4 son parámetros de escala; p_x y p_y son parámetros de segregación de las variables x y y , respectivamente.

Las derivadas de las funciones marginales $F(x)$ y $F(y)$, doble Gumbel están dadas por, Domínguez *et al.*, (2009):

$$\frac{\partial F(x)}{\partial x} = \frac{p_x}{c_1} e^{-e^{-\left(\frac{x+a_1}{c_1}\right)}} e^{-\left(\frac{x+a_1}{c_1}\right)} + (1-p_x) \frac{1}{c_2} e^{-e^{-\left(\frac{x+a_2}{c_2}\right)}} e^{-\left(\frac{x+a_2}{c_2}\right)}$$

2.2-21

$$\frac{\partial F(y)}{\partial y} = \frac{p_y}{c_3} e^{-e^{-\left(\frac{y+a_3}{c_3}\right)}} e^{-\left(\frac{y+a_3}{c_3}\right)} + (1-p_y) \frac{1}{c_4} e^{-e^{-\left(\frac{y+a_4}{c_4}\right)}} e^{-\left(\frac{y+a_4}{c_4}\right)}$$

2.2-22

Para obtener los 11 parámetros de esta función de distribución con método de máxima verosimilitud se pueden usar como apoyo las ecuaciones 2.2-14 a 2.2-21 para el uso de la función de densidad que se aplica en la función de verosimilitud L .

La precisión en los cálculos de la FDDGB $f(x, y)$ para la evaluación de la función objetivo, que para este caso se trata del método de máxima verosimilitud, es de suma importancia, debido a que por el método por mínimos cuadrados o de momentos se hace impráctico, a causa de que el número de ecuaciones no lineales necesarias para encontrar varios parámetros se incrementa conforme al número de momentos necesarios también se incrementan.

2.2.3.3. Período de retorno conjunto

Al no conocer la forma del hidrograma, ya que se hace por lo general, con el proceso de mayoración, Domínguez *et al.*, (1980), se puede decir, que no se conoce el período de retorno del hidrograma completo dada la definición arbitraria de su forma y por ende, su volumen mismo, que carece de una base probabilística, Ramírez (2010), por tal motivo, el término “estimación de avenidas de diseño”, basado en el enfoque hidrométrico, es sinónimo de “estimación de gastos máximos de diseño”.

Por un lado, se ha reconocido que el gasto máximo del hidrograma es una variable aceptable para el diseño solamente cuando la variación temporal del almacenamiento de la cuenca no es un factor importante en el proceso de generación de escurrimiento, McCuen (2005), pero, para el caso de vasos de almacenamiento, es bien sabido que éstos exhiben una diferente sensibilidad al gasto pico que al volumen de almacenamiento.

En algunos casos, la respuesta de un embalse al transitar la avenida es más sensible al volumen de la misma que al gasto pico. Algunos autores han concluido que una parametrización bivariada en función del gasto pico y del volumen es suficiente para caracterizar una avenida con fines de diseño en vasos.

El concepto de período de retorno T , para una variable aleatoria X , correspondiente a un evento de magnitud x , se define como el inverso de la probabilidad de que en un año cualquiera, el evento x sea igualado o excedido, Dueñas (2016).

$$T = \frac{1}{P(X > x)}$$

2.2-23

Es decir, el inverso de dicho período de retorno, en términos de la probabilidad de no excedencia es

$$\frac{1}{T} = P(X > x) = 1 - F(x)$$

2.2-24

Para el caso de una función bivariada, Ramírez y Aldama consideran que el evento correspondiente al gasto máximo anual x y el volumen máximo anual y son excedidos simultáneamente, de tal manera que el período de retorno es, Dueñas (2016):

$$T = \frac{1}{P(X > x, Y > y)}$$

2.2-25

La caracterización probabilista de un hidrograma biparamétrico puede construirse a través de una función bivariada $F_{Q_p, V}(Q_p, V)$, la cual representa la probabilidad de no excedencia conjunta de gasto pico y volumen de escurrimientos del hidrograma. Se puede demostrar que el período de retorno asociado con esta probabilidad, denominado período de retorno conjunto, está dado por Ramírez y Aldama (2010):

$$Tr = \frac{1}{1 - F_{Q_p}(Q_p) - F_v(V) + F_{Q_p, V}(Q_p, V)}$$

2.2-26

En donde la función bivariada $F_{Q_p, V}(Q_p, V)$, puede construirse mediante el modelo logístico propuesto por Gumbel (1960) como:

$$F_{Q_p, V}(Q_p, V) = \exp\left\{-\left[(-\ln F_{Q_p}(Q_p))^m + (-\ln F_v(V))^m\right]^{\frac{1}{m}}\right\}$$

2.2-27

Donde, las funciones de gasto y volumen, respectivamente son:

$$F_{Q_p}(Q_p) = \int_{-\infty}^{\infty} F_{Q_p, V}(Q_p, V) dV$$

2.2-28

$$F_v(V) = \int_{-\infty}^{\infty} F_{Q_p, V}(Q_p, V) dQ_p$$

2.2-29

Que son las funciones de distribución marginales para gastos pico y volumen, respectivamente.

El parámetro m es un parámetro de asociación que depende de la correlación lineal entre gastos y volúmenes. De acuerdo con la naturaleza de las variables tratadas es posible utilizar distribuciones marginales tipo Gumbel o Gumbel mixta (si hay evidencia de la existencia de poblaciones mezcladas).

La función de distribución de probabilidad bivariada y por ende las distribuciones marginales se obtienen a partir de los registros existente de gastos máximos anuales y volúmenes de escurrimiento máximos anuales en el sitio en cuestión.

Debe notarse que para un período de retorno conjunto, la Ecuación 2.2-28 se satisface por una infinidad de pares de valores (Q_p, V) . Para resolver de forma única este problema, se propone determinar cuál de esos pares de valores produce los efectos más desfavorables sobre la presa por diseñar o revisar. Por ejemplo si $máx_{(Q_p, V)} \tilde{F}_{qv}(\tilde{Q}_p, \tilde{V})$ representa la máxima elevación de la superficie libre del agua que se alcanza dentro del vaso una vez transitado el hidrograma, entonces la avenida de diseño queda definida al resolver el problema de optimización no lineal planteado por:

$$Máx Z_m(Q_p, V)$$

2.2-30

Sujeto a:

$$Tr = \frac{1}{1 - F_{Q_p}(Q_p) - F_v(V) + F_{Q_p, V}(Q_p, V)}$$

2.2-31

Para un período dado.

Este proceso de solución incorpora, de forma natural, las características del vaso en cuestión dentro del proceso de estimación de su propia avenida de diseño, ya que la curva elevaciones-capacidades, las características de la obra de excedencias y la política de operación de ésta, se ven involucradas

en el proceso de tránsito de la avenida por el vaso, que es la herramienta usada para evaluar los impactos sobre el vaso.

2.3. Estimación de parámetros

Las técnicas para la estimación de parámetros, se aplican, en la presente tesis a manera de función objetivo, ya que permiten evaluar los parámetros que se quieren optimizar para la búsqueda del mejor ajuste de la función en cuestión en función los datos trabajados.

Se buscan ajustar a una función de distribución de probabilidad los datos hidrológicos medidos. Esta función está conformada por la variable aleatoria analizada, como por ejemplo gastos, escurrimientos, precipitación, entre otros y por los parámetros que describen dicha función, según Gómez *et al.*, (2010).

Modelos aplicados en la hidrología

Los modelos descritos en este apartado también son aplicables para resolver problemas en la hidrología. “Los modelos pueden ser del tipo determinístico, parámetros o estocástico” Escalante y Reyes (2008).

Cada estimador de un parámetro es una función de los valores de la muestra, los cuales son observaciones de una variable aleatoria. Así, el propio parámetro estimado es una variable aleatoria que tiene su propia distribución muestral. Un estimador que se obtiene a partir de un grupo de valores puede considerarse como un valor observado de una variable aleatoria, Escalante y Reyes (2008).

Las técnicas de estimación mas empleadas en la hidrología, según Aparicio (2010), se enumeran a continuación:

2.3.1. Método de momentos

Este procedimiento se emplea para estimar uno o más parámetros de una función.

El *k*-ésimo momento de la distribución, tomado respecto al origen y a la media es

$$M_k = \int_{-\infty}^{\infty} x^k f(x) dx$$

2.3-1

Con respecto al origen y

$$M_k = \int_{-\infty}^{\infty} (x - \mu)^k f(x) dx$$

2.3-2

Con respecto a la media.

Los primeros cuatro momentos (insesgados) de la muestra son

$$M_1 = \frac{1}{n} \sum_{i=1}^n x_i = x$$

2.3-3

$$M_2 = S^2 = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - x)$$

2.3-4

$$M_3 = S^3 C_s$$

2.3-5

Donde C_s es el coeficiente de sesgo definido como

$$C_s = \frac{n}{(n-1)(n-2)} \sum_{i=1}^n \frac{(x_i - x)^3}{S^3}$$

2.3-6

$$M_4 = S^4 K$$

2.3-7

Donde K es la curtosis que se calcula como

$$K = \frac{n^3}{(n-1)(n-2)(n-3)} \sum_{i=1}^n \frac{(x_i - x)^4}{S^4}$$

2.3-8

La estimación de los parámetros desconocidos por este método se logra al igualar los momentos correspondientes a la población (función de distribución de probabilidad) con los momentos de la muestra, obteniendo así n ecuaciones con n incógnitas (los parámetros buscados). Este método, a pesar de ser de fácil aplicación, requiere del cálculo de integrales, las cuales contienen a la función de densidad $f(x)$, haciendo de esta manera complejo el procedimiento.

2.3.2. Método de máxima verosimilitud

Si se tienen n observaciones aleatorias x_1, x_2, \dots, x_n , su distribución conjunta es

$$f(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_m) \text{ o } f(x_1, x_2, \dots, x_n; \theta)$$

2.3-9

Como la muestra es aleatoria, los valores de las observaciones son independientes y entonces su función de distribución conjunta es:

$$f(x_1, x_2, \dots, x_n; \theta) = f(x_1; \theta) \cdot f(x_2; \theta) \dots f(x_n; \theta)$$

2.3-10

La ecuación anterior es proporcional a la probabilidad de que la muestra, o una muestra aleatoria particular, pertenezca a la distribución, y se conoce con el nombre de “función de verosimilitud (L)”

$$L = \prod_{i=1}^n f(x_i; \theta)$$

2.3-11

Donde θ es el vector de parámetros desconocidos.

Los valores de los parámetros que maximizan L se conocen como estimadores de máxima verosimilitud.

Para poder estimar los parámetros que maximicen la función de verosimilitud es necesario calcular las derivadas de dicha función con respecto a cada uno de los parámetros a estimar e igualar cada derivada a cero y así obtener un sistema de ecuaciones, el cual será necesario resolver para obtener los parámetros buscados, esto es

$$\frac{\partial L}{\partial \theta_i} = 0, \forall i$$

2.3-12

Dado que L es un producto de funciones, la derivación de tales productos es complicada, por lo que es más fácil encontrar el valor de los parámetros que maximicen a la función logarítmica de L , ya que $\ln L$ es una función monótona creciente de L , por lo que los valores de los parámetros que maximicen a L corresponden también a los valores de los parámetros que maximizan a $\ln L$, por lo que la ecuación anterior queda

$$\frac{\partial \ln L}{\partial \theta_i} = 0, \forall i$$

2.3-13

Cuando el número de parámetros por determinar es mayor a uno, el sistema de ecuaciones que se obtiene con la Ecuación 2.3-13 es un sistema no lineal, por lo que para resolver dicho sistema es necesario recurrir a métodos numéricos, Gómez *et al.*, (2010).

2.3.3. Método de mínimos cuadrados

El método de mínimos cuadrados consiste en hallar una función analítica de forma sencilla que coincida con casi todos los pares ordenados de los datos en cuestión, cabe mencionar que la curva ajustada no necesariamente tiene que pasar por todos los puntos de los datos analizados. De modo tal, que la ecuación buscada debe satisfacer la condición de minimizar la suma de las desviaciones (d_i) del comportamiento de cada par de datos discretos, con respecto al comportamiento del modelo propuesto, elevados al cuadrado; es decir:

$$\sum_{i=1}^n (d_i)^2 = 0$$

2.3-14

El caso más simple de una aproximación por mínimos cuadrados es la regresión lineal, la cual consiste en ajustar a una línea recta un conjunto de datos discretos, menciona Quintana *et al.*, (2005).

Sin embargo, también existen modelos no lineales, los cuales, mediante cierto tratamiento matemático se pueden comportar como lineales, por ejemplo, cuando un cierto conjunto de datos discretos tienen una tendencia no lineal, es posible proponer un modelo exponencial o bien también se puede buscar que dichos datos se ajusten a una ecuación potencial.

2.4. Análisis del error

En la aplicación de cálculos numéricos es importante tener en cuenta la aproximación de los resultados calculados con el objetivo buscado. Es por tal motivo, que la dispersión del conjunto de valores obtenidos respecto a los medidos puede incrementar debido a diversos motivos, dependiendo del problema y forma de resolución del mismo. Esta dispersión se puede medir por medio del cálculo de su error.

En general, el incremento del error se debe al método de optimización elegido y también por el método de programación del mismo. El conocimiento en cuanto a la medición del error nos ayuda a construir algoritmos numéricos adecuados.

Hay diversas formas de medir el error; las que se aplican en esta tesis son las siguientes:

2.4.1. Suma de error

Para un conjunto de n datos medidos, la suma del error corresponde a la sumatoria de la diferencia de los n datos medidos menos los calculados, de la forma siguiente:

$$\text{Suma del error} = \sum_{i=1}^n (y_{\text{medidos}} - y_{\text{calculados}})$$

2.4-1

Donde:

y_{medidos} , son los datos medidos

$y_{\text{calculados}}$, son los datos calculados

n es el número de datos de la muestra

2.4.2. Suma de los errores al cuadrado

Es la sumatoria del cuadrado de la diferencia entre los datos medidos y calculados (Ver Ecuación 2.4-2).

$$\text{Suma de los errores al cuadrado} = \sum_{i=1}^n (y_{\text{medidos}} - y_{\text{calculados}})^2$$

2.4-2

2.4.3. Error medio cuadrático

A diferencia del método anterior, el error medio cuadrático es el promedio de los n errores al cuadrado, Ecuación 2.4-3.

$$\text{Error medio cuadrático} = \frac{1}{n} \sum_{i=1}^n (y_{\text{medidos}} - y_{\text{calculados}})^2$$

2.4-3

2.4.4. Coeficiente de determinación (R^2)

El coeficiente de determinación se calcula a través de la Ecuación 2.4-4

$$R^2 = \frac{\sigma^2 y_{\text{medidos}} - \sigma^2 \text{error}}{\sigma^2 y_{\text{medidos}}}$$

2.4-4

Donde

$\sigma^2 y_{\text{medidos}}$ es la varianza de los datos medidos

$\sigma^2 y_{\text{medidos}}$ es la varianza del error

Error = $y_{\text{medidos}} - y_{\text{calculados}}$

2.5. Métodos de optimización

Un método de optimización se puede describir como aquel que tiene un conjunto de algoritmos o procedimientos que tienen la finalidad de obtener los mejores resultados posibles dados ciertas condiciones.

Goldberg (1999), habla de tres tipos de métodos de optimización: los basados en cálculo, los enumerativos y los aleatorios.

Los métodos basados en cálculos han sido estudiados arduamente y a su vez se dividen en dos subclases: directos e indirectos. Los métodos directos buscan un óptimo local, por medio de la resolución de ecuaciones no lineales mediante la búsqueda del gradiente hacia la función objetivo igualada a cero. Siendo éstos una generalización de la notación de cálculo elemental de puntos extremos, como se muestra en la Figura 2-2. Dada una función suave y sin restricciones, puede ser posible encontrar un punto de inflexión mediante la búsqueda por restricciones, es decir, que los métodos basados en búsqueda directa, se dirigen a un óptimo local mediante una función que se vaya moviendo en relación a un gradiente local.

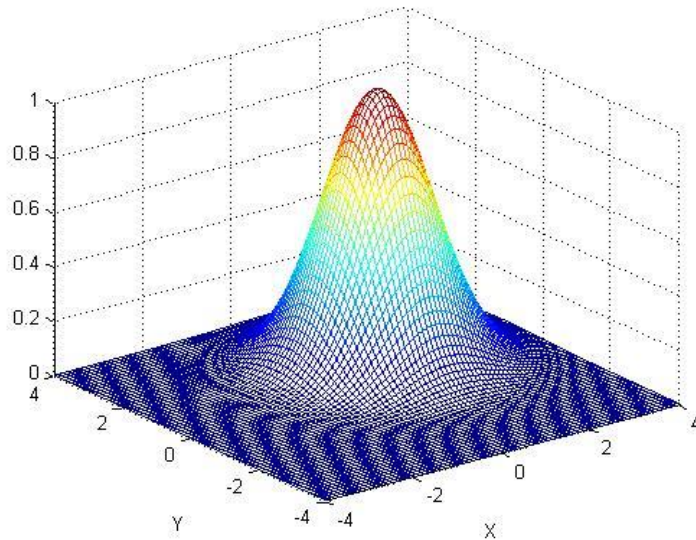


Figura 2-2. Función pico-simple.

Encontrar un óptimo en una función como la que se muestra en la Figura 2-2, es sencillo, pues se sabe que sólo se tiene un punto en el cual la derivada es cero.

Sin embargo, también existen casos de funciones con dos o más máximos (Figura 2-3), en las cuales, el campo inicial de búsqueda puede garantizar o no, la convergencia a un resultado óptimo.

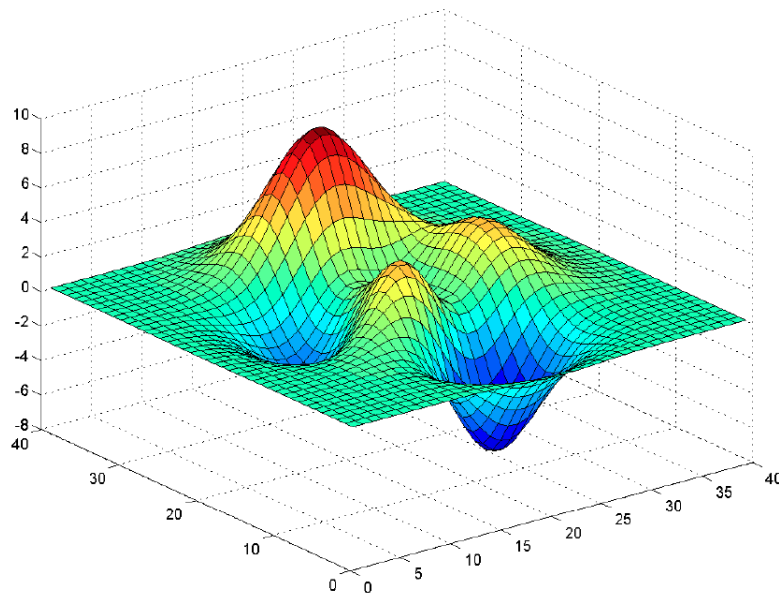


Figura 2-3. Función con múltiples picos

En la búsqueda de puntos óptimos en una función y un dominio similares a la Figura 2-3, se pueden encontrar óptimos ya sea locales o globales.

Este tipo de métodos tienden a buscar el óptimo dentro de un vecindario de puntos cercanos al punto deseado; es decir, con restricciones en el dominio para iniciar la búsqueda, por lo que se puede concluir que la eficacia del método basado en derivadas depende del campo inicial de búsqueda. Como segunda instancia, este método depende de la existencia de derivadas (valores de pendientes bien definidas).

Dichas técnicas de optimización se enfocan a maximizar o minimizar una función dada, la cual está sujeta a un conjunto de restricciones. Existen diversas técnicas de optimización, basadas en búsquedas directas, en gradientes y en aproximaciones polinomiales. Es necesario buscar cuál técnica se adapta al problema en cuestión. Inclusive, si se permiten aproximaciones numéricas de las derivadas, esta representa una grave deficiencia. En el mundo real, las funciones que se desean optimizar, por lo general presentan discontinuidades o un vasto campo de búsqueda. Por ejemplo, en un registro de datos de precipitación (Ver Figura 2-4), se puede presentar ruido, que en la hidrología representa, falta de información o malas lecturas de la misma.

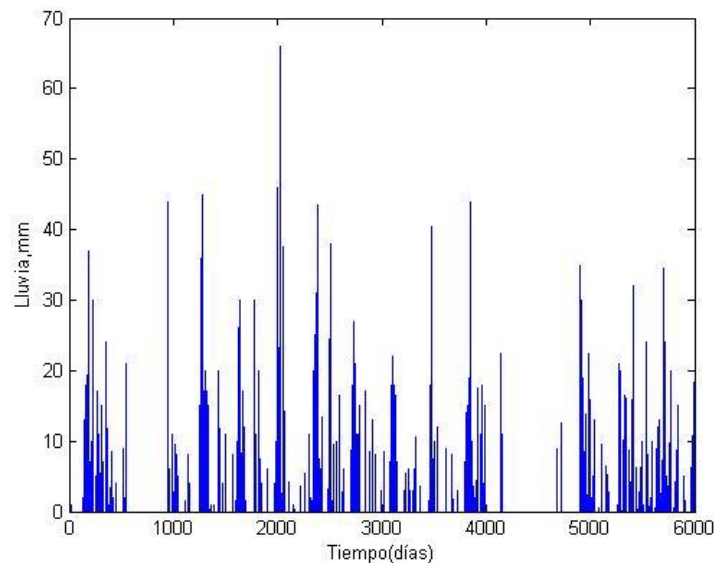


Figura 2-4. Ejemplo de función de análisis real. Datos de Lluvia en Tabasco.

Por tal motivo, diversos métodos dependen de los requerimientos en el uso de las restricciones tales como continuidad y la existencia de funciones que se puedan derivar, lo que representa muchas limitaciones respecto a problemas de dominio. Los algoritmos de optimización estocásticos han tenido buenos resultados para la búsqueda con dominio discontinuo, siendo los Algoritmos genéticos (AG) un ejemplo de búsqueda estocástica.

Los esquemas mencionados, en conjunto con la implementación de modelos híbridos y permutaciones han sido ampliamente empleados para optimizar diversos tipos de funciones, en la Figura 2-5, se puede observar que un método de optimización robusto nos garantiza llegar al óptimo global, a diferencia de un método tradicional, Goldberg (1999).

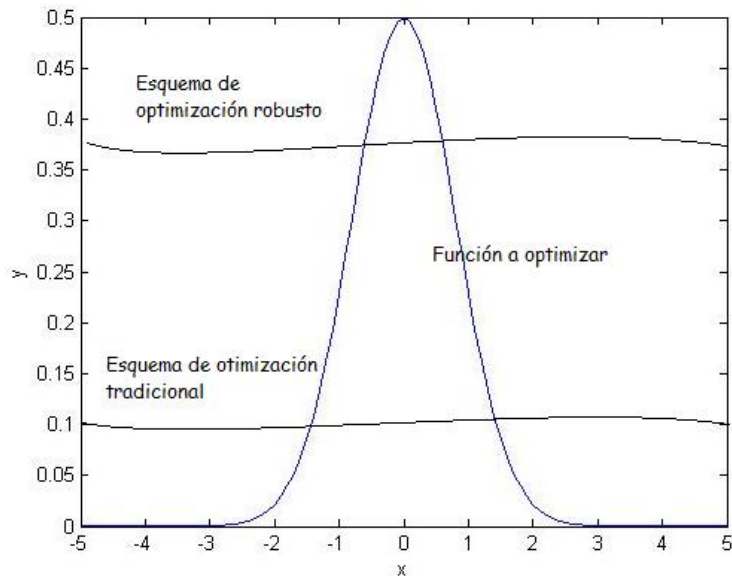


Figura 2-5. Diversos métodos de optimización y su problemática de dominio en la búsqueda de óptimos globales.

Los métodos de optimización heurísticos, han demostrado ser técnicas muy competitivas en problemas de alta complejidad, menciona Menchaca (2008).

2.5.1. Algoritmos genéticos

Se entiende por algoritmo genético (AG) como el conjunto de algoritmos ordenados de tal manera que el proceso en conjunto sea una búsqueda estocástica para la optimización de parámetros de una función cualquiera, dada. Dicho algoritmo está basado en mecanismos de selección natural; es decir, combinan la supervivencia de los mejores individuos dentro de un conjunto de ellos, los más aptos, con un intercambio de información estructurado y aleatorio, de tal forma que se imitan los procesos de evolución biológica.

Los algoritmos genéticos, se basan en los mecanismos que han permitido a los individuos adaptarse a su entorno, tales como la recombinación de su material genético (ADN) y las mutaciones que en ellos se lleven a cabo ya sea que estas sean benéficas o perjudiciales.

Tales mecanismos dotan a los individuos de nuevas características fisiológicas o en cuanto fenotipo; mismas, que son puestas a prueba mediante la selección natural, Darwin (1859).

La naturaleza provee de la variabilidad de las especies, entendiéndose como “una modificación debida directamente a las condiciones físicas de la vida” Darwin (1859).

En cuanto a la selección natural, ésta se puede “mostrar cómo la lucha por la existencia está relacionada con la selección natural”, derivado de variabilidad natural de las especies, incluyendo también la “dependencia de un ser respecto de otro, incluyendo no sólo la vida del individuo, sino también el éxito al dejar descendencia” Darwin (1859), promoviendo así, que las mejores características pasen a las siguientes generaciones, IIMAS-UNAM (2005).

Entonces la evolución consiste en que los procesos de recombinación y mutación se repiten a lo largo de varias generaciones, dando como resultado la adaptación de una especie a su ambiente.

La programación evolutiva se inspira en los mecanismos descritos aquí brevemente con la finalidad de realizar una búsqueda de las mejores soluciones aplicados a problemas de búsqueda y optimización, tal y como se describe en la Figura 2-6.

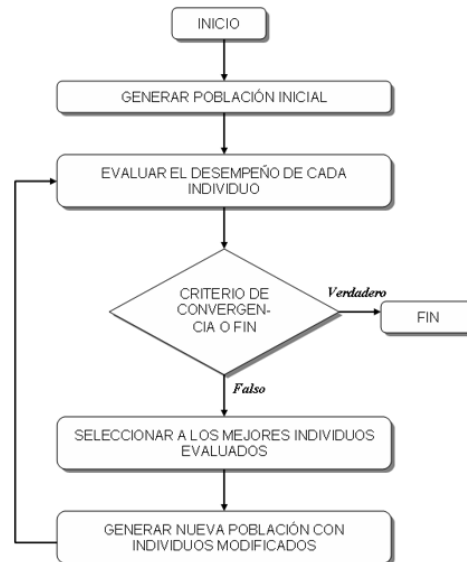


Figura 2-6. Representación del algoritmo común para las técnicas de computación evolutiva, IIMAS-UNAM (2005).

Los algoritmos genéticos fueron desarrollados por John Holland, colaboradores y estudiantes en la Universidad de Michigan. El principal objetivo de los algoritmos genéticos es el de robustecer el balance entre eficiencia y eficacia dentro de los métodos de optimización, tratando de imitar a la naturaleza en cuanto al tema de la evolución, ya que en este campo ha demostrado la supervivencia de la especie en cuanto a la mutación de éstas.

La principal diferenciación que tiene ese método con los tradicionales, es la robustez que se ha encontrado. Los algoritmos genéticos se diferencian de los métodos de optimización tradicionales y procedimientos de búsqueda en cuatro puntos:

1. Los AGs funcionan con un “set” de algoritmos codificados, mas no en algoritmos en sí mismos.
2. El espacio de búsqueda de los AGs es una población de puntos, no de un un punto en específico.
3. Los AGs usan una matriz de información (función objetivo), mas no en derivadas u otro conocimiento general.
4. AGs emplean reglas de probabilidad dinámica, mas no reglas determinísticas.

Por ejemplo al buscar el máximo de la función $f(x) = x^2$, (ver Figura 2-7) con AG el primer paso en el proceso de la optimización es de codificar a la variable como una cadena finita de caracteres o

bits, a este paso se le puede llamar como la codificación del individuo , al cual se le llamará “ genotipo”, mismo que puede ser codificado y decodificado. A diferencia de otros métodos, que encuentran un óptimo mediante procesos iterativos.

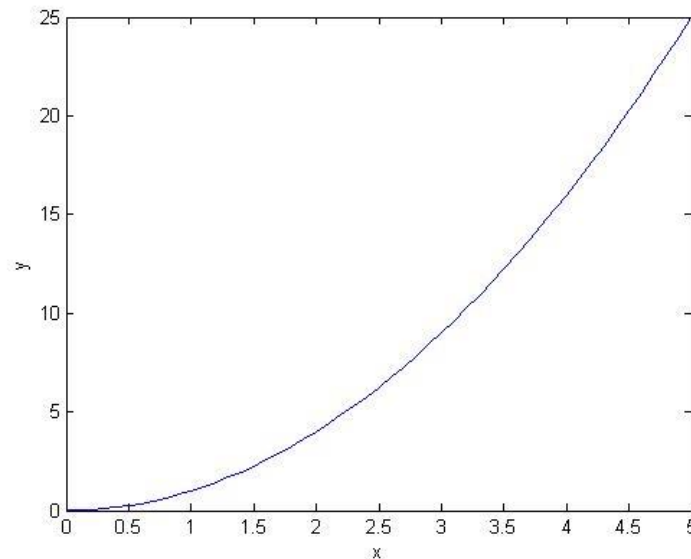


Figura 2-7. Ejemplo de una función simple para su optimización, la función $f(x) = x^2$ con un dominio entre 0 y 5.

Un algoritmo genético de código simple (AGs), puede generar una cadena de códigos con ceros y unos, y puede ser, por ejemplo, constituida por cinco elementos.

Por ejemplo, nuestra cadena para la codificación de ajustes, puede ser: 11110, donde los primeros cuatro bits son de “prendido” y el último “apagado”, de esta manera los AGs pueden explorar varios códigos similares de muchas maneras diferentes; como resultado, en gran parte no se tienen las grandes limitaciones que otros métodos requieren, por ejemplo de continuidad, existencia de derivadas, unimodalidad, entre otros.

En diversos métodos de optimización movemos un punto en el espacio de decisión a otro empleando alguna regla de transición. Esta forma de moverse de un punto a otro es peligrosa, ya que desde el inicio el camino de búsqueda puede ser erróneo, desembocando así a un óptimo falso en el espacio de búsqueda. Por el contrario, los AGs, funcionan con una población inicial de individuos (cadenas de bits) que acotan el espacio de búsqueda, reduciendo con esto la probabilidad de llegar a un óptimo local, debido a que la búsqueda de puntos de inflexión se hace en paralelo, tomando en cuenta el espacio global.

Un algoritmo genético empieza en una población de cadenas y después de eso, genera poblaciones sucesivas de cadenas. Por ejemplo para el problema de cinco bits como cadena inicial, de manera aleatoria y para un AG simple estándar (de 5 bits), la generación para cuatro generaciones ($n = 4$):

01101
11000
01000
10011

Las generación de poblaciones, representan la variabilidad genética que se tiene en la naturaleza. Para garantizar la eficiencia del método de búsqueda de la optimización de funciones, un AG requiere únicamente de una función objetivo para valuar las cadenas de valores (individuos) que representen los parámetros a optimizar.

Se asigna un determinado número de bits a cada parámetro y se realiza una discretización de la variable representada por cada gen. El número de bits asignados a cada parámetro dependerá del grado de ajuste que se desee alcanzar. No todos los parámetros tienen que estar codificados con el mismo número de bits. La Figura 2-8 muestra un ejemplo de un individuo binario que codifica tres parámetros.

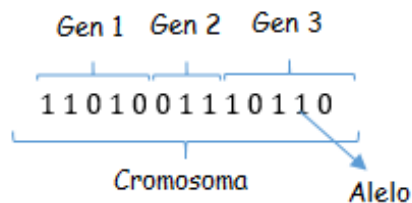


Figura 2-8. Individuo genético binario.

A diferencia de muchos métodos, un AGs utiliza las reglas de la probabilidad de la búsqueda, pero no se debe confundir con que es la toma de decisiones lanzada por una moneda. Los algoritmos genéticos emplean las decisiones aleatorias para buscar el espacio de soluciones con menor probabilidad de error.

2.5.1.1. Algoritmo genético simple

El mecanismo del AGs es asombrosamente sencillo, ya que prácticamente consiste en la copia de cadenas de información e intercambio parcial de las mismas. La sencillez de las operaciones y el efecto potente que tiene, son dos de los principales enfoques del método.

Recordando la población inicial a manera de ejemplo que se detalló anteriormente:

01101
 11000
 01000
 10011

Cabe recordar, que la población inicial fue elegida aleatoriamente. Y ahora se definirán las operaciones necesarias para iniciar con esta población y generar igualmente, poblaciones sucesivas.

Un algoritmo genético simple ofrece buenos resultados en prácticamente muchos problemas y éste se compone de tres operaciones, Goldeberg (1999).

1. Reproducción
2. Cruzamiento
3. Mutación

Antes de la reproducción, el individuo o cadena de caracteres debe codificarse, como se mencionó anteriormente, en cadenas de bits, en el cual, cada bit, recibe el nombre de alelo y a la cadena completa se llama cromosoma, misma que se conforma por conjunto de subcadenas o genes.

2.5.1.1.1. Reproducción

La reproducción es el procedimiento en el cual, la cadena de individuos son copiados con su función objetivo, se conoce como función de ajuste (f) y es una medida para evaluar los parámetros buscados, ya sea que éstos requieran ser maximizados o minimizados. De acuerdo a esta copia de cadenas en f quiere decir que la cadena con mayor valor tiene una mayor probabilidad de contar con mayores descendientes a la siguiente generación. Esta operación es una versión artificial de la selección natural, por ejemplo, esta función es determinada por las habilidades de cada especie para poder sobrevivir a los predadores, enfermedades o cualquier contingencia natural y de esta manera asegurar la supervivencia de la especie. Dicho de otra manera, dichas habilidad son dotadas a los individuos de manera aleatoria.

El operador de reproducción es implementado en los AG de diversas formas. Quizá la forma sencilla de implementarlo es mediante el uso de una rueda de ruleta, propuesta por DeJong, donde cada afluyente de las cadenas de individuos (ver Figura 2-9), tienen una parte de la rueda de la ruleta de acuerdo a su aptitud relativa (f_{rn}), la cual se traduce como la probabilidad de ser seleccionado y de supervivencia; mientras que la aptitud acumulada (f_{an}) es el rango de selección de cada individuos, el cual se obtiene de sumar sucesivamente la aptitud relativa.

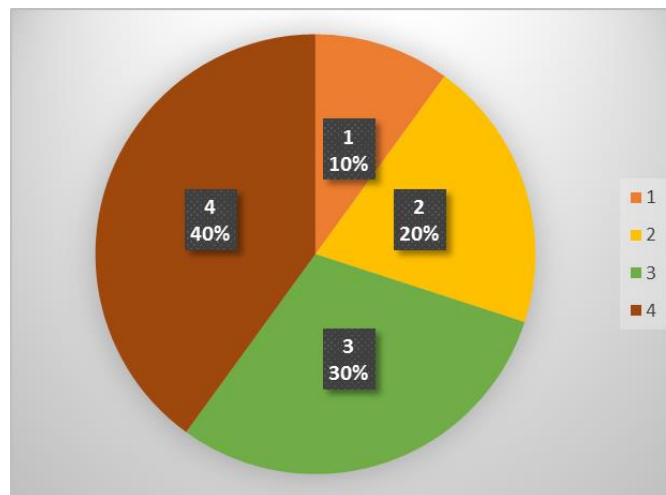


Figura 2-9. Selección simple, por medio de ruleta. Generación para cuatro individuos, con números aleatorios entre 0 y 1 para crear la población de selección.

La correspondiente ponderación para cada individuo se muestra en la Tabla 2-1. La ponderación de cada peso para cada individuo, toma una parte del total de la ruleta y ésta se hace girando a la misma cuatro veces, para este caso. Por ejemplo, para el individuo número uno, al girar la ruleta, el valor de la función de ajuste tomó un valor de 169, el cual corresponde al 14.4 % del total: es decir, que el individuo número uno tiene una probabilidad del 0.144 de reproducirse. De esta manera, cada individuo tiene una probabilidad de reproducción en función de la función de ajuste.

Tabla 2-1. Problema simple de evaluación con la función de ajuste de individuos.

No	Cadena	Fitness	% de total
1	1101	169	14.4
2	11000	576	30.9
3	1000	64	5.5
4	10011	361	49.2
Total		1170	100

Además de la reproducción por ruleta existen otros operadores, Arroyo (2003).

- ✓ Selección por torneo
- ✓ Selección basada en ranking
- ✓ Selección sexual

2.5.1.1.2. Cruza

Después de llevar a cabo el operador genético de la reproducción, se realiza un cruzamiento simple (ver Figura 2-10), el cual se lleva a cabo en dos pasos. El primer paso consiste en que los individuos recién generados son apareados aleatoriamente, mientras que el segundo paso, consiste, en que cada pareja de cadenas codificadas sufre un cruzamiento de la siguiente manera: en una determinada posición k a lo largo de la cadena es seleccionado uniformemente de manera aleatoria entre el alelo 1 y el largo de la cadena *menos uno* $[1, l - 1]$. Dos nuevas cadenas fueron creadas por el intercambio de todos los caracteres entre la posición $k + 1$ y la última posición l , incluida. Por ejemplo, se considera la cadena A_1 y A_2 , del ejemplo de la población inicial:

$$A_1 = 0\ 1\ 1\ 0\ | 1$$

$$A_2 = 1\ 1\ 0\ 0\ | 0$$

Y suponga que la elección aleatoria entre el 1 y el número 4 es de $k = 4$, al cual se le conoce como probabilidad de cruza P_c , tal y como se indica en los ejemplos anteriores con el símbolo $x |$ de separación. El resultado de la operación de cruzamiento son dos nuevas cadenas de bits donde el símbolo "prima (´)" indica que corresponde a la cadena como parte de la nueva generación, según Golberg (1999):

$$A_1' = 0\ 1\ 1\ 0\ | 0$$

$$A_2' = 1\ 1\ 0\ 0\ | 1$$

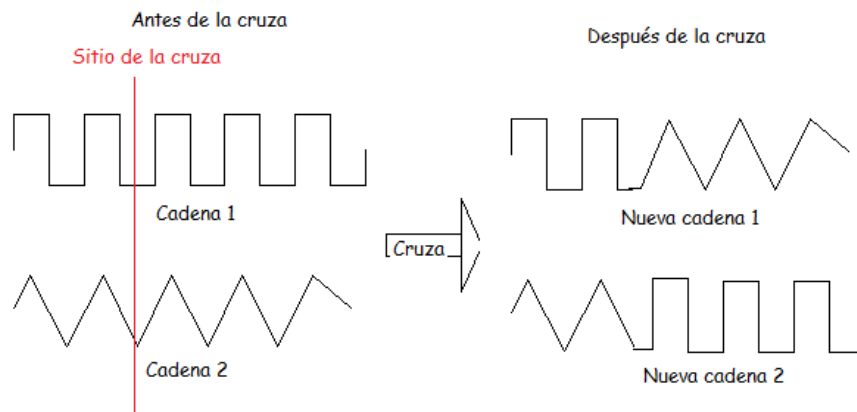


Figura 2-10. Esquema de la operación genética de cruce parcial para dos cadenas, empleando la probabilidad de Cruzamiento P_c .

Al proceso de cruce también se le conoce como recombinación y también se puede definir como el intercambio de información genética entre una pareja de individuos. Por ejemplo si al generar un número aleatorio, este es menor o igual a la probabilidad de cruce, el individuo es seleccionado para cruzarse.

A diferencia de la Figura 2-10, la cruce también puede ser en un punto, el punto de cruce se selecciona de igual manera que el descrito anteriormente por IMMAS-UNAM (2005), considerando para este caso, el cambio de 1 por 0 sólo en el sitio de cruce.

La idea básica de la cruce es que, el individuo *hijo* contenga las bondades de los padres, la cual se espera que sea mayor. Existen multitud de operadores de cruce, los más empleados se enlistan a continuación, Arroyo (2003).

- ✓ Cruce en un punto
- ✓ Cruce en dos puntos
- ✓ Cruce uniforme

2.5.1.1.3. Mutación

El objetivo de este operador genético es garantizar mejores resultados en el algoritmo genético a través de la frecuencia de cruce. También es considerado, de la misma forma que en la naturaleza, como el segundo mecanismo de adaptación menciona Goldberg (1999).

El proceso de mutación es en el cual un alelo de un gen es aleatoriamente reemplazado, generándose una nueva estructura. Si al generar un número aleatorio, este es menor o igual que la probabilidad de mutación (P_m), el bit de la cadena es mutado. La mutación en el caso de un algoritmo genético simple se hará invirtiendo un valor de la cadena de 0 a 1 ó viceversa, IMMAS-UNAM (2005).

Aunque se pueden seleccionar los individuos directamente de la población actual y mutarlos después antes de introducirlos en la nueva población, la mutación se suele utilizar de manera conjunta con el operador de cruce. Generalmente la probabilidad de mutación es menor al 1%,

debido a que los individuos suelen tener una evaluación de fitness menor después de mutados; es decir, que la mutación se realiza con el objetivo de garantizar que ningún alelo (o bit) dentro del espacio de búsqueda no tenga probabilidad de ser examinado. Los operadores de mutación más empleados se enlistan a continuación (Arroyo Apaza, 2003):

- ✓ Reemplazo aleatorio
- ✓ Mutación para representación de orden

2.5.1.2. En síntesis

El mecanismo general de un AGs, que, durante la ejecución del mismo, los padres deben ser seleccionados para la reproducción, los cuales, posteriormente, se cruzarán generando dos hijos, sobre cada uno de los cuales actuará un operador de mutación. El resultado de la combinación de las anteriores funciones será un conjunto de individuos, los cuales en la evolución del AG formarán parte de la siguiente población. En la Figura 2-11, basada en (Arroyo Apaza, 2003), se muestra el modelo básico del AG.

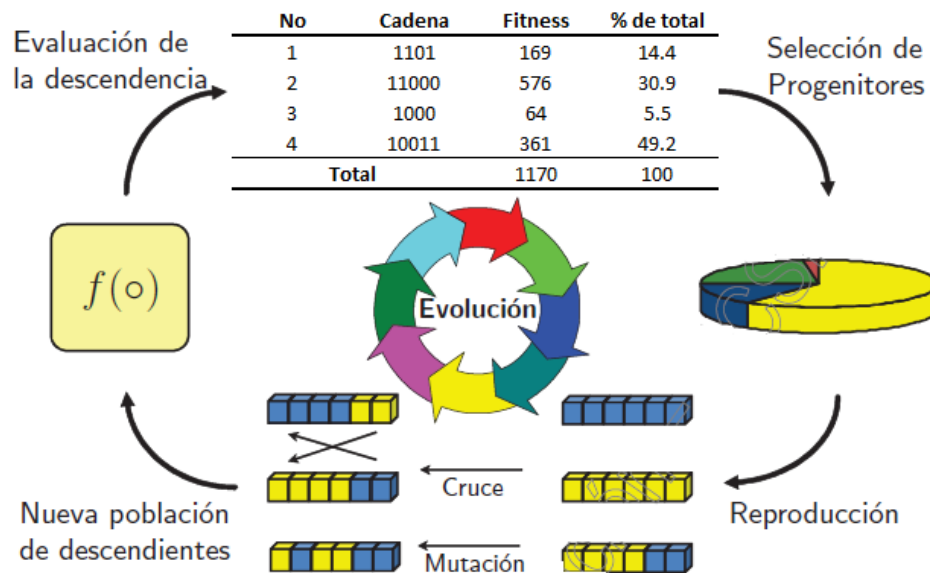


Figura 2-11. Esquema básico del funcionamiento de un AG simple.

2.5.2. Nelder-Mead

La idea es introducida por Spendley *et al.* (1962), como un seguimiento de operaciones de optimización mediante la evaluación de un conjunto de puntos que forman el simplex en un espacio n -dimensional, y continuamente forma nuevos puntos mediante la reflexión de un punto de un punto en el hiperplano de los puntos restantes según Nelder y Mead (1965).

Se le llama simplex a la figura geométrica formada por n vértices; es decir, por $n+1$ puntos en un espacio n -dimensional, donde cada uno de estos vértices son el número de parámetros a optimizar más uno. Por ejemplo, si la función objetivo tiene dos parámetros a optimizar, entonces se tiene un espacio de dos dimensiones y se dice que el simplex es un triángulo; mientras que un espacio de

tres dimensiones el simplex es un tetraedro, y así, sucesivamente.

La idea básica del método es comparar el valor de la función objetivo con los $n+1$ vértices de un simplex general, para, posteriormente mover el simplex gradualmente. Esto se logra aplicando los operadores de reflexión, expansión o contracción (ya sea interna o externa) y encogimiento hacia el punto óptimo de manera iterativa. Lo anterior se logra debido a que el algoritmo emplea la propiedad de que puede generarse un nuevo simplex sobre cualquier cara del simplex previo, proyectando cualquier vértice elegido a una distancia apropiada, mediante el centroide de los vértices restantes del viejo simplex X_c , tomando todos los vértices, excepto el peor (P) de ellos, a partir de reflejar los vértices (X_r), y con la finalidad de obtener un vértice nuevo X_n , (ver Figura 2-12).

Para encontrar el centroide (X_c) del simplex es necesario evaluar cada vértice en la función con el objetivo de encontrar el mejor vértice (M), el segundo mejor (SM), y así sucesivamente hasta el peor de ellos (P).

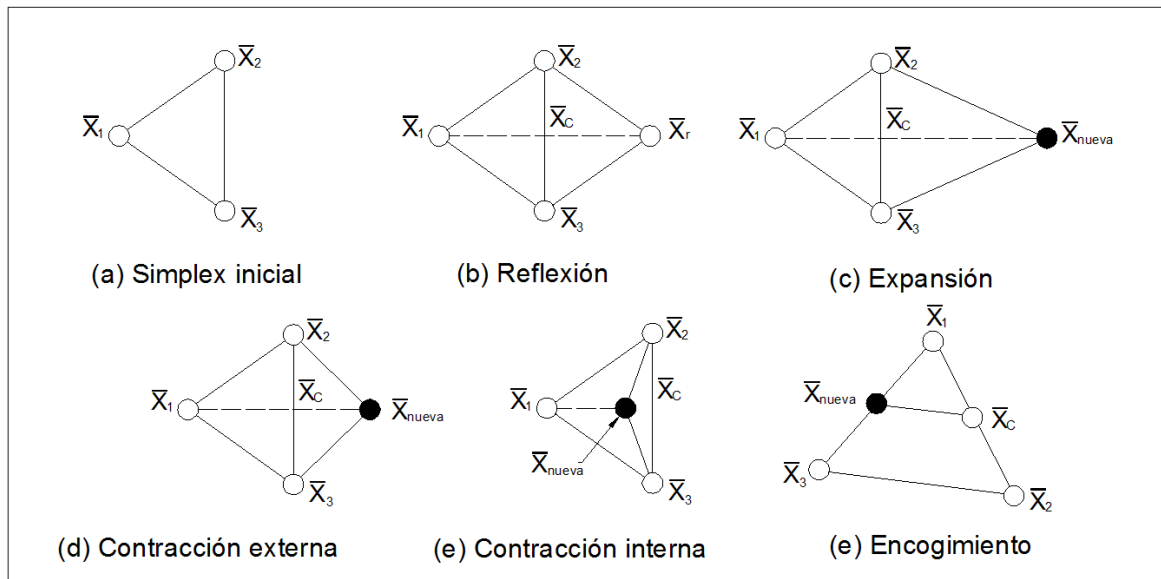


Figura 2-12. Nelder-Mead, donde X_n , es el nuevo vértice generado, Menchaca (2008) y Herrera et al., (2012).

Una vez encontrado el centro geométrico del simplex, la primera operación será la de reflexión.

Para evaluar las operaciones que conforman el método sólo es necesario elegir el mejor vértice (M), un vértice intermedio denominado “ B ” (bueno) y el peor vértice del simplex (P).

2.5.2.1. Reflexión

Una vez evaluados los puntos en la función objetivo, éstos se pueden ordenar del mejor vértice al peor y entonces se puede decir que la función decrece del mejor al peor vértice; por lo tanto, el objetivo principal de la reflexión será el de tratar de obtener un vértice con un valor más pequeño en la función objetivo, ya que se esperan mejores resultados al proyectar el vértice peor (P) a lo largo del segmento del mejor vértice al vértice segundo mejor o también denominado Bueno (B).

La operación de reflexión para encontrar el nuevo vértice (X_r) es la Ecuación 2.5-1, Herrera, Arganis y Domínguez (2014).

$$X_r = X_C + \omega(X_C - P) \quad 2.5-1$$

En donde:

ω = coeficiente de reflexión

X_C = Centro geométrico MB

El centro geométrico, el cual considera todos los vértices del simplex, excepto el peor de los vértices pertenecientes al simplex generado, puede calcularse según la Ecuación 2.5-2.

$$X_C = \frac{1}{n} \sum_{i=1, i \neq P}^{n+1} X_i \quad 2.5-2$$

2.5.2.2. Expansión

Si dentro del proceso observamos que el nuevo punto, correspondiente a la reflexión es mejor que el peor del simplex general; entonces, la función objetivo decrecerá si nos movemos del centroide al vértice reflejado mediante la Ecuación 2.5-3 Herrera, Arganis y Domínguez (2014).

$$E = X_C + \beta(X_R - X_C) \quad 2.5-3$$

En donde:

β = coeficiente de expansión

X_C = Centro geométrico MB

2.5.2.3. Contracción

Hay dos tipos de contracciones, una llamada interna, (ver Ecuación 2-48), Herrera *et al.*, (2012) y se aplica en caso que dentro del proceso, la función de la reflexión resulte mayor a la función del peor vértice (P), por lo que se considera que condujo al simplex a una mala región, como se menciona en Herrera, Arganis y Domínguez (2014).

$$T = X_C \mp \gamma(X_r - X_C) \quad 2.5-4$$

Donde

γ = coeficiente de contracción y $0 \leq \beta \leq 1$ y se define según la Ecuación 2.5-5

$$\gamma = \frac{\text{distancia entre } X_n \text{ y } X_C}{\text{distancia entre } P \text{ y } X_C} \quad 2.5-5$$

Por otro lado, la contracción interna (ver Ecuación 2.5-6) se aplica en el caso en el cual el proceso haya conducido a la condición en la que la evaluación del punto reflejado sea mayor al punto medio elegido y a su vez, que la función del vértice reflejado sea menor que la función del peor vértice.

$$T = X_C - \gamma(X_r - X_C)$$

2.5-6

2.5.2.4. Encogimiento

El encogimiento es la operación que se realiza al final del proceso, con la finalidad de acercar cada vértice al mejor de ellos. La expresión para el encogimiento (X_i) está dado por la Ecuación 2.5-7 Herrera, Arganis y Domínguez (2014).

$$X_i = \delta(X_1 + X_i)$$

2.5-7

Donde $i = 2, 3, \dots, n + 1$

El encogimiento se realiza en dirección del mejor vértice y se aplica si el valor de la función en la contracción no es menor que el valor en el peor de ellos.

El valor de los coeficientes ω, β, γ y δ que son usados típicamente son de 1, 2, 0.5 y 0.5 respectivamente, menciona Pérez (2005).

2.5.3. Algoritmo Nelder-Mead

El algoritmo empleado es el que se describe en el libro de “Métodos Numéricos con Matlab” de Mathews y Fink (2003). Estos autores muestran un algoritmo que es computacionalmente eficiente. En cada paso se determina un nuevo vértice para reemplazar al vértice peor P . los detalles del proceso de decisión lógica se explica en la Tabla 2-2.

Tabla 2-2. Decisiones lógicas en el algoritmo Nelder-Mead. Mathews y Fink (2003).

Si $f(X_r) < f(SM)$, Entonces se hace el caso (i), que corresponde a la reflexión y expansión Si no, se hace el caso (ii), el cual corresponde a contraer o extender	
<p>Caso (i) Si $f(M) < f(X_r)$ Entonces Se reemplaza P con X_r Si no Se calcula expansión (e) y F_e Si $f(E) < f(O)$ Entonces Se reemplaza P por E Si no Se reemplaza P por X_r Fin de si Fin de si Fin de si (caso (i))</p>	<p>Caso (ii) Si $f(X_r) < f(P)$ Entonces Se reemplaza P con X_r Si no Se calcula contracción (X_{oc}) y $F_{X_{oc}}$ O bien contracción interna (X_{ic}) y $F_{X_{ic}}$ Si $f(X_c) < f(P)$ Entonces Se reemplaza P por C Si no Se calcula encogimiento (X_{en}) y $F_{X_{en}}$ Y se reemplaza M con X_{en} Fin de si Fin de si Fin de si (caso (ii))</p>

El código completo se puede consultar con la autora de este trabajo o en el Instituto de Ingeniería UNAM).

2.6. Método híbrido

En la actualidad existe una enorme cantidad de problemas, en todas las áreas, que requieren de su optimización, por lo que existen, también una gran gama de modelos matemáticos aplicados a problemas de optimización con restricciones, los cuales están formados por una función objetivo y un conjunto de restricciones de igualdad y desigualdad.

El manejo adecuado de las funciones objetivo y de las restricciones no es una tarea fácil de resolver debido principalmente a que las técnicas matemáticas y/o computaciones presentan limitaciones en su empleo, por ejemplo, cuando la función objetivo y/o las restricciones no son diferenciales o éstas son discontinuas o no se pueden expresar de manera algebraica o bien, cuando la región factible es disjunta.

Por tal motivo, esta tesis busca un algoritmo híbrido que pueda aproximarse a una buena solución, mediante la búsqueda de los parámetros o conjunto de parámetros que optimicen la función que describa mejor un conjunto de datos dado.

Técnicas como los algoritmo genéticos (AG) y el método geométrico Nelder-Mead (N-M) han demostrado, en los últimos años, ser una herramienta que han podido resolver funciones con diversas características, resolviendo así con un costo computacional bajo (al momento de hibridizarse), dicho problema.

El modelo híbrido consiste en iniciar la búsqueda con un algoritmo genético simple en un tiempo computacional corto en un cierto número de iteraciones y los resultados obtenidos con este método son el dominio que acota la búsqueda para iniciar con el proceso Nelder-Mead. Es importante medir los tiempos computacionales debido a que uno de los objetivos principales del método es el de reducir los tiempos de búsqueda

2.6.1. Función objetivo

La función objetivo representa las técnicas para la estimación de parámetros que se quieren optimizar para la búsqueda del mejor ajuste de la función en cuestión, según los datos que se estén procesando.

2.6.1.1. Selección de la función objetivo en los problemas abordados

Debido a que los problemas de optimización son de diversa naturaleza, la función objetivo es diferente, dependiendo la función que se requiera ajustar al conjunto de datos que se requiera.

A continuación se describen los métodos de optimización con base a la función que se desea ajustar dependiendo el caso de aplicación:

- ✓ Para la función exponencial, la función objetivo (ver Ecuación 2.6-1) será la suma de los cuadrados de la diferencia entre los rangos correspondientes al vector *de datos medidos* y los calculados con el método Nelder-Mead.

$$FO = \sum_{i=1}^y (y_i - y_0)^2$$

2.6-1

Donde:

Y_i = son las ordenadas de los datos observados

Y_i = son las ordenadas de los datos que se van generando al aplicar el método N-M

- ✓ Para el primer caso de análisis, de la función Gumbel, la función objetivo empleada fue máxima verosimilitud.
- ✓ Para el segundo caso de análisis, de la función de densidad doble Gumbel bivariada y debido a que la FDDGB consta de 11 parámetros, el método para optimizarla por mínimos cuadrados o de momentos se hace impráctico, a causa de que el número de ecuaciones no lineales necesarias para encontrar los 11 parámetros se incrementa conforme al número de momentos necesarios también se incrementan. Por tal motivo el método para optimizar fue por máxima verosimilitud.

El código de la función objetivo se realizó en Matlab (MAtrix LABoratory), mismo que se puede consultar con la autora de la presente tesis o en el Instituto de Ingeniería de la UNAM.

2.6.2. Calibración de los modelos de optimización con la función exponencial

Se utilizó una función exponencial (Ec. 2.6-2) para verificar el correcto funcionamiento de los algoritmos codificados. El primer paso fue usar el método de Nelder-Mead con el que se optimizaron los parámetros de la función exponencial. El ejercicio consiste en proponer distintos valores para la variable X (ver Tabla 2-3), se calcula primero el vector y_0 , correspondiente a la función exponencial, estimado con los parámetros de la Ec. 2.6-2 propuestos inicialmente como: $a = 1.0$ y $b = 0.5$.

$$y = ae^{bx}$$

2.6-2

Después se calculan los parámetros de ajuste de la función exponencial (Ec. 2.6-2) a los valores de X (Tabla 2-3), mediante el método Nelder-Mead y con éstos se estima de nuevo la función, a la que llamaremos y_f . Finalmente tanto los valores para la función exponencial (y_0) se compararon con los estimados (y_f) (Tabla 2-3). Los valores aleatorios propuestos para iniciar con el método Nelder-Mead se muestran en la Tabla 2-4.

Tabla 2-3. Valores propuestos para la calibración del método Nelder-Mead, valores iniciales (Y_o) y valores finales (Y_f).

X	Y_o	Y_f
1	1.65	1.65
2	2.72	2.72
3	4.48	4.48
4	7.39	7.39
5	12.18	12.18
6	20.09	20.09
7	33.12	33.12
8	54.60	54.60
9	90.02	90.02
10	148.41	148.41
Coeficiente de correlación		1.00

Tabla 2-4. Límites para valores propuestos de a y b.

Parámetros propuestos	M	B	P
a	0.50	1.80	1.10
b	0.10	0.50	1.10

El método se implementó para un total de 100 iteraciones y un tiempo computacional correspondiente a 0.091 segundos y el método convergió a los valores exactos

En las siguientes gráficas se muestran los resultados (Fig. 2-13) y los vértices optimizados (Fig. 2-14).

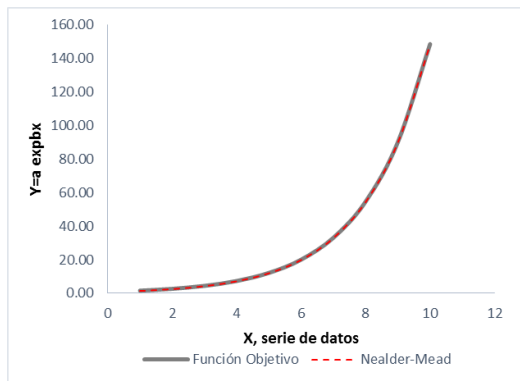


Figura 2-13. Calibración del modelo N-M. Ajuste de parámetros de función exponencial.

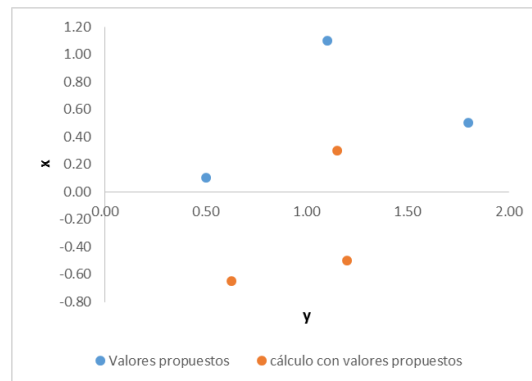


Figura 2-14. Vértices encontrados con N-M, correspondientes a la función exponencial.

2.7. Implementación del modelo híbrido

El método consiste en iniciar la búsqueda con AGs para distinto número de generaciones e individuos y al término de éste se inicia con el proceso Nelder-Mead con la finalidad de reducir tiempos computacionales y acotar el dominio inicial de búsqueda. El proceso de implementación se describe a continuación.

Primero se utiliza un algoritmo genético simple (AGs), originalmente codificado en el Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS) y con adecuaciones realizadas en

el Instituto de Ingeniería, UNAM usando el software MATLAB R2012® (Abreviatura de MATrix LABoratory), el cual es una herramienta de software matemático que ofrece un entorno de desarrollo integrado (IDE) con un lenguaje de programación propio, lenguaje M, diseñado para realizar cálculos matemáticos y de simulación.

Con el algoritmo genético se genera una población inicial cuyos valores son tomados como medios para acotar tanto el límite superior como inferior de dicha población.

Posteriormente se realiza el proceso de generación de poblaciones, mutación, selección y finalmente maximización de los parámetros de la función de distribución en cuestión. Como resultados se obtiene un matriz de dimensiones que el simplex del Nelder-Mead requiere, a la cual se le llamó "M".

La matriz de valores iniciales "M" tiene un tamaño de $n \times m$, donde n corresponde al número de renglones que representan el número de parámetros a optimizar y m , es el número de columnas que representan los vértices que conforman el simplex para Nelder-Mead.

El número de vértices de la matriz de parámetros iniciales "M" es igual al número de parámetros de la función de distribución más uno y cada uno de estos vértices fue generado con AGs para un número determinado de individuos y de generaciones.

Este paso tiene la importancia de acotar el dominio de búsqueda para el siguiente proceso del método híbrido (Nelder-Mead).

Con la matriz de datos iniciales "M", se inicia el proceso de búsqueda con el método Nelder-Mead.

Se midió el tiempo en el cual se generó el simplex inicial para el Nelder-Mead y se le sumó el tiempo para el algoritmo Nelder-Mead necesario para que el modelo en conjunto converja a un óptimo; es decir el mejor vértice.

Finalmente, con el mejor vértice, se puede ajustar la función de distribución deseada y con esto la estimación de eventos para un período de retorno dado. El proceso híbrido se puede observar en el diagrama de bloques de la Figura 2-15.

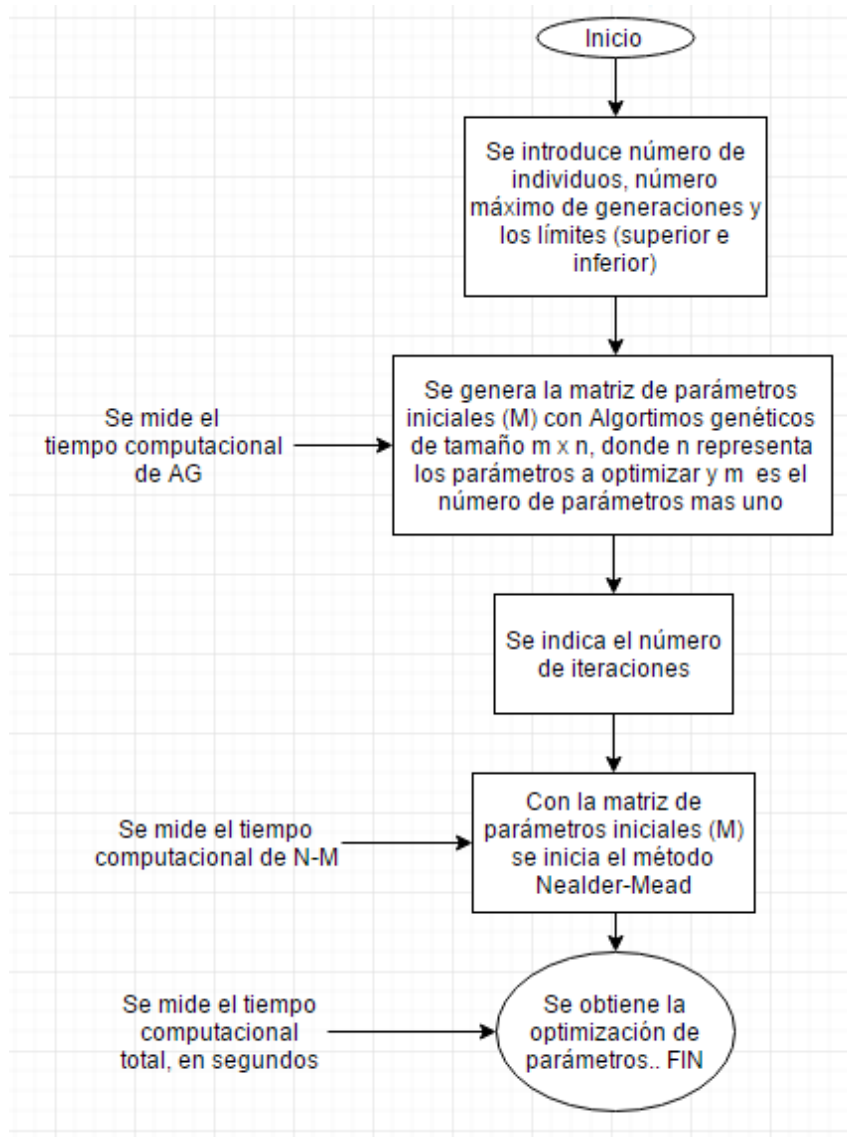


Figura 2-15. Proceso del método híbrido.

3. Aplicaciones y resultados

3.1. Caso de aplicación I. Gumbel

3.1.1. Descripción del sitio de estudio

El área de estudio corresponde al aeropuerto de la Ciudad de México en el Estado de México, para la estación número 15008 “Atenco”, dentro de la región hidrológica del Pánuco, (ver Figura 3-1).

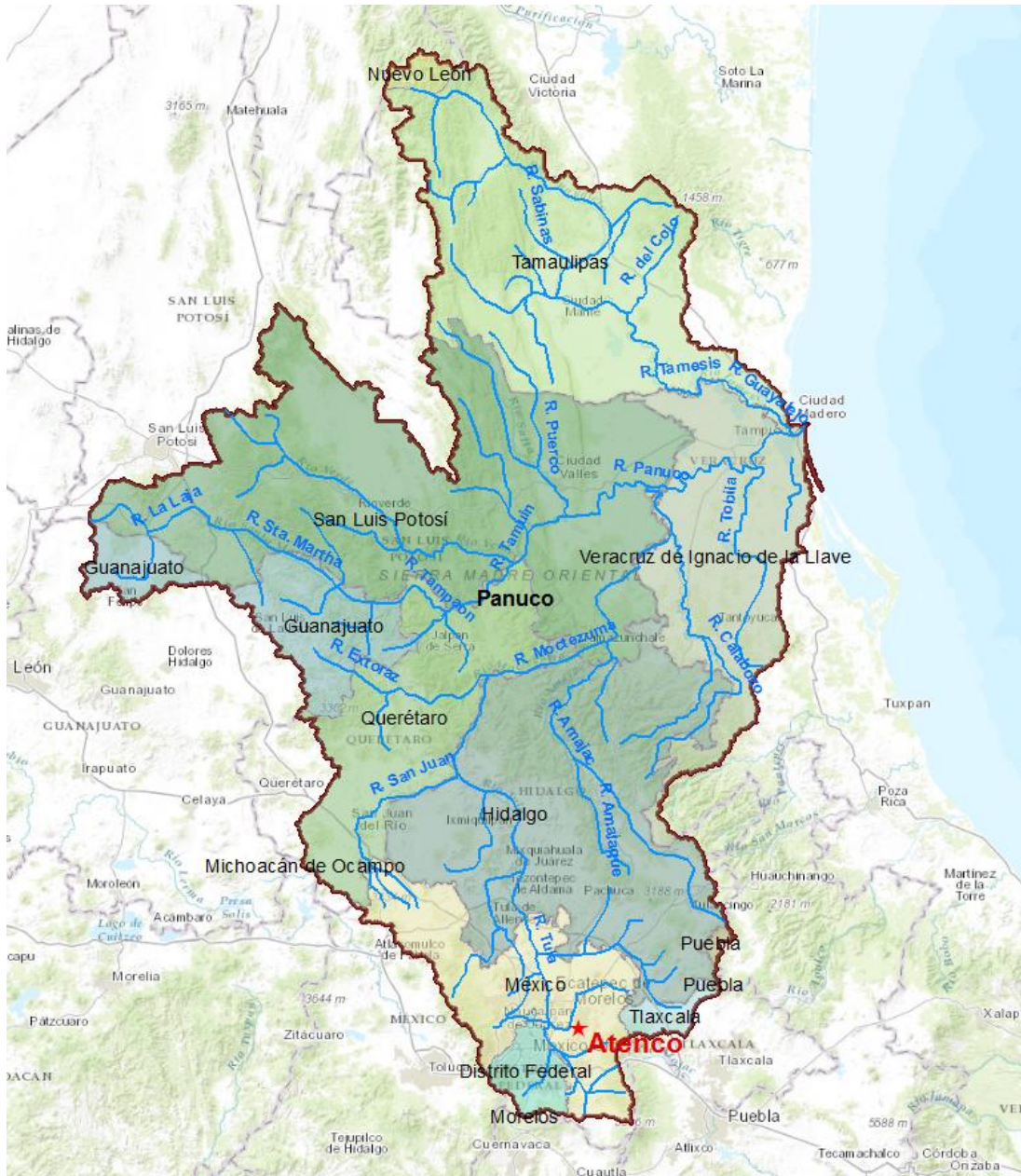


Figura 3-1. Localización de la estación número 15008 de la estación “Atenco”, Estado de México.

3.1.2. Base de Datos

Los datos recabados corresponden a la precipitación medida en la estación número 15008 de la estación “Atenco”, en el Estado de México. Considerando un registro de precipitación con una longitud de 576 datos de precipitación diarios en milímetros y para un período comprendido entre los años 1961 y 2008. En la Tabla 3-1 se observan los datos de precipitación máximos anuales.

Tabla 3-1. Precipitación máxima anual para el análisis de ajuste de función de distribución Gumbel.

N	Año	Precipitación máx anual, mm	N	Año	Precipitación máx anual, mm
1	1961	41.1	25	1985	42.6
2	1962	34.2	26	1986	51.3
3	1963	49.3	27	1987	39.1
4	1964	73.2	28	1988	75.5
5	1965	38.3	29	1989	37.5
6	1966	35.5	30	1990	30.8
7	1967	51.5	31	1991	50
8	1968	38.8	32	1992	43.9
9	1969	38	33	1993	62
10	1970	34.5	34	1994	38.3
11	1971	33.5	35	1995	60
12	1972	35.8	36	1996	43
13	1973	33.8	37	1997	25.8
14	1974	25.1	38	1998	60
15	1975	40.2	39	1999	58.8
16	1976	37.7	40	2000	30.7
17	1977	30.8	41	2001	40.2
18	1978	40.4	42	2002	32.5
19	1979	40.4	43	2003	26.5
20	1980	29.5	44	2004	49.8
21	1981	27.4	45	2005	45.5
22	1982	28.7	46	2006	45.5
23	1983	33.6	47	2007	85.5
24	1984	34	48	2008	26

Siendo sus datos estadísticos los presentados en la Tabla 3-2.

Tabla 3-2. Estadísticos de los datos de precipitación.

Media	14.4814
Desvest	13.3364
coef asim	1.2908
Varianza	177.5509

Se optimizaron los dos parámetros de la función de distribución Gumbel a los datos de precipitación, con la función objetivo de máxima verosimilitud. El proceso de implementación se describe a continuación.

3.1.3. Implementación del modelo híbrido

Se generó una matriz M , conformada por tres vértices, debido a que la función de distribución Gumbel tiene dos parámetros. La matriz de parámetros iniciales (M) es necesaria para saltar al segundo proceso (Nelder-Mead).

La conformación de la matriz M se realizó con AGs, corresponde al primer proceso del método híbrido. Para la generación de M , fue necesario introducir un límite superior y uno inferior a los AGs, los cuales serán necesarios para darle la longitud necesaria a la generación aleatoria inicial de los números binarios. Los límites se pueden observar en la Tabla 3-3.

Tabla 3-3. Parámetros iniciales de búsqueda al método híbrido.

Parámetro	Lím inferior	Lím superior
α	0.07610	0.108
β	10.65290	10.685

Posteriormente se llevaron a cabo varias pruebas combinando distinto número de individuos y de generaciones. La mejor combinación resultó de 200 individuos y 1500 generaciones. El tiempo computacional requerido para obtener la matriz M fue de 59.35 segundos para AGs. La matriz generada en este primer proceso del método híbrido se puede ver en la Tabla 3-4.

Tabla 3-4. Matriz de parámetros iniciales para Nelder-Mead, generada con 200 individuos y 1500 generaciones en 59.35 s.

Parámetro	No de vector		
	1	2	3
α	0.09588	0.09606	0.09722
β	10.69230	10.60230	10.76490

Una vez que se tienen los vértices necesarios, se inicia con el proceso Nelder-Mead, el cual se implementó con 70 iteraciones. Finalmente la solución a la que convergió el modelo híbrido en un tiempo computacional de 0.46 segundos fue la siguiente:

$$\alpha = 0.1058 \text{ y } \beta = 10.67288$$

El tiempo total en el cual, el método híbrido convergió al resultado fue de 59.82 segundos. El valor de la función objetivo (máxima verosimilitud) para el resultado final es de -1.962×10^3 .

3.1.4. Ajuste de la función de distribución Gumbel

Una vez que se optimizaron los parámetros tanto de ubicación como de escala de la función Gumbel, se calculó la función de densidad de distribución Gumbel.

Para comparar los resultados del ajuste de la función de distribución Gumbel fue necesario calcular las frecuencias mediante esta función y se graficaron contra las medidas.

Primero se le atribuyeron un número de orden a cada dato y con esto se calculó un frecuencia como

$$F = \frac{m}{n + 1}$$

3.1-1

Donde

m: es el número total de datos de registro.

n: es el número de orden, correspondiente a cada datos de precipitación.

Posteriormente, con los datos ya optimizados ($\alpha = 0.1058$ y $\beta = 10.67288$), se calculó la frecuencia tipo Gumbel.

A modo de comparación se realizó el ejercicio de ajustar la función de distribución Gumbel a los datos de precipitación correspondientes a la estación Atenco empleado AGs solamente. El método se empleó para 200 individuos y 5 mil generaciones, el tiempo computacional para el ajuste de la función Gumbel fue de 109.17 segundos y la mejor solución a la que se convergió fue de: $\alpha = .09637608$ y $\beta = 10.780322$.

En la Tabla 3-5 se muestran el análisis de frecuencias para los primeros datos de precipitación a manera de ejemplo.

Tabla 3-5. Análisis de Frecuencias para la distribución Gumbel con los datos optimizados por medio del método híbrido y AGs.

Fecha	x, mm	F relativa	F estimada (híbrido)	Error (Híbrido)	F estimada (AG)	Error (AG)
29/02/1972	0.2	0.001965	0.048445	0.0465	0.06252	0.060551
31/07/1984	0.2	0.003929	0.048445	0.0445	0.06252	0.058587
31/10/1984	0.2	0.005894	0.048445	0.0426	0.06252	0.056622
30/09/1998	0.2	0.007859	0.048445	0.0406	0.06252	0.054657
31/01/1974	0.3	0.009823	0.050013	0.0402	0.06420	0.054377
31/05/2003	0.3	0.011788	0.050013	0.0382	0.06420	0.052413
30/11/1977	0.4	0.013752	0.051615	0.0379	0.06591	0.052161
31/10/1978	0.6	0.015717	0.054918	0.0392	0.06943	0.053709
30/11/1978	0.6	0.017682	0.054918	0.0372	0.06943	0.051744
30/11/1994	0.7	0.019646	0.056621	0.0370	0.07123	0.051579
31/01/1973	0.8	0.021611	0.058357	0.0367	0.07305	0.051442
31/08/1979	0.8	0.023576	0.058357	0.0348	0.07305	0.049477
31/05/1995	0.8	0.025540	0.058357	0.0328	0.07305	0.047513
28/02/1966	1.0	0.027505	0.061933	0.0344	0.07679	0.049290
31/01/1968	1.0	0.029470	0.061933	0.0325	0.07679	0.047325
31/03/1968	1.0	0.031434	0.061933	0.0305	0.07679	0.045361
31/01/1972	1.0	0.033399	0.061933	0.0285	0.07679	0.043396
31/07/1979	1.0	0.035363	0.061933	0.0266	0.07679	0.041431
31/08/1985	1.0	0.037328	0.061933	0.0246	0.07679	0.039467
30/04/2004	1.0	0.039293	0.061933	0.0226	0.07679	0.037502
31/07/1982	1.1	0.041257	0.063772	0.0225	0.07871	0.037451
30/09/1988	1.1	0.043222	0.063772	0.0206	0.07871	0.035487
31/01/1970	1.2	0.045187	0.065646	0.0205	0.08065	0.035465
28/02/1971	1.2	0.047151	0.065646	0.0185	0.08065	0.033500
31/03/1977	1.2	0.049116	0.065646	0.0165	0.08065	0.031535

Se calcularon los estadísticos de los datos de la tabla anterior (Tabla 3-6).

Tabla 3-6. Estadísticos calculados para los datos de precipitación de la estación Atenco, de las frecuencias y de los errores estimados.

Fecha	x, mm	F relativa	$-\ln(-\ln(F))$	F estimada (híbrido)	Error (Híbrido)	F estimada (AG)	Error (AG)
media	16.419882	0.50000	0.572461	0.495251	0.0268	0.4876	0.0198
desvest	13.018767	0.28811	1.259081	0.306084	0.0175	0.2919	0.0180
coef asim	1.303233	0.00000	1.014873	0.037751	0.5488	0.0890	0.6978
coef de var	1.261247	1.73546	0.454665	1.618019	1.5332	1.6706	1.1001
Varianza	169.488286	0.08301	1.585285	0.093688	0.0003	0.0852	0.0003

También se graficaron las frecuencia relativas contra las estimadas mediante los parámetros optimizados por medio del método híbrido, con AGsy se comparan con una función identidad con el objetivo de ver qué tanto se aproximan a ésta. Los resultados se observan en la Fig. 3-2).

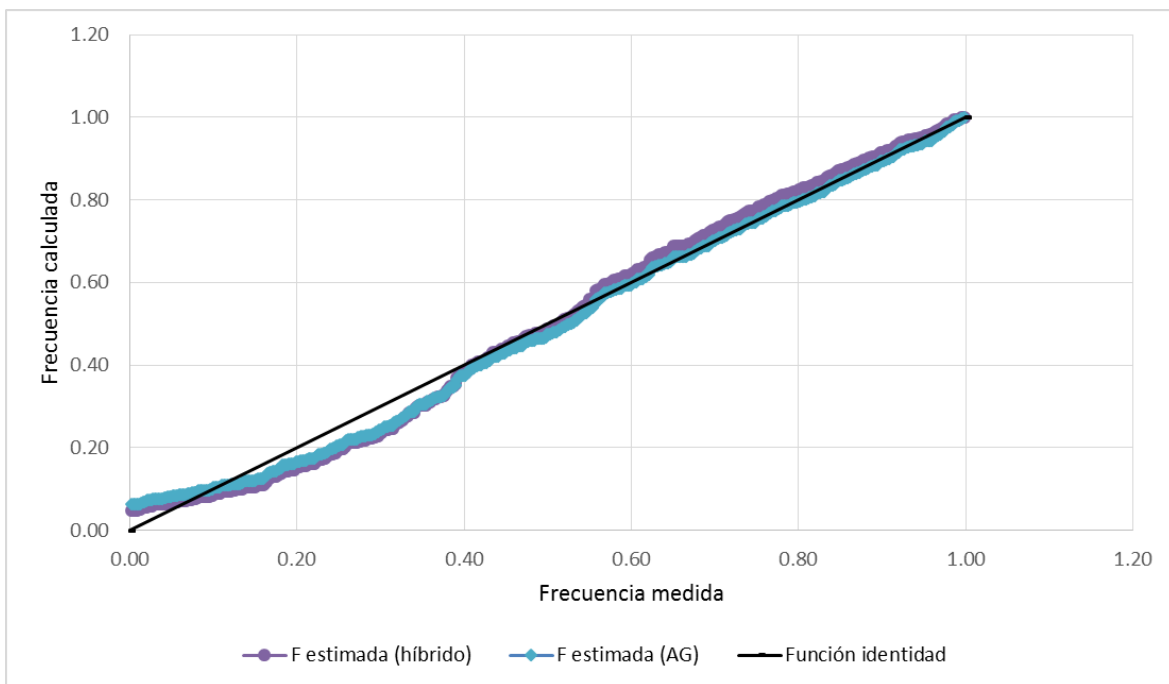


Figura 3-2. Frecuencia relativa vs Frecuencia calculada para la Función de distribución tipo Gumbel aplicada a datos de precipitación de la estación Atenco, Edo. de México.

Los datos correspondientes a la Tabla 3-6 permitieron medir el coeficiente de determinación (R^2) entre la frecuencia relativa y la estimada. Primero se calculó el coeficiente de determinación para los parámetros ajustados con el método híbrido, la cual resultó como sigue

$$R^2_{\text{híbrido}} = \frac{\sigma^2 y_{\text{medidos}} - \sigma^2 \text{error}}{\sigma^2 y_{\text{medidos}}} = \frac{0.083 - 0.0003061}{0.083}$$

$$R^2_{\text{híbrido}} = 0.9963$$

También se calculó el coeficiente de determinación de las frecuencias estimadas para los datos optimizados solamente empleando AG:

$$R^2_{AGs} = \frac{\sigma^2 y_{medidos} - \sigma^2 error}{\sigma^2 y_{medidos}} = \frac{0.083 - 0.0003245}{0.083}$$

$$R^2_{AGs} = 0.9961$$

En la Tabla 3-7 se hace una síntesis de los resultados.

Tabla 3-7. Síntesis de resultados para la optimización de los parámetros de la función Gumbel aplicada a datos de precipitación de la estación Atenco, Edo. de México.

Parámetros	Método híbrido	Método AGs
a	0.1058	0.0964
b	10.6729	10.7803
t₁, s	36.9215	109.1772
t₂, s	0.6596	-
t_t, s	37.5811	109.1772
Nindividuos₁	200.0000	100.0000
Ngeneraciones₁	1500.0000	1500.0000
Nitreciones₂	70.0000	-
R²	0.9963	0.9961

Donde t_1 representa el tiempo computacional correspondiente al primer proceso del método híbrido (AGs), t_2 representa el tiempo computacional correspondiente al segundo proceso del método híbrido (Nelder-Mead) y t_t el tiempo total del método híbrido. Es por tal motivo, que en la segunda columna, el tiempo computacional correspondiente al proceso Nelder-Mead no aplica ya que esta columna corresponde a la optimización de los parámetros empleando solamente AGs a modo de comparación.

3.2. Caso de aplicación II. Función de distribución doble gumbel bivariada aplicada a la Presa Infiernillo.

3.2.1. Descripción del sitio de estudio

El área de estudio se encuentra dentro de la región hidrológica número 18, de las 37 que conforman la República Mexicana, localizada entre los paralelos 17° 13' y 20° 04' de latitud Norte y los meridianos 97° 25' y 103° 20' de longitud Oeste. Cuenta con una superficie hidrológica de 117,405 kilómetros cuadrados, equivalente al 6% del territorio nacional, según la Secretaria del Medio Ambiente y Recursos Naturales (2007), ver Fig. 3-3.

La región hidrológica, está limitada por las Sierras Madre del Sur y la de Juárez, así como por el eje neovolcánico, tiene la forma de una depresión muy alargada con valles muy angostos, cuyo territorio está formado en su mayor parte por elevaciones con fuertes pendientes y un arreglo geológico poco propicio para el control y almacenamiento de los grandes escurrimientos que se

presentan en la región hidrológica, ya que cuenta con un potencial importante de escurrimientos consistentes en más de 900 milímetros al año.

La Región Hidrológica número 18 comprende los siguientes estados de la República Mexicana: Estado de Morelos y parcialmente a los Estados de Tlaxcala, Puebla, México, Oaxaca, Guerrero, Michoacán y Jalisco, así como muy pequeñas porciones la Ciudad de México y del Estado de Veracruz; con un total de 422 municipios y una superficie total de 117,305.9 kilómetros cuadrados, reporta la Secretaria del Medio Ambiente y Recursos Naturales (2007).

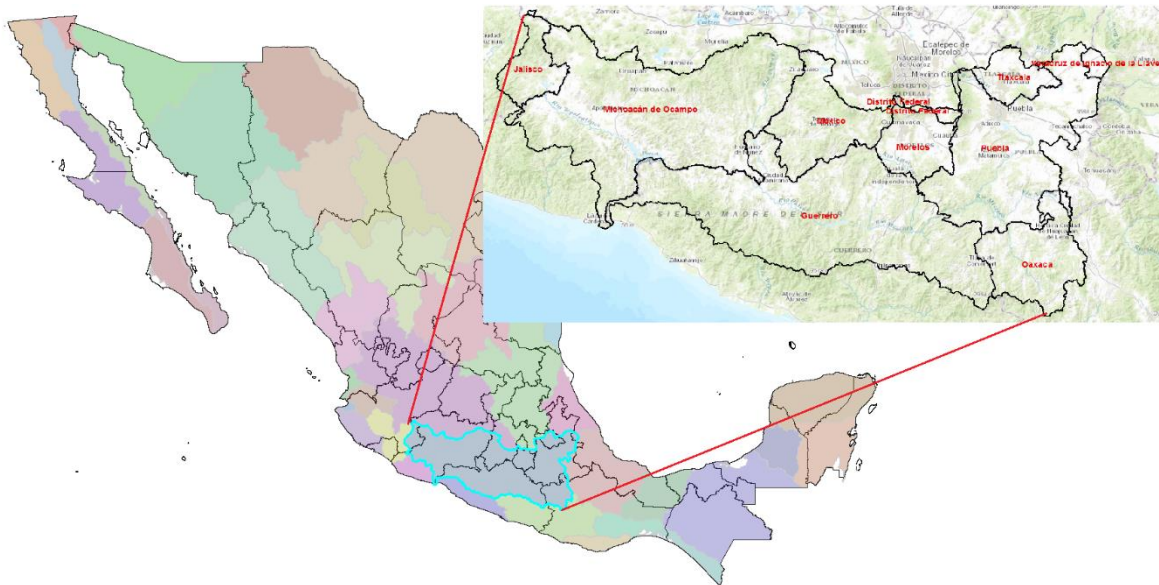


Figura 3-3. Localización de la Región Hidrológica número 18 Balsas.

3.2.2. Sistema hidrológico

La región hidrológica número 18, cuyo nombre correspondiente es “Balsas”, a su vez está conformada por 15 cuencas hidrológicas, como se puede observar en la figura anterior, las cuales se dividieron partir de las condiciones físicas que las definen y de las estructuras de control existentes, ya sean presas o estaciones hidrométricas, excepto las cuencas cerradas definidas exclusivamente por sus límites físicos. Por otro lado, para fines de planeación y manejo administrativo, la región hidrológica número 18 “Balsas”, se ha dividido en las subregiones hidrológicas alto, medio y bajo Balsas. La primera está integrada por las cuencas hidrológicas río Libres Oriental, río Alto Atoyac, río Nexapa, río Bajo Atoyac, río Mixteco, río Tlapaneco y río Amacuzac. La subregión hidrológica del medio Balsas está formada por las cuencas hidrológicas río Medio Balsas y río Cutzamala. Por su parte, las cuencas hidrológicas del río Tacámbaro, río Cupatitzio, río Zirahuén, río Paracho, río Tepalcatepec y río Bajo Balsas, están consideradas en la subregión hidrológica bajo Balsas, como se observa en la Fig. 3-4.

Las quince cuencas hidrológicas que integran la región hidrológica número 18 “Balsas” son: río Alto Atoyac, río Amacuzac, río Tlapaneco, río Nexapa, río Mixteco, río Bajo Atoyac, río Cutzamala, río Medio Balsas, río Cupatitzio, río Tacámbaro, río Tepalcatepec, río Bajo Balsas, río Paracho-Nahuatzen, río Zirahuén y río Libres Oriental, las cuencas hidrológicas río Libres Oriental, río

mil m³/s y además con riesgo a que se produzca cavitación. La importancia de esta presa radica en que alimenta al sistema central de energía del país.

Las turbinas de generación fueron instaladas en dos etapas. La primera, entró en servicio entre 1965 y 1966, tuvo una capacidad instalada de 624 MW. La segunda, entró en servicio hasta 1973, totalizando 960 MW de capacidad instalada. Para el año 2005 contaba con una capacidad efectiva de 1 000 MW, una generación bruta anual de 2 749 GWh y un factor de planta del 31.4%, según la Secretaría de energía. Las seis turbinas de la central que operaban desde 1964, fueron modernizadas en 2009 por solicitud del Gobierno Federal, lo que incrementó considerablemente su capacidad de generación a 1 120 MW siendo la segunda generadora de electricidad en el país después de Chicoasén, menciona Gómez (2015).



Figura 3-5. Presa Infiernillo, Michoacán, Gómez (2015).

3.2.4. Base de datos

Para la aplicación de la función de distribución doble Gumbel bivariada se emplearon gastos y escurrimientos correspondientes a la Presa El Infiernillo y de las estaciones hidrométricas: La Pastoria, Los Panches, Los Pinzanes y La Caimanera recabados del Banco Nacional de Datos de Aguas Superficiales (BANDAS) de CONAGUA (2017), correspondientes a estaciones hidrométricas del país.

Los datos de escurrimiento pertenecen a los registros de cuenca total, considerando los gastos medios diarios por cuenca propia reportados en la suma de las estaciones hidrométricas: Los Pinzones, Panches, La Pastoría y Caimanera, ver tabla 3-8, afectados por el factor 1.3 (obtenidos al comparar estas sumas y diferencia contra los períodos de retorno comunes de la presa El Infiernillo) en el período comprendido entre 1965 a 1994. También se utilizaron los datos del período

comprendido entre 1998 a 2013 de los gastos medios diarios por aportaciones totales registrados en la presa por CFE, Gómez (2015).

Tabla 3-8. Datos de las estaciones (BANDAS) de CONAGUA (2017).

CLAVE	NOMBRE	Lat. N, m	Long .W, m
18481	LA CAIMANERA	18.464	100.898
18487	LOS PINZANES	18.543	101.102
18494	LOS PANCHES	18.849	102.136
18495	LA PASTORIA	18.930	102.051

El cálculo de los datos de gasto pico y del volumen pico, Dueñas (2016), ocupó el método bivariado de Ramírez y Aldama (2000 y 2002), quienes proponen:

$$\tilde{Q}_p = \frac{Q_p - u_q}{\alpha_q} \quad 3.2-1$$

$$\tilde{V} = \frac{V - u_v}{\alpha_v} \quad 3.2-2$$

Donde \tilde{Q}_p y \tilde{V} son los valores normalizados del gasto pico y el volumen pico, respectivamente y u_q , α_q , u_v y α_v son los parámetros de ubicación y escala para la distribución de mejor ajuste de cada variable, Dueñas (2016).

1.7.4.1. Estimación de eventos de diseño

De acuerdo con la Ec.2.2-24, para un período de retorno conjunto, existe un número finito de pares de valores para el gasto pico y el volumen pico, Ramírez y Aldama (2000) proponen determinar cuál de esos pares produce los efectos más adversos en la obra hidráulica que se desee diseñar.

Una vez obtenidos los parámetros de la función bivariada, para un período de retorno dado, la función a optimizar (ver Ecuación 2.2-30) se resuelve dando varios valores a Q_p , y encontrando para cada uno el valor de V que satisface la ecuación 2.2-31 (Dueñas Puebla, 2016).

Para cada pareja de valores Q_p, V obtenidos, se puede generar una avenida de diseño de gran volumen, de gran gasto de pico o intermedia (Fig. 3-6); para identificar la paraje más desfavorable se simula el tránsito de casa una de las avenidas por el vaso de la presa.

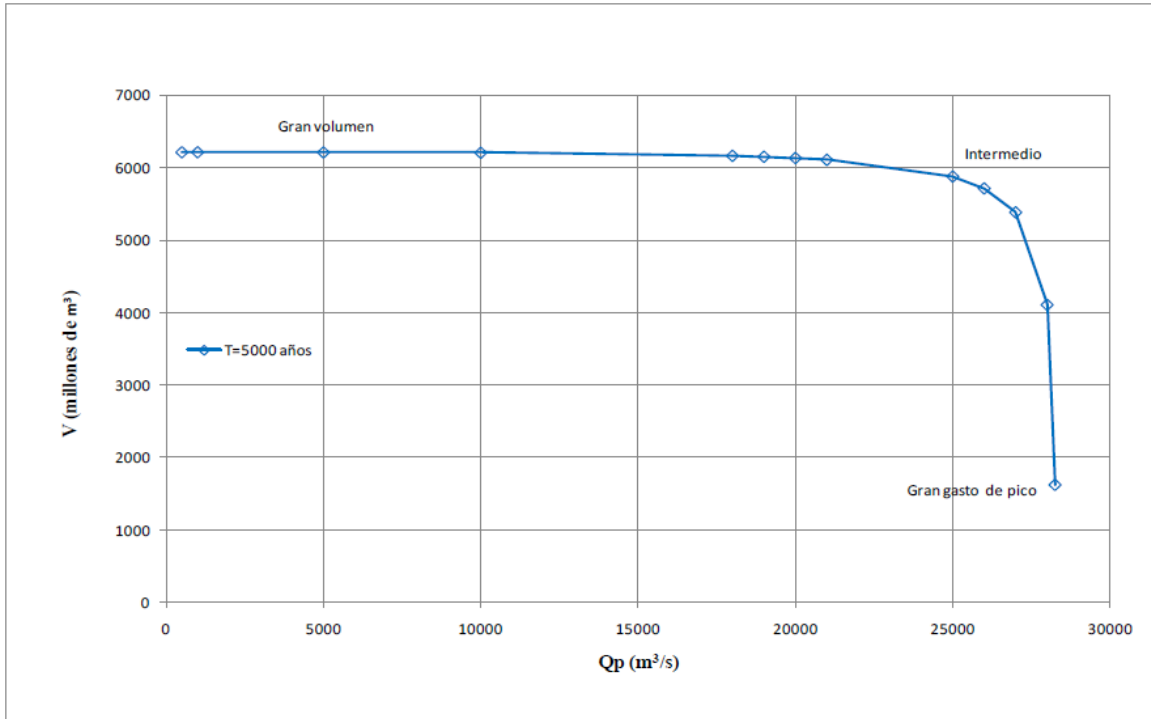


Figura 3-6. Curvas de eventos de diseño Q_p , V , Dueñas (2016).

Finalmente, para el cálculo de los volúmenes de escurrimiento, se implementó el método de Ramírez Aldama y Dueñas (2016) lo presenta en su tesis de maestría para un cálculo de 46 avenidas correspondientes a los gastos máximos anuales (1965-1984 y 1988-2013); cada una con distintas duraciones, ya que el inicio y final de las avenidas se consideró de manera aproximada por apreciación de los datos de gastos medios diarios. Se calculó el volumen de cada avenida y se obtuvo un registro de parejas de valores para gasto pico y volumen.

Los datos para el análisis se presentan en la Tabla 3-9, correspondientes a los gastos máximos anuales (1965-1984 y 1988-2013); calculados para distintas duraciones, ya que el inicio y final de las avenidas se consideró de manera aproximada por apreciación de los datos de gastos medios diarios, y los datos reportados corresponden a los cálculos del volumen de cada avenida para un registro de parejas de valores de gasto y volumen realizadas en la tesis de Dueñas (2016).

Tabla 3-9. Registro de gastos de pico y volumen de avenidas máximas históricas Presa Adolfo López Mateos (C.H. Infiernillo).

Año	Q m³/s	V (Millones de m³)	Año	Q m³/s	V (Millones de m³)
1965	4254.61	3826.31	1988	5928.70	3113.26
1966	2550.66	1192.73	1989	2713.76	2165.98
1967	14109.10	6867.26	1990	2152.94	1705.48
1968	2681.26	1892.40	1991	2550.06	1628.46
1969	5940.10	4241.13	1992	5069.59	2015.87
1970	3671.70	3094.23	1993	3674.30	2350.30
1971	5603.00	3282.71	1994	1754.57	1337.64
1972	2905.35	2201.09	1995	5513.89	3820.60
1973	7142.70	4026.61	1996	3019.68	2547.25
1974	2897.50	2131.06	1997	3572.92	483.05
1975	4740.20	2911.51	1998	4063.66	1447.00
1976	9720.60	3623.57	1999	3510.42	1651.50
1977	3162.50	1808.04	2000	5949.07	2465.85
1978	3230.00	2116.07	2001	8118.06	2250.15
1979	2264.75	2004.88	2002	3740.74	491.05
1980	2591.29	2324.75	2003	3147.41	524.55
1981	4203.20	3397.73	2004	2827.88	2324.64
1982	1088.11	839.78	2005	3519.05	2733.61
1983	2059.30	1490.48	2006	2106.23	1463.55
1984	7408.70	4157.61	2007	3737.44	1345.01
1985	2344.28	1860.64	2008	5242.71	3141.10
1986	1816.77	1090.68	2009	2282.33	1701.36
1987	2688.33	1323.89			

Estadísticos	Q m³/s	V (Millones de m³)
media	4072.654	2320.276
desvest	2354.456	1193.731
coef asim	2.176	1.276
coef de var	0.578	0.514
coef correl	0.791	

Tabla 3-10. Resumen de errores estándar para gasto (m³/s)

Función	Momentos		Máxima Verosimilitud	
	2 parámetros	3 parámetros	2 parámetros	3 parámetros
Normal	1126.949	-----	1126.949	-----
Lognormal	611.668	592.31	770.013	817.886
Gumbel	744.787	-----	10318.724	-----
Exponencial	547.331	-----	1373.409	-----
Gamma	738.048	602.189	882.285	783.836
Doble Gumbel	495.33			

Tabla 3-11. Resumen de errores estándar para volumen (millones de m³)

Función	Momentos		Máxima Verosimilitud	
	2 parámetros	3 parámetros	2 parámetros	3 parámetros
Normal	376.881	-----	376.881	-----
Lognormal	266.586	261.571	242.479	273.534
Gumbel	259.159	-----	265.336	-----
Exponencial	312.123	-----	933.941	-----
Gamma	269.774	268.648	272.796	11111
Doble Gumbel	291.327			

Siendo el mínimo error estándar para gasto y volumen de 495.33 y de 242.479 respectivamente, calculados con la función doble Gumbel para el gasto y log normal (max. Ver.) 2 p. para el volumen.

Sin embargo se aplicó la distribución de probabilidad Doble Gumbel para ambos datos debido a la naturaleza de la función a optimizar y se procedió a calcular los valores de escurrimientos extrapolados para los diferentes períodos de retorno.

Los parámetros estadísticos de esta distribución para ambos valores se presentan en las Tablas 3-12 y 3-13:

Tabla 3-12. Parámetros estadísticos para Q y V con AX

Datos	Q, m3/s	V, millones de m3
Media	4072.654	2320.276
Desviación	2381.061	1207.22
Asimetría	2.176	1.276
Error estándar	340.777	218.44

Tabla 3-13. Parámetros de la función Doble Gumbel con AX para Q y V

Parámetros	para Q, m3/s	para V, millones de m3
$\alpha 1$	0.00096	0.001273
$\beta 1$	2778.749	1708.2983
$\alpha 2$	0.00033	0.000474
$\beta 2$	7139.797	4264.7297

Los valores extrapolados se presentan en la Tabla 3-14

Tabla 3-14. Valores de gasto (Q) y volumen (V), extrapolados para distintos períodos de retorno (Tr), en años.

No	Tr	Q cal, m ³ /s	V cal, millones de m ³
1	2	3441.48	2072.7
2	5	5274.2	3117.71
3	10	7339.78	3952.46
4	20	10003.95	4980.4
5	50	13324.44	6756.49
6	100	15681.55	8260.02
7	200	17978.1	9767.8
8	500	20970.97	11740.48
9	1000	23219.95	13210.11
10	2000	25434.34	14679.73
11	5000	28409.91	16578.94
12	10000	30762.69	18025.95

Se estiman los parámetros de la función de probabilidad de distribución Doble Gumbel: α_1 , β_1 , α_2 y β_2 , para cada valor de probabilidad. Los resultados obtenidos se muestran en la Tabla 3-16 para volumen y en la Tabla 3-17 para los datos de gastos.

Tabla 3-15. Resultados para los cinco parámetros de la función de densidad Doble Gumbel para gasto (m³/s)

Pq	a1	b1	a2	b2	Error Estándar de ajuste
0.7	0.001239	2580.3862	0.000397	5467.3974	450.829
0.71	0.001239	2580.3862	0.000405	5634.7553	459.54
0.72	0.001243	2586.6692	0.00041	5746.7708	461.508
0.73	0.001206	2612.5359	0.000402	5804.819	450.828
0.74	0.001198	2619.491	0.000404	5911.6322	450.105
0.75	0.001163	2645.6859	0.000392	5971.3457	436.395
0.76	0.001139	2653.9876	0.000393	6032.8475	431.792
0.77	0.001105	2653.9876	0.000378	6093.7854	414.486
0.78	0.001104	2689.8617	0.000381	6222.186	414.351
0.79	0.001071	2689.8617	0.000366	6285.0363	396.442
0.8	0.00106	2700.2531	0.00037	6455.8247	397.601
0.81	0.001039	2727.5284	0.000355	6521.035	379.43
0.82	0.000998	2727.5284	0.000338	6586.9041	359.327
0.83	0.000994	2738.6957	0.000339	6777.0189	356.82
0.84	0.000964	2766.3593	0.000322	6914.6198	339.808
0.85	0.000963	2778.7489	0.000329	7139.7971	340.466
0.86	0.000934	2806.817	0.00031	7284.7639	319.947
0.87	0.000932	2819.61	0.000324	7616.1295	327.382
0.88	0.000905	2819.61	0.000302	7770.7678	302.52
0.89	0.000892	2859.458	0.000303	8070.3145	297.943
0.9	0.000857	2859.458	0.000274	8317.3481	269.355
0.91	0.000823	2888.3414	0.000245	8571.0862	247.04
0.92	0.00081	2902.4579	0.00025	9069.3868	240.435

Tabla 3-16. Resultados para los cinco parámetros de la función de densidad Doble Gumbel para volumen, en millones de m³

Pv	a1	b1	a2	b2	Error Estándar de ajuste
0.7	0.001691	1445.0793	0.001041	3217.8707	272.622
0.71	0.001691	1445.0793	0.001041	3283.2065	275.492
0.72	0.001688	1454.9038	0.001087	3319.6415	276.684
0.73	0.001654	1469.5997	0.001065	3353.1732	275.244
0.74	0.001655	1479.7852	0.001079	3381.3121	274.804
0.75	0.001622	1494.7326	0.001047	3415.4668	272.835
0.76	0.001605	1504.8964	0.001062	3455.7601	273.066
0.77	0.001589	1504.8964	0.00103	3490.6668	271.211
0.78	0.001565	1530.0466	0.001049	3511.1925	269.966
0.79	0.001534	1530.0466	0.001017	3546.6591	267.604
0.8	0.001522	1540.0658	0.001038	3604.5547	268.737
0.81	0.001507	1555.622	0.000997	3640.9644	266.128
0.82	0.001477	1571.3353	0.000958	3677.7418	263.682
0.83	0.00145	1581.5577	0.000972	3708.5572	261.965
0.84	0.001435	1581.5577	0.000925	3783.8559	261.347
0.85	0.001413	1608.114	0.000918	3804.203	257.948
0.86	0.001385	1608.114	0.000873	3842.6293	254.482
0.87	0.001385	1618.9358	0.000855	3895.256	252.482
0.88	0.001358	1635.2887	0.000789	3934.602	247.786
0.89	0.001344	1646.3337	0.000792	4002.3695	245.672
0.9	0.001317	1662.9634	0.000709	4042.7975	239.239
0.91	0.001304	1679.761	0.000616	4124.8827	234.245
0.92	0.001286	1691.2154	0.00058	4179.8615	226.887
0.93	0.001273	1708.2983	0.000474	4264.7297	218.144

De las tablas anteriores se hace una síntesis de los parámetros de la función empleada, en la cual se puede observar los 10 parámetros de ajuste a optimizar; es decir, cinco correspondientes al gasto y cinco al volumen, los resultados se pueden observar en la Tabla 3-18.

Tabla 3-17. Resultados de la optimización del ajuste para la Función de distribución de probabilidad Doble Gumbel

Parámetros de Ajuste	Q (m ³ /s) anual	V (millones de m ³)
α_1	0.00081	0.001273
β_1	2902.4579	1708.2983
α_2	0.00025	0.000474
β_2	9069.3868	4264.7297
P	0.92	0.93
Error Estándar de ajustes	240.435	218.144

Es importante destacar que la función de distribución doble Gumbel Bivariada (FDDGB) consta de once parámetros, los que serán necesarios optimizar; es decir, los 10 parámetros mostrados en la Tabla anterior más uno de asociación llamado m .

Posteriormente se va a llevar a cabo el análisis de las frecuencias.

Primero, se calculó la frecuencia empírica para los gastos pico y volumen de la avenida presentados de la Tabla 3-9, siguiendo el siguiente procedimiento: el registro se ordena de mayor a menor en función del gasto pico, garantizando de esta manera el gasto mayor; para asignar el número de orden es necesario contabilizar el número de veces que un dato de volumen es menor o igual a sí mismo (ver k en la Tabla 3-18) en el registro completo. Una vez que se obtiene el valor de número de orden (k), entendido como el número de valores menores o iguales al gasto pico y al volumen de la avenida o ambos (ver Tabla 3-18), se calcula el valor de la frecuencia empírica como $k/(n+1)$, donde n es el número total de datos que contiene el registro.

Como segundo paso se obtiene la Función teórica (FDDGB), empleando los parámetros de ajuste encontrados con el programa AX, mediante el empleo de sus funciones marginales tal y como se observa en la Tabla 3-9.

Tabla 3-18. Cálculo de frecuencias para la función de distribución de densidad doble Gumbel bivariada para los datos de gasto pico y volumen del hidrograma.

N	t, años	Gasto máximo, m3/s	Volumen máximo	k :No. Valores <=Q,V o ambos	Fempírica k/(n+1)	Marginal FQ	Marginal FV	Fteórica (AX)
1	1965	14109.10	6867.26	45	0.9783	0.9801	0.9810	0.9717
2	1966	9720.60	3623.57	39	0.8478	0.9505	0.8703	0.8601
3	1967	8118.06	2250.15	26	0.5652	0.9291	0.5683	0.5641
4	1968	7408.70	4157.61	43	0.9348	0.9140	0.9142	0.8769
5	1969	7142.70	4026.61	42	0.9130	0.9067	0.9055	0.8657
6	1970	5949.07	2465.85	30	0.6522	0.8542	0.6419	0.6194
7	1971	5940.10	4241.13	44	0.9565	0.8537	0.9192	0.8320
8	1972	5928.70	3113.26	35	0.7609	0.8530	0.7993	0.7532
9	1973	5603.00	3282.71	37	0.8043	0.8298	0.8270	0.7593
10	1974	5513.89	3820.60	40	0.8696	0.8225	0.8893	0.7910
11	1975	5242.71	3141.10	36	0.7826	0.7976	0.8041	0.7227
12	1976	5069.59	2015.87	25	0.5435	0.7792	0.4769	0.4517
13	1977	4740.20	2911.51	33	0.7174	0.7383	0.7597	0.6549
14	1978	4254.61	3826.31	41	0.8913	0.6613	0.8898	0.6469
15	1979	4203.20	3397.73	38	0.8261	0.6519	0.8433	0.6250
16	1980	4063.66	1447.00	10	0.2174	0.6251	0.2321	0.2104
17	1981	3740.74	491.05	2	0.0435	0.5559	0.0086	0.0081
18	1982	3737.44	1345.01	9	0.1957	0.5551	0.1913	0.1674
19	1983	3674.30	2350.30	29	0.6304	0.5404	0.6038	0.4395
20	1984	3671.70	3094.23	34	0.7391	0.5398	0.7958	0.5118
21	1985	3572.92	483.05	1	0.0217	0.5162	0.0082	0.0076
22	1986	3519.05	2733.61	32	0.6957	0.5029	0.7180	0.4577
23	1987	3510.42	1651.50	14	0.3043	0.5008	0.3196	0.2535
24	1988	3230.00	2116.07	22	0.4783	0.4283	0.5173	0.3299
25	1989	3162.50	1808.04	17	0.3696	0.4102	0.3883	0.2609
26	1990	3147.41	524.55	3	0.0652	0.4062	0.0104	0.0091
27	1991	3019.68	2547.25	31	0.6739	0.3714	0.6668	0.3354
28	1992	2905.35	2201.09	25	0.5435	0.3400	0.5501	0.2817
29	1993	2897.50	2131.06	23	0.5000	0.3378	0.5232	0.2725
30	1994	2827.88	2324.64	27	0.5870	0.3187	0.5950	0.2767
31	1995	2713.76	2165.98	24	0.5217	0.2875	0.5367	0.2398
32	1996	2688.33	1323.89	7	0.1522	0.2806	0.1832	0.1121
33	1997	2681.26	1892.40	19	0.4130	0.2787	0.4248	0.2050
34	1998	2591.29	2324.75	28	0.6087	0.2546	0.5950	0.2250
35	1999	2550.66	1192.73	6	0.1304	0.2439	0.1363	0.0805
36	2000	2550.06	1628.46	13	0.2826	0.2437	0.3096	0.1501
37	2001	2344.28	1860.64	18	0.3913	0.1915	0.4112	0.1456
38	2002	2282.33	1701.36	15	0.3261	0.1766	0.3415	0.1225
39	2003	2264.75	2004.88	20	0.4348	0.1724	0.4723	0.1419
40	2004	2152.94	1705.48	16	0.3478	0.1471	0.3433	0.1049
41	2005	2106.23	1463.55	11	0.2391	0.1371	0.2390	0.0799
42	2006	2059.30	1490.48	12	0.2609	0.1273	0.2502	0.0773
43	2007	1816.77	1090.68	5	0.1087	0.0828	0.1043	0.0309
44	2008	1754.57	1337.64	8	0.1739	0.0732	0.1885	0.0410
45	2009	1088.11	839.78	4	0.0870	0.0120	0.0458	0.0038

Posteriormente se grafican las frecuencias empíricas contras las estimadas para la función bivariada (Figuras 3-8) y se grafica también la función identidad para observar que tan cerca están a la misma.

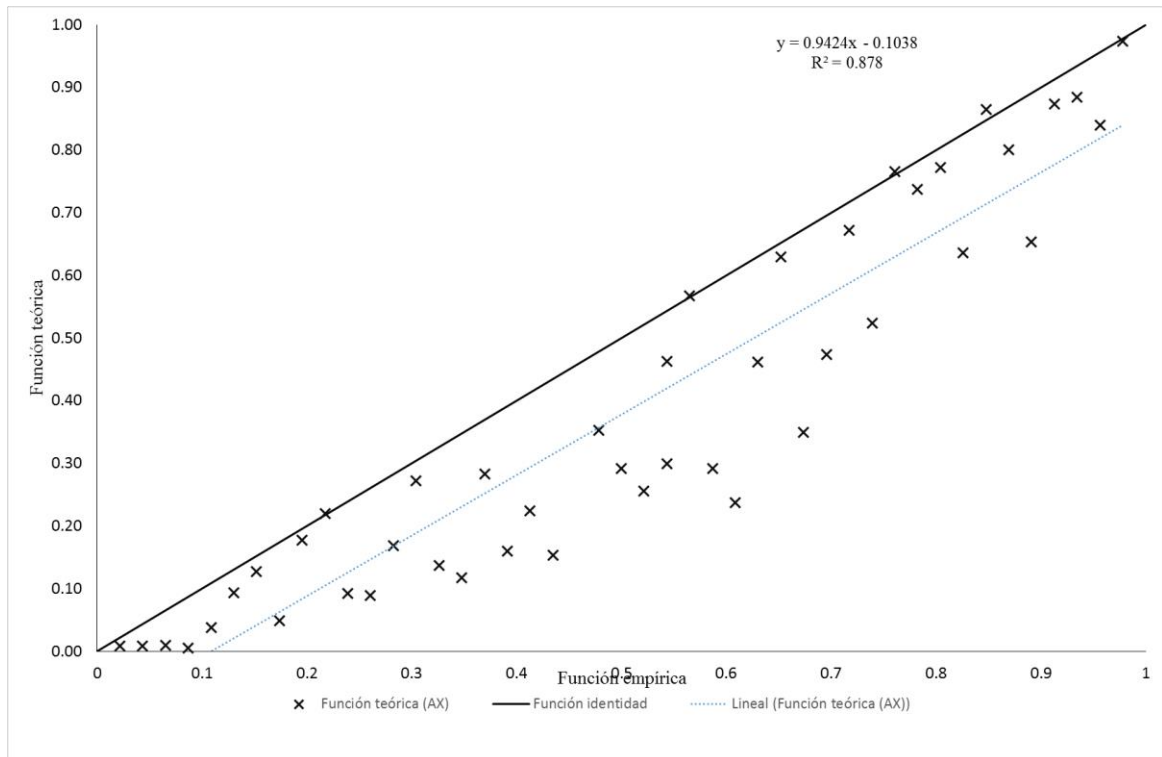


Figura 3-8. Análisis para el ajuste de la función de probabilidad de la función bivariada para datos de gasto y volumen. Presa El Infiernillo, Mich.

3.2.6. Ajuste de parámetros

El segundo paso en el proceso consiste en obtener los parámetros de ajuste para la función de distribución bivariada, tanto para los datos de gasto y de volumen. El ajuste se realiza mediante las ecuaciones presentadas en la Tabla 3-19:

Tabla 3-19. Ecuaciones para realizar el ajuste de los parámetros para la FDDGB (equivalencias con las marginales obtenidas con el programa AX)

Parámetros para gasto	Parámetros para volumen
$a_1 = \alpha_1$	$a_3 = -\beta_1$
$c_1 = 1/\alpha_1$	$c_3 = 1/\alpha_1$
$a_2 = -\beta_2$	$a_4 = -\beta_2$
$c_2 = 1/\alpha_2$	$c_2 = 1/\alpha_2$

Las probabilidades se conservan iguales. Los resultados del ajuste de parámetros se observan en la Tabla 3-20.

Tabla 3-20. Obtención de los once parámetros de la FDDGB. Equivalencia bivariada

Parámetros de gastos	a ₁	-2902.4579
	c ₁	1234.567901
	a ₂	-9069.3868
	c ₂	4000
Parámetros de volúmenes	a ₃	-1708.2983
	c ₃	785.5459544
	a ₄	-4264.7297
	c ₄	2109.704641
Probabilidades	P _q	0.92
	P _v	0.93
Parámetro de asociación	m	2.18735096

Los valores de la tabla anterior servirán para dar los valores iniciales al algoritmo híbrido.

3.2.7. Implementación del modelo híbrido

3.2.7.1. Primer paso del modelo híbrido: AG

La población inicial para este caso de estudio corresponde a la codificación de los parámetros a optimizar (11 en total). El límite superior como inferior de dicha población (ver Tabla 3-21).

Tabla 3-21. Creación de límites para evaluación de la función objetivo en AG

Parámetros de la FDDGB		Limite Inferior	Parametros del ajuste para la FDDGB	Limite Superior
Parámetros de gastos	a ₁	-3,000.00	-2,902.46	-2,500.00
	c ₁	1,100.00	1,234.57	1,300.00
	a ₂	-12,000.00	-9,069.39	-1,000.00
	c ₂	3,500.00	4,000.00	4,500.00
Parámetros de volúmenes	a ₃	-2,500.00	-1,708.30	-1,650.00
	c ₃	700.00	785.55	800.00
	a ₄	-4,300.00	-4,264.73	-4,000.00
	c ₄	2,100.00	2,109.70	2,200.00
Probabilidades	P _q	0.89	0.90	0.92
	P _v	0.88	0.90	0.93
Asociación	m	2.00	2.19	2.30

Se realizó el proceso de generación de poblaciones, evaluación de la función objetivo, selección, cruza, mutación y finalmente maximización de los parámetros de la FDDGB con el código de AGs,

Como resultados se obtuvo un vector de 11 renglones y una columna, los renglones corresponden al número de parámetros a optimizar. La matriz de datos iniciales M tiene dimensiones de 11×12 , donde cada renglón representa cada uno de los once parámetros a optimizar y cada columna representa un vector para un cierto número de generaciones. Se contabilizó el tiempo de generación.

La matriz M representa el simplex inicial para el segundo proceso del método híbrido; es decir, el método Nelder-Mead. Para la coinformación de la matriz de datos iniciales M , se llevó el análisis de distintos casos, con la finalidad de comparar tiempos computacional y eficiencia del método.

3.2.7.1.1. Caso I. Para la generación de matriz de parámetro iniciales " M "

El primero caso, al cual se nombró "Modeo híbrido 1" se llevó a cabo con la conformación de la matriz M para 200 individuos y generaciones de 10, 20, 30, ... y 120 para cada uno de los 12 vértices. Para su generación se tomó un tiempo de total de 31.96 segundos (ver tabla 3-22).

Se decidió emplear este número de generaciones por el motivo de que el tiempo se redujera mucho más que con las 5 000 generaciones empleados para llegar a un óptimo en los 178.41 segundos sólo con AGs.

Tabla 3-22. Generación de M , con AGs y un tiempo total de cálculo de 32 segundos

No. De vértices =	1	2	3	4	5	6	7	8	9	10	11	12
MAXGEN = Máximo número de generaciones	10	20	30	40	50	60	70	80	90	100	110	120
a_1	-2504.06	-2505.97	-2501.95	-2500.23	-2501.97	-2500.23	-2500.49	-2500.15	-2500.12	-2500.00	-2500.03	-2500.02
c_1	1297.92	1293.73	1280.86	1297.97	1296.34	1299.47	1298.90	1299.83	1298.25	1299.97	1299.60	1299.82
a_2	-1863.43	-1055.39	-1021.82	-1001.34	-1011.75	-1018.46	-1006.71	-1000.67	-1000.67	-1004.36	-1001.01	-1000.00
c_2	4009.38	4103.29	3505.80	3657.26	3599.70	3691.44	3509.61	3509.06	3512.18	3509.25	3500.12	3504.73
a_3	-2438.75	-2465.27	-2496.68	-2498.13	-2499.61	-2499.77	-2500.00	-2499.71	-2499.77	-2499.97	-2499.90	-2499.92
c_3	799.18	799.86	799.89	799.41	799.67	799.95	799.93	799.85	799.78	799.93	799.94	799.90
a_4	-4207.59	-4298.28	-4294.58	-4295.23	-4291.05	-4293.82	-4299.91	-4298.71	-4299.46	-4299.81	-4299.66	-4299.90
c_4	2149.19	2183.69	2199.10	2179.69	2197.87	2186.77	2195.93	2193.62	2191.10	2196.28	2194.43	2197.83
P_q	0.90	0.90	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89	0.89
P_v	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88	0.88
m	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00
$t, \text{seg} =$	0.442	0.814	1.174	1.590	2.113	2.534	2.867	3.245	3.669	4.287	4.508	4.722
$t, \text{mín} =$	0.007	0.014	0.020	0.026	0.035	0.042	0.048	0.054	0.061	0.071	0.075	0.079

3.2.7.1.2. Caso II. Para la generación de matriz de parámetro inicales “ M ”

El segundo caso de análisis (Modelo híbrido 2) se llevó a cabo con la conformación de la matriz M , para 200 individuos pero esta vez, para generaciones de 50, 100, 200, 250, 300, 500, 700, 800, 900, 1500, 2000 y 2500 correspondientes a la generación para cada uno de los 12 vértices. Su generación se tomó un tiempo de total de 100.75 segundos (ver Tabla 3-23).

Este caso de análisis se decidió llevar a cabo porque es un campo inicial de búsqueda más cercano al óptimo realizado con el AG, considerando las cinco mil iteraciones, buscando con esto un acercamiento más próximo al óptimo local.

Tabla 3-23. Generación de M , con AGs y un tiempo total de cálculo de 102 segundos

No Vértice	1	2	3	4	5	6	7	8	9	10	11	12
MAXGEN = Máximo número de generaciones	50	100	200	250	300	500	700	800	900	1500	2000	2500
a_1	-2794.20	-2830.39	-2785.33	-2796.78	-2794.29	-2815.73	-2783.35	-2763.45	-2787.91	-2777.35	-2778.79	-2790.32
c_1	1104.80	1117.80	1106.87	1106.56	1117.74	1111.28	1101.15	1105.95	1102.97	1100.95	1100.14	1103.65
a_2	-7591.88	-7715.41	-7671.10	-7443.16	-7657.67	-7982.64	-7416.30	-7588.85	-7555.28	-7540.18	-7583.48	-7571.40
c_2	3563.33	3564.03	3506.35	3501.59	3504.15	3532.50	3504.27	3500.89	3509.92	3504.30	3500.89	3501.04
a_3	-1730.16	-1756.67	-1726.68	-1733.76	-1720.64	-1735.89	-1728.47	-1722.30	-1731.38	-1726.19	-1730.03	-1735.40
c_3	798.18	799.94	799.91	799.57	799.99	799.81	799.95	799.94	799.68	799.79	799.92	799.97
a_4	-4015.86	-4023.04	-4006.12	-4000.33	-4003.35	-4001.39	-4000.60	-4007.74	-4001.68	-4000.06	-4000.46	-4001.17
c_4	2108.88	2107.65	2103.54	2101.40	2100.22	2101.26	2100.34	2107.89	2102.50	2100.26	2100.11	2100.18
P_q	0.89	0.89	0.89	0.89	0.90	0.90	0.89	0.89	0.89	0.89	0.89	0.89
P_v	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93	0.93
m	2.11	2.08	2.10	2.08	2.10	2.09	2.10	2.12	2.10	2.10	2.10	2.10
$t, \text{seg} =$	1.056	1.333	1.901	2.138	2.477	3.844	5.589	6.526	7.356	14.755	21.691	32.084
$t, \text{min} =$	0.018	0.022	0.032	0.036	0.041	0.064	0.093	0.109	0.123	0.246	0.362	0.535

3.2.7.1.3. Caso III. Para la generación de matriz de parámetro iniciales " M "

El tercer caso se llevó a cabo mediante el análisis de un conjunto de combinaciones entre número de individuos y el número de generaciones. Este caso ya considera el número de iteraciones para Nelder-Mead. Las combinaciones probadas fueron las mostradas en la Tabla 3-24:

Tabla 3-24. Combinaciones empleadas para la conformación de la matriz de parámetros iniciales, del caso III.

N	Algoritmos genéticos		Nealder-Mead	Tiempos computacionales			Max Ver
	N Individuos	N Generaciones	N Iteraciones	† AG, s	† NM, s	† total, s	Valor de FO
1	200	5000		178.406	0.000	178.41	
2	200	2500	60	158.476	0.566	159.04	16.40
3	100	500	50	25.024	0.545	25.57	17.88
4	100	200	110	6.057	1.255	7.31	31.07
5	200	2000	50	124.992	1.010	126.00	20.11
6	200	1000	100	60.819	1.315	62.13	17.02

Como se puede observar los mejores resultados, según el valor de la función objetivo (por máxima verosimilitud) son los casos 2, 3, 5 y 6, mismos que se emplearan para el siguiente proceso dentro del método híbrido.

3.2.7.2. Segundo paso del modelo híbrido: N-M

Las matrices de los tres casos descritos anteriormente representan el campo inicial de búsqueda para el segundo proceso del método híbrido (Nelder-Mead). A modo de ejemplo ilustrativo, se tomó la matriz “M” del caso I.

Primero se evaluaron los vértices con la función objetivo elegido y posteriormente se ordenaron los vértices de acuerdo con el valor de esta. A la matriz ordenada se le llamó A, tal como se muestra en la Tabla 3-25.

Tabla 3-25. Proceso Nelder-Mead para ordenamiento de vértices.

Parámetros	Vértices de la matriz ordenada "A"											
	Mejor	2do Mejor	3er Mejor	4to Mejor	5to Mejor	6to Mejor	7mo Mejor	8vo Mejor	9no Mejor	10mo Mejor	11mo Mejor	Peor
a_1	-2504.10	-2506.00	-2502.00	-2502.00	-2500.50	-1219.40	-2500.00	-2500.20	453.87	-3091.40	-2500.20	-1034.60
c_1	1297.90	1293.70	1280.90	1296.30	1298.90	648.49	1299.60	1298.00	-220.17	1609.60	1299.80	555.31
a_2	-1863.40	-1055.40	-1021.80	-1011.80	-1006.70	-147.07	-1001.00	-1001.30	516.56	-1242.20	-1000.70	-45.00
c_2	4009.40	4103.30	3505.80	3599.70	3509.60	1294.50	3500.10	3657.30	-760.90	4495.80	3509.10	996.80
a_3	-2438.80	-2465.30	-2496.70	-2499.60	-2500.00	-1260.80	-2499.90	-2498.10	408.82	-3093.20	-2499.70	-1082.40
c_3	799.18	799.86	799.89	799.67	799.93	392.33	799.94	799.41	-142.76	989.42	799.85	333.71
a_4	-4207.60	-4298.30	-4294.60	-4291.10	-4299.90	-2145.00	-4299.70	-4295.20	739.41	-5313.50	-4298.70	-1834.00
c_4	2149.20	2183.70	2199.10	2197.90	2195.90	1106.10	2194.40	2179.70	-381.54	2715.30	2193.60	941.18
P_q	0.90	0.90	0.89	0.89	0.89	0.99	0.89	0.89	0.99	0.99	0.89	0.99
P_v	0.88	0.88	0.88	0.88	0.88	0.99	0.88	0.88	0.99	0.99	0.88	0.99
m	2.00	2.00	2.00	2.00	2.00	0.98	2.00	2.00	-0.36	2.47	2.00	0.83
FO	16.505300	16.618500	16.673300	16.702200	16.707700	16.708400	16.709700	16.709800	16.709800	16.709800	16.710000	16.710400

Una obtenidos los vértices ordenados se calcula el centro geométrico sin considerase el peor vértice. Posteriormente se obtiene el vector de reflexión. Ambos vectores se pueden consultar en la Tabla 3-26.

Tabla 3-26. Vector del centro geométrico y reflexión para el primer caso de análisis, FDDGB.

Parámetros	Centro	Reflexión
	Xc	Xr
a_1	1031.70	-1108.20
c_1	-785.24	577.93
a_2	2811.30	-456.93
c_2	-1984.80	1684.40
a_3	634.64	-1108.40
c_3	-3411.30	354.79
a_4	1741.80	-1902.20
c_4	0.92	973.57
P_q	0.91	0.99
P_v	1.59	0.99
m	3.38	0.89

Una vez que se calculan los dos vectores anteriores se empieza con el algoritmo Nelader-Mead (para el cual se considero el algoritmo de Herrera, 2012, basado en Mathews & Kurtis, 2000), realizando las operaciones adecuadas de expansión, contracción o encogimiento, según sea el caso. También se realizó una comparación con AX. En la Tabla 3-27 se observan los resultados obtenidos para los casos propuestos.

Tabla 3-27. Parámetros optimizados de la función de densidad Doble Gumbel Bivariada con diversos métodos

Parámetro	AG	AX	Caso i	Caso ii	Caso iii				
					200 Ind, 2500 gen	100 Ind, 500 gen	200 Ind, 2000 gen	200 Ind, 1000 gen	
Q, m ³ /s	a ₁	-2,902.46	-2,902.46	-2,484.29	-2,618.50	-2,779.50	-2,781.60	-2,779.60	-2,802.50
	c ₁	1,297.92	1,234.57	1,299.66	1,262.30	1,098.40	1,100.00	510.60	1,060.80
	a ₂	-1,863.43	-9,069.39	3,938.72	-5,328.00	-7,197.40	-7,204.00	-7,201.40	-7,125.30
	c ₂	4,009.38	4,000.00	1,422.94	2,231.60	-1,256.20	1,920.80	1,811.60	1,043.10
V, m ³	a ₃	-2,438.75	-1,708.30	-2,819.99	-1,107.50	-1,631.40	717.20	-1,633.40	-1,187.00
	c ₃	799.18	785.55	803.55	813.52	804.84	830.32	1,136.30	1,200.30
	a ₄	-4,207.59	-4,264.73	-4,821.92	-3,571.10	-3,000.10	-1,949.00	-3,000.00	-2,083.50
	c ₄	2,149.19	2,109.70	2,437.20	1,879.20	808.16	170.34	-977.64	1,194.80
P _q	0.90	0.92	0.84	0.98	0.84	0.88	0.89	1.06	
P _v	0.88	0.93	0.88	0.93	0.82	0.82	0.82	0.85	
m	2.00	2.19	2.00	1.82	2.07	2.07	2.07	2.08	
Medición del tiempo									
TAG, s	178.41		39.32	101.96	103.54	25.02	87.69	60.82	
TNM, s	0.000		0.550	0.576	0.566	0.545	1.010	1.315	
Ttotal, s	178.41		39.87	102.53	104.11	25.57	88.70	62.13	

3.2.8. Ajuste de la función de distribución de densidad doble Gumbel bivariada

Finalmente se calculan las frecuencias teóricas de la función de distribución doble Gumbel Bivariada con los once parámetros calculados de cada combinación planteada. La FDDGB se calcula mediante sus funciones marginales.

Como primer paso es necesario ordenar los datos de gasto y volumen, de forma descendente en función del gasto. Posteriormente es importante analizar las frecuencias con las que cada dato de volumen se presenta en el registro total menores o iguales al máximo anterior (k). Luego se calcula el valor de la función empírica como $k/(n+1)$. En seguida se calculan las funciones marginales para el gasto y para el volumen con los parámetros calculados mediante el modelo híbrido y para cada caso de estudio planteado y finalmente se obtiene la FDDGB mediante sus funciones marginales.

Los cálculos correspondientes se pueden observar en la Tabla 3-28.

Tabla 3-28. Cálculo de la función de distribución Doble Gumbel Bivariada para los casos propuestos del modelo híbrido.

N	t, años	Gasto máximo, m3(s)	Volumen máximo	k :No. Valores <=Q,V o ambos	Fempírica k/(n+1)	Marginal FQ	Marginal FV	Fteórica (AX)	Marginal FQ	Marginal FV	Fteórica (2500 gen, 200 ind)	Marginal FQ	Marginal FV	Fteórica (Caso I)	Marginal FQ	Marginal FV	Fteórica (AG)	Marginal FQ	Marginal FV	F(2000 gen y 200 ind)	Marginal FQ	Marginal FV	Fteórica (1000 gen, 200 ind)
1	1965	14109.10	6867.26	45	0.9783	0.9801	0.9801	0.9801	0.8400	0.9973	0.8400	0.9969	0.9999	0.9881	0.9988	1.0000	0.9649	0.9975	0.8102	0.8102	0.9900	0.9898	0.9862
2	1966	9720.60	3623.57	39	0.8478	0.9801	0.9801	0.9801	0.8387	0.8667	0.7993	0.9712	0.9956	0.9149	0.9562	0.9998	0.8429	0.9751	0.7154	0.7150	0.9900	0.8588	0.8587
3	1967	8118.06	2250.15	26	0.5652	0.9801	0.9801	0.9801	0.8536	0.5292	0.5204	0.9398	0.9491	0.7362	0.8885	0.9990	0.5501	0.9488	0.5718	0.5710	0.9900	0.6247	0.6247
4	1968	7408.70	4157.61	43	0.9348	0.9801	0.9801	0.9801	0.8767	0.9267	0.8603	0.9186	0.9977	0.9344	0.8607	0.9999	0.8619	0.9332	0.7412	0.7371	0.9900	0.9068	0.9066
5	1969	7142.70	4026.61	42	0.9130	0.9801	0.9801	0.9801	0.8858	0.9148	0.8618	0.9092	0.9973	0.9268	0.8528	0.9999	0.8508	0.9270	0.7350	0.7303	0.9900	0.8967	0.8965
6	1970	5949.07	2465.85	30	0.6522	0.9801	0.9801	0.9801	0.9049	0.6002	0.5953	0.8534	0.9740	0.7684	0.8166	0.9993	0.6092	0.9006	0.6079	0.6033	0.9900	0.6740	0.6739
7	1971	5940.10	4241.13	44	0.9565	0.9801	0.9801	0.9801	0.9048	0.9335	0.8869	0.8528	0.9979	0.9025	0.8162	0.9999	0.8182	0.9004	0.7451	0.7349	0.9900	0.9128	0.9125
8	1972	5928.70	3113.26	35	0.7609	0.9801	0.9801	0.9801	0.9047	0.7744	0.7614	0.8522	0.9917	0.8497	0.8157	0.9997	0.7421	0.9001	0.6826	0.6757	0.9900	0.7930	0.7929
9	1973	5603.00	3282.71	37	0.8043	0.9801	0.9801	0.9801	0.8990	0.8094	0.7908	0.8309	0.9933	0.8502	0.7994	0.9997	0.7491	0.8936	0.6953	0.6866	0.9863	0.8173	0.8171
10	1974	5513.89	3820.60	40	0.8696	0.9801	0.9801	0.9801	0.8963	0.8926	0.8557	0.8244	0.9965	0.8696	0.7941	0.9999	0.7790	0.8917	0.7252	0.7144	0.9804	0.8787	0.8779
11	1975	5242.71	3141.10	36	0.7826	0.9801	0.9801	0.9801	0.8849	0.7805	0.7604	0.8024	0.9920	0.8215	0.7754	0.9997	0.7147	0.8858	0.6849	0.6753	0.9589	0.7972	0.7951
12	1976	5069.59	2015.87	25	0.5435	0.9801	0.9801	0.9801	0.8749	0.4463	0.4422	0.7865	0.8818	0.6517	0.7611	0.9987	0.4475	0.8815	0.5267	0.5223	0.9421	0.5660	0.5649
13	1977	4740.20	2911.51	33	0.7174	0.9801	0.9801	0.9801	0.8491	0.7270	0.7011	0.7507	0.9891	0.7670	0.7283	0.9996	0.6510	0.8704	0.6643	0.6530	0.9025	0.7605	0.7497
14	1978	4254.61	3826.31	41	0.8913	0.9801	0.9801	0.9801	0.7923	0.8933	0.7738	0.6820	0.9966	0.7393	0.6629	0.9999	0.6374	0.8397	0.7255	0.6995	0.8220	0.8793	0.7957
15	1979	4203.20	3397.73	38	0.8261	0.9801	0.9801	0.9801	0.7848	0.8306	0.7396	0.6733	0.9942	0.7233	0.6547	0.9998	0.6186	0.8347	0.7027	0.6783	0.8117	0.8324	0.7635
16	1980	4063.66	1447.00	10	0.2174	0.9801	0.9801	0.9801	0.7630	0.2329	0.2280	0.6485	0.7619	0.4419	0.6308	0.9976	0.2125	0.8185	0.4000	0.3940	0.7818	0.4063	0.3967
17	1981	3740.74	491.05	2	0.0435	0.9801	0.9801	0.9801	0.7038	0.0132	0.0131	0.5827	0.6496	0.1041	0.5675	0.9939	0.0082	0.7619	0.2214	0.2179	0.7015	0.1454	0.1423
18	1982	3737.44	1345.01	9	0.1957	0.9801	0.9801	0.9801	0.7031	0.1964	0.1902	0.5820	0.7546	0.3893	0.5668	0.9973	0.1713	0.7611	0.3766	0.3666	0.7006	0.3763	0.3587
19	1983	3674.30	2350.30	29	0.6304	0.9801	0.9801	0.9801	0.6901	0.5629	0.5083	0.5677	0.9636	0.5811	0.5531	0.9991	0.4465	0.7459	0.5893	0.5537	0.6830	0.6482	0.5707
20	1984	3671.70	3094.23	34	0.7391	0.9801	0.9801	0.9801	0.6895	0.7702	0.6382	0.5671	0.9915	0.6200	0.5525	0.9997	0.5091	0.7452	0.6811	0.6247	0.6823	0.7901	0.6452
21	1985	3572.92	483.05	1	0.0217	0.9801	0.9801	0.9801	0.6682	0.0127	0.0125	0.5438	0.6481	0.1004	0.5301	0.9938	0.0077	0.7180	0.2205	0.2151	0.6535	0.1436	0.1391
22	1986	3519.05	2733.61	32	0.6957	0.9801	0.9801	0.9801	0.6560	0.6798	0.5687	0.5307	0.9855	0.5761	0.5175	0.9995	0.4596	0.7013	0.6446	0.5777	0.6373	0.7284	0.5849
23	1987	3510.42	1651.50	14	0.3043	0.9801	0.9801	0.9801	0.6541	0.3095	0.2895	0.5285	0.7749	0.4511	0.5154	0.9981	0.2627	0.6985	0.4471	0.4209	0.6346	0.4657	0.4197
24	1988	3230.00	2116.07	22	0.4783	0.9801	0.9801	0.9801	0.5859	0.4824	0.4092	0.4550	0.9172	0.4754	0.4448	0.9989	0.3405	0.5864	0.5467	0.4571	0.5434	0.5917	0.4577
25	1989	3162.50	1808.04	17	0.3696	0.9801	0.9801	0.9801	0.5684	0.3689	0.3215	0.4362	0.8003	0.4265	0.4268	0.9984	0.2735	0.5530	0.4823	0.4013	0.5200	0.5099	0.4019
26	1990	3147.41	524.55	3	0.0652	0.9801	0.9801	0.9801	0.5645	0.0157	0.0152	0.4320	0.6556	0.1047	0.4227	0.9941	0.0095	0.5453	0.2254	0.2055	0.5148	0.1529	0.1403
27	1991	3019.68	2547.25	31	0.6739	0.9801	0.9801	0.9801	0.5304	0.6255	0.4585	0.3955	0.9788	0.4534	0.3877	0.9993	0.3392	0.4748	0.6200	0.4219	0.4693	0.6913	0.4383
28	1992	2905.35	2201.09	25	0.5435	0.9801	0.9801	0.9801	0.4991	0.5123	0.3858	0.3621	0.9394	0.4093	0.3557	0.9990	0.2903	0.4059	0.5629	0.3525	0.4277	0.6128	0.3838
29	1993	2897.50	2131.06	23	0.5000	0.9801	0.9801	0.9801	0.4970	0.4878	0.3717	0.3598	0.9216	0.4032	0.3534	0.9989	0.2821	0.4010	0.5496	0.3449	0.4249	0.5955	0.3766
30	1994	2827.88	2324.64	27	0.5870	0.9801	0.9801	0.9801	0.4777	0.5544	0.3929	0.3392	0.9605	0.3964	0.3337	0.9991	0.2830	0.3571	0.5849	0.3210	0.3993	0.6423	0.3682
31	1995	2713.76	2165.98	24	0.5217	0.9801	0.9801	0.9801	0.4460	0.5001	0.3496	0.3055	0.9311	0.3605	0.3013	0.9989	0.2478	0.2844	0.5563	0.2564	0.3574	0.6042	0.3254
32	1996	2688.33	1323.89	7	0.1522	0.9801	0.9801	0.9801	0.4390	0.1891	0.1586	0.2980	0.7530	0.2679	0.2941	0.9973	0.1228	0.2683	0.3718	0.2017	0.3481	0.3700	0.2447
33	1997	2681.26	1892.40	19	0.4130	0.9801	0.9801	0.9801	0.4370	0.4006	0.2957	0.2959	0.8302	0.3358	0.2921	0.9985	0.2164	0.2639	0.5007	0.2301	0.3455	0.5331	0.2995
34	1998	2591.29	2324.75	28	0.6087	0.9801	0.9801	0.9801	0.4122	0.5544	0.3490	0.2695	0.9605	0.3355	0.2668	0.9991	0.2300	0.2089	0.5849	0.1953	0.3129	0.6423	0.2939
35	1999	2550.66	1192.73	6	0.1304	0.9801	0.9801	0.9801	0.4012	0.1459	0.1209	0.2577	0.7421	0.2244	0.2555	0.9969	0.0896	0.1853	0.3426	0.1426	0.2983	0.3315	0.2038
36	2000	2550.06	1628.46	13	0.2826	0.9801	0.9801	0.9801	0.4010	0.3007	0.2249	0.2575	0.7733	0.2837	0.2553	0.9980	0.1625	0.1850	0.4418	0.1592	0.2981	0.4591	0.2459
37	2001	2344.28	1860.64	18	0.3913	0.9801	0.9801	0.9801	0.3466	0.3887	0.2461	0.1998	0.8178	0.2548	0.1997	0.9985	0.1546	0.0850	0.4939	0.0791	0.2272	0.5244	0.2044
38	2002	2282.33	1701.36	15	0.3261	0.9801	0.9801	0.9801	0.3311	0.3284	0.2123	0.1833	0.7794	0.2317	0.1837	0.9982	0.1324	0.0628	0.4584	0.0581	0.2071	0.4800	0.1819
39	2003	2264.75	2004.88	20	0.4348	0.9801	0.9801	0.9801	0.3268	0.4423	0.2545	0.1787	0.8774	0.2419	0.1793	0.9987	0.1485	0.0572	0.5244	0.0544	0.2015	0.5631	0.1868
40	2004	2152.94	1705.48	16	0.3478	0.9801	0.9801	0.9801	0.3003	0.3300	0.1987	0.1506	0.7799	0.2035	0.1520	0.9982	0.1134	0.0292	0.4594	0.0276	0.1676	0.4811	0.1498
41	2005	2106.23	1463.55	11	0.2391	0.9801	0.9801	0.9801	0.2898	0.2390	0.1541	0.1394	0.7630	0.1795	0.1412	0.9976	0.0892	0.0211	0.4038	0.0196	0.1542	0.4111	0.1314
42	2006	2059.30	1490.48	12	0.2609	0.9801	0.9801	0.9801	0.2796	0.2489	0.1552	0.1287	0.7648	0.1722	0.1306	0.9977	0.0861	0.0147	0.4100	0.0138	0.1413	0.4190	0.1222
43	2007	1816.77	1090.68	5	0.1087	0.9801	0.9801	0.9801	0.2338	0.1155	0.0763	0.0800	0.7324	0.1047	0.0828	0.9965	0.0369	0.0012	0.3208	0.0011	0.0842	0.3019	0.0676
44	2008	1754.57	1337.64	8	0.1739	0.9801	0.9801	0.9801	0.2239	0.1939	0.1116	0.0695	0.7541	0.1099	0.0724	0.9973	0.0470	0.0005	0.3750	0.0005	0.0723	0.3741	0.0631
45	2009	1088.11	839.78	4	0.0870	0.9801	0.9801	0.9801	0.1666	0.0564	0.0350	0.0084	0.7037	0.0232	0.0097	0.9956	0.0051	0.0000	0.2723	0.0000	0.0069	0.2317	0.0059

La tabla anterior nos permite evaluar el error de una manera diferente a la función objetivo de máxima verosimilitud a través del coeficiente de determinación (R^2), al comparar la frecuencia empírica con las frecuencias calculadas para cada conjunto de parámetros correspondiente a cada combinación propuesta para ejecutar la metodología planteada en esta tesis.

De esta forma, se puede concluir en qué momento o en qué número de iteraciones, combinada con cierto número de individuos correspondiente al primer proceso del método híbrido (algoritmo genético), se entra al segundo proceso, correspondiente a Nelder-Mead y así concluir en un resultado óptimo en un tiempo computacional eficiente.

Con la finalidad de visualizar los resultados para función de densidad de distribución doble Gumbel bivariada, se grafica la frecuencia teórica contra las frecuencias calculadas a través de los distintos casos descritos anteriormente. A modo de comparación se grafica también la función identidad.

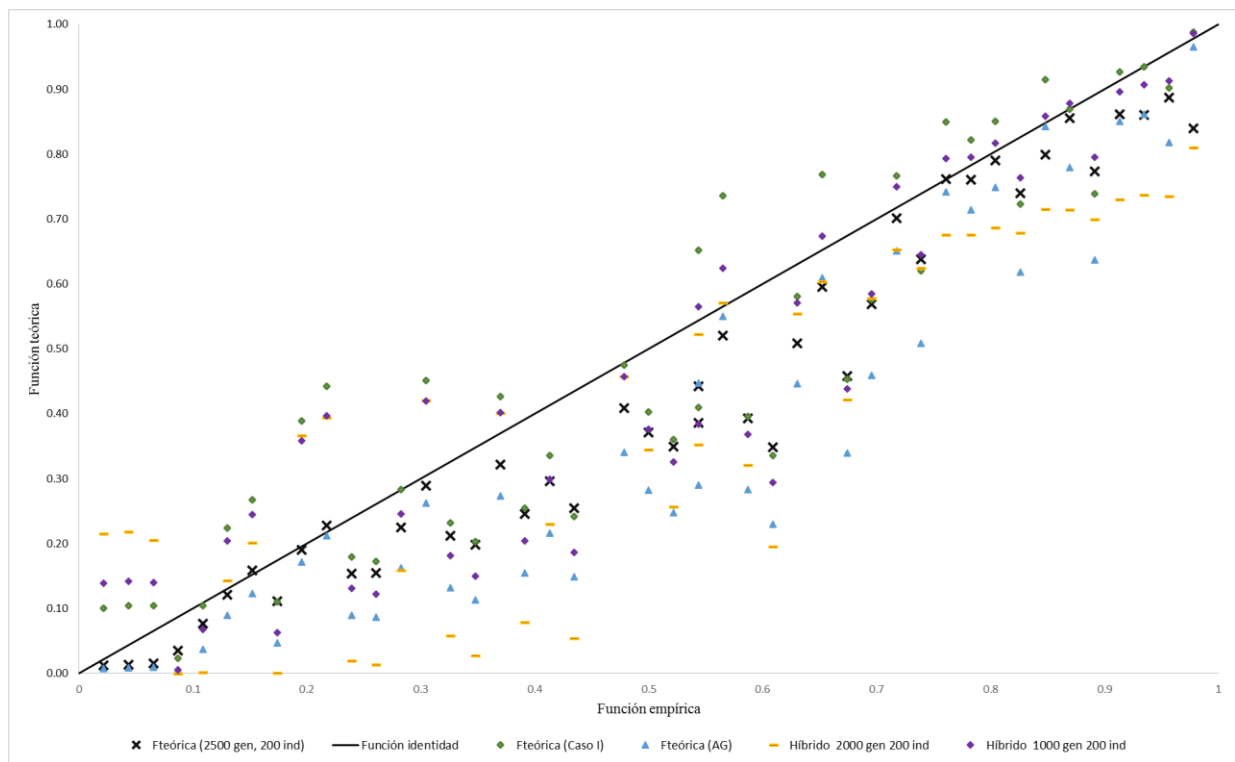


Figura 3-9. Gráfica de la función de densidad de distribución doble Gumbel Bivariada por AG y por el método híbrido, con el algoritmo N-M basado en Mathews y Fink (2003), de Herrera (2012).

3.2.8.1. Medición del error

Se calculan los estadísticos de los resultados, tanto para el cálculo de la FDDGB mediante AGs solamente y con el algoritmo híbrido; es decir para los tres casos de estudio y cinco combinaciones en total, en la Tabla 3-29 se observan los resultados.

Tabla 3-29. Estadísticos de los datos medidos y calculados.

	Fempírica k/(n+1)	Marginal FQ	Marginal FV	Fteórica (AX)	Marginal FQ	Marginal FV	Fteórica (2500 gen, 200 ind)	Marginal FQ	Marginal FV	Fteórica (Caso I)	Marginal FQ	Marginal FV	Fteórica (AG)	Marginal FQ	Marginal FV	F(2000 gen y 200 ind)	Marginal FQ	Marginal FV	Fteórica (1000 gen, 200 ind)
media	0.50193	0.49288	0.50667	0.36839	0.60615	0.49291	0.42077	0.50661	0.87583	0.48235	0.49236	0.99841	0.36924	0.54492	0.53338	0.38936	0.58490	0.58700	0.45703
desvest	0.28232	0.28251	0.28814	0.29130	0.22867	0.28392	0.27501	0.28298	0.11620	0.28206	0.27033	0.00157	0.28395	0.34556	0.15535	0.25734	0.31746	0.22021	0.28362
coef asim	-0.02108	0.16696	-0.10970	0.53489	-0.20061	0.04851	0.28140	0.08777	-0.44925	0.23747	0.08490	-1.59678	0.51580	-0.35713	-0.32558	-0.05554	-0.10445	-0.20713	0.28866
coef var	0.56247	0.57317	0.56870	0.79075	0.37725	0.57600	0.65358	0.55857	0.13267	0.58475	0.54904	0.00157	0.76902	0.63414	0.29125	0.66093	0.54276	0.37514	0.62058
Varianza	0.07971	0.07981	0.08303	0.08486	0.05229	0.08061	0.07563	0.08008	0.01350	0.07956	0.07308	0.00000	0.08063	0.11941	0.02413	0.06622	0.10078	0.04849	0.08044

Con la finalidad de determinar la dispersión de los datos entre la frecuencia medida y la calculada se obtuvo el error a través del coeficiente de determinación, mismo que funciona como un índice para evaluar qué momento o número de generaciones son las adecuadas para seguir con el proceso Nelder-Mead y obtener un resultado local óptimo en tiempo computacional corto; es decir, mientras más cercano a la unidad sea el coeficiente de determinación, mejor resultado presenta el resultado local óptimo.

Es importante mencionar que el coeficiente de determinación se calcula según la Ecuación (2.43) debido a que el modelo no es lineal y es un estadístico cuyo principal propósito es del probar hipótesis, en este caso, probar qué combinación hace posible mejorar el ajuste de la FDDGB a los datos de escurrimiento mediante el método híbrido . A continuación se describen los coeficientes de determinación para cada caso propuesto:

- ✓ El coeficiente de determinación calculado para los valores optimizados por Algoritmos Genéticos fue de 0.873.
- ✓ El coeficiente de determinación calculado para los valores optimizados por el método híbrido para 2 500 generaciones y 200 individuos en AG y 50 iteraciones en Nelder-Mead fue de 0.949.
- ✓ El coeficiente de determinación calculado para los valores optimizados por el método híbrido para 500 generaciones y 100 individuos en AG y 110 iteraciones en Nelder-Mead fue de 0.834.
- ✓ El coeficiente de determinación calculado para los valores optimizados por el método híbrido para 2000 generaciones y 200 individuos en AG y 50 iteraciones en Nelder-Mead fue de 0.727.

- ✓ Finalmente, el coeficiente de determinación calculado para los valores optimizados por el método híbrido para 1000 generaciones y 200 individuos en AG y 50 iteraciones en Nelder-Mead fue de 0.839.

3.2.8.2. En síntesis

A manera de resumen, en la siguiente tabla se presentan los resultados (Tabla 3-30):

Tabla 3-30. Síntesis de resultados para método híbrido aplicado a la función de distribución Doble Gumbel Bivariada.

Parametros del ajuste para la FDDGB								
No. Caso	1.00	2.00	3.00	4.00	5.00	6.00	7.00	
Parámetro	AG	Caso i	Caso ii	Caso iii				
				200 Ind, 2500 gen	100 Ind, 500 gen	200 Ind, 2000 gen	200 Ind, 1000 gen	
Q, m ³ /s	a ₁	-2,902.46	-2,484.29	-2,618.50	-2,779.50	-2,781.60	-2,779.60	-2,802.50
	c ₁	1,297.92	1,299.66	1,262.30	1,098.40	1,100.00	510.60	1,060.80
	a ₂	-1,863.43	3,938.72	-5,328.00	-7,197.40	-7,204.00	-7,201.40	-7,125.30
	c ₂	4,009.38	1,422.94	2,231.60	-1,256.20	1,920.80	1,811.60	1,043.10
V, m ³	a ₃	-2,438.75	-2,819.99	-1,107.50	-1,631.40	717.20	-1,633.40	-1,187.00
	c ₃	799.18	803.55	813.52	804.84	830.32	1,136.30	1,200.30
	a ₄	-4,207.59	-4,821.92	-3,571.10	-3,000.10	-1,949.00	-3,000.00	-2,083.50
	c ₄	2,149.19	2,437.20	1,879.20	808.16	170.34	-977.64	1,194.80
P _q	0.90	0.84	0.98	0.84	0.88	0.89	1.06	
P _v	0.88	0.88	0.93	0.82	0.82	0.82	0.85	
m	2.00	2.00	1.82	2.07	2.07	2.07	2.08	
Medición del tiempo								
TAG, s	178.41	39.32	101.96	103.54	25.02	87.69	60.82	
TNM, s	0.000	0.550	0.576	0.566	0.545	1.010	1.315	
Ttotal, s	178.41	39.87	102.53	104.11	25.57	88.70	62.13	
Medición del error								
R ²	0.873	0.778	0.945	0.949	0.834	0.727	0.839	

En este trabajo se probaron los parámetros de la función de distribución doble Gumbel bivariada optimizados con el método híbrido propuesto y se comparó su eficiencia en función del tiempo de cómputo requerido para la optimización de dichos parámetros y respecto al coeficiente de determinación entre las frecuencias medidas y las estimadas. El uso de 200 individuos en la primera parte del método híbrido (algoritmos genéticos) y un número mayor a las dos mil generaciones para la conformación de la matriz de parámetros iniciales en conjunto con un número mayor a 50 iteraciones para el proceso Nelder-Mead probaron ser más eficientes.

se observa que a medida que el número de iteraciones aumenta, el coeficiente de determinación disminuye, mientras que el tiempo empleado para optimizar los parámetros de ajuste de la función

de distribución doble Gumbel bivariada aumenta. En la Figura 3.10 se puede ver la variación del coeficiente de correlación según el caso de aplicación.

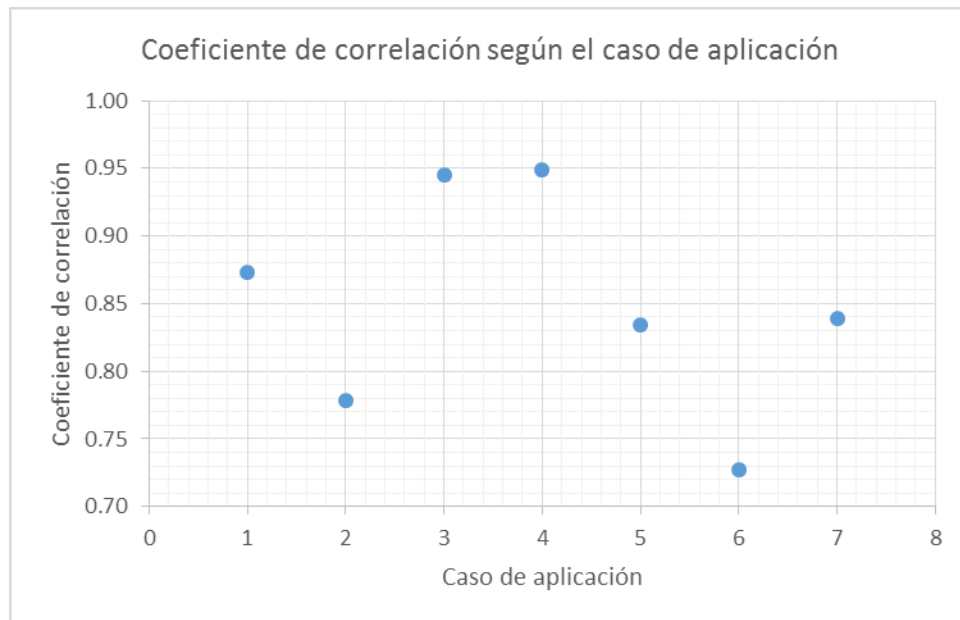


Figura 3-10. Coeficiente de determinación según el caso de aplicación

En la Figura 3-10 se observa que a las 5 000 iteraciones (primer caso de aplicación) el coeficiente de determinación disminuye; esto se debe a que es el tiempo computacional empleado sin el método híbrido; es decir, solo empleando AG (Tabla 3.30). Por otro lado el tiempo computacional empleado en los casos de 4 al 7 (ver Tabla 3.30), se encuentra sujeto a la combinación entre el número de individuos, de generaciones para AG y el número de iteraciones empleadas en el proceso con Nelder-Mead.

4. Conclusiones y recomendaciones

El algoritmo Nelder-Mead converge a un óptimo local, mientras que los algoritmos genéticos convergen a un óptimo global, siendo el método Nelder-Mead, más veloz ya que se trata de un método geométrico. Debido a esto fue necesario la implementación del modelo híbrido, el cual inició primero con algoritmos genético y después de un cierto tiempo óptimo se detuvo el programa y estos resultados arrojados fueron el punto de partida para el método Nelder-Mead, buscando con esto el ahorro de tiempos computacionales, debido principalmente a que los equipos de cómputo no siempre se encuentran en estado óptimo o no son los adecuados para hacer este tipo de procesamientos matemáticos.

El método N-M es sensible a los parámetros iniciales de búsqueda, mismos que se establecen con AGs. Un buen resultado en un tiempo computacional adecuado depende de las condiciones de operación para AG y NM; tanto en el número de individuos y número mínimo de generaciones para AG como del número de iteraciones en el proceso N-M. Por lo que el método ha mostrado ser iterativo y particular al caso de aplicación que se esté realizando.

En el ejercicio de calibración para el método N-M con la función exponencial se llegó a la solución exacta mediante la conformación de una matriz aleatoria como datos iniciales en un tiempo de 0.091 segundos y con un total de 100 iteraciones.

En cuanto al caso de aplicación uno, la función de distribución Gumbel se ajustó a un registro de precipitación correspondiente a la estación número 15008 "Atenco" dentro de la región hidrológica del Pánuco, sobre el río Tula. El ajuste de la función Gumbel mediante el método híbrido consistió en primero calcular la matriz de parámetros iniciales con AG en un tiempo computacional de 59.35 s para 200 individuos y 1500 generaciones, posteriormente, con el método de N-M se llegó a la solución para los parámetros de ajuste en 0.46 s y 70 iteraciones, siendo dichos valores de $\alpha = 0.1058$ y $\beta = 10.67288$ en un tiempo total de un minuto. En cuanto a la medición del error para esta caso de aplicación, el valor de la función objetivo, por máxima verosimilitud se reportó en -1.962×10^3 . A modo de comparación también se ajustaron los datos empleando solamente AG para 5 000 generaciones y 200 individuos, permitiendo calcular un coeficiente de determinación para AG de 0.9963 y para el método híbrido de 0.9961.

Para el segundo caso de aplicación, correspondiente a la función de distribución doble Gumbel Bivariada ajustada a datos de escurrimiento en la Presa Infiernillo, en estado de Michoacán considerando los gastos medios diarios por cuenca propia reportados en la suma de las estaciones hidrométricas: Los Pinzones, Panches, La Pastoría y Caimanera, la mejor solución se llevó a cabo de la siguiente manera: para la conformación de la matriz de parámetros iniciales (AG) para 200 individuos y 2500 generaciones en un tiempo de 158.48 s y para el proceso de Nelder-Mead para 60 iteraciones en 0.57 s, sumando un tiempo total del método híbrido correspondiente de 159 s y un valor reportado de la función objetivo (máxima verosimilitud) de 16.4. A modo de comparación, el problema propuesto también se calculó con AG y AX, que son métodos ya conocidos, reportando un valor de coeficiente de determinación de 0.949, 0.873 y 0.865, respectivamente. Los valores por estos tres métodos se reportan en la siguiente tabla.

Parámetro	AG	AX	200 Ind, 2500 gen	
Q, m³/s	a₁	-2,902.46	-2,902.46	-2,779.50
	c₁	1,297.92	1,234.57	1,098.40
	a₂	-1,863.43	-9,069.39	-7,197.40
	c₂	4,009.38	4,000.00	-1,256.20
V, m³	a₃	-2,438.75	-1,708.30	-1,631.40
	c₃	799.18	785.55	804.84
	a₄	-4,207.59	-4,264.73	-3,000.10
	c₄	2,149.19	2,109.70	808.16
P_q	0.90	0.92	0.84	
P_v	0.88	0.93	0.82	
m	2.00	2.19	2.07	
Medición del tiempo				
TAG, s	178.41		103.54	
TNM, s	0.000		0.566	
Ttotal, s	178.41		104.11	
Medición del error				
R2	0.873	0.865	0.949	
Niteraciones AG	50000.000		2500.000	
iteraciones N-M			60.000	

Figura 4-1. Síntesis de resultados para método híbrido aplicado a la función de distribución Doble Gumbel Bivariada

Como se puede observar en la Figura 2, de resultados, las diferencias entre realizar la optimización de los parámetros de la FDDGB con el algoritmo simple AG y el modelo híbrido no son grandes; sin embargo, el tiempo computacional prácticamente se reduce a la mitad o inclusive menos.

Por último, en cuanto al número de iteraciones para Nelder-Mead, dependen de qué tanto se aproxime el simplex que se esté generando de la solución óptima.

5. Referencias

- Abdulla, F. A., & Lettenmainer, D. P. (30 de Agosto de 1996). Development of regional parameter estimation equations for a macroscale hydrologic model. *ELSEVIER. Journal of Hydrology*, 197, 230-257. doi:PII S0022-1694(96)03262-3
- Aldama, A. A., & Ramírez, A. I. (2005). Aspectos de hidrología de presas: diseño y operación. En G. A. Paz Soldán, H. Marengo, & F. I. Arreguín, . *Las presas y el hombre*, 65-76.
- Arganis Juárez, M. L. (2004). *Operación óptima de un sistema de presas en cascada para generación hidroeléctrica tomando en cuenta condiciones reales de operación y el uso de muestras sintéticas para el pronóstico*. México, D.F.: Tesis doctoral. Universidad Nacional Autónoma de México.
- Arganis Juárez, M. L., Domínguez Mora, R., Rodríguez, V., & Preciado, J. (2014). Estimación de la función de distribución bivariada de avenidas históricas máximas anuales.
- Arroyo Apaza, V. M. (2003). *Modelo de un algoritmo genético con selección discriminatoria de individuos bajo un esquema de ponderación de probabilidades de mutación*. (U. c. Pablo, Ed.) Arequipa, Perú: Tesis para obtener el título de ingeniero en informática.
- Barati, R. (Noviembre de 2011). Parameter estimation of nonlinear Muskingum models using Nelder-Mead simplex algorithm. *JOURNAL OF HYDROLOGIC ENGINEERING @ ASCE*, 16(11), 946-954. doi:10.1061/(ASCE)HE.1943-5584.0000379.
- Bonilla Petricolet, A., Lira Padilla, M. G., & Soto Becerra, C. A. (Octubre de 2005). Aplicación del método de optimización de recocido simulado en la regresión de isotermas de adsorción. *Rev. Int. Contam. Ambiental*, 21(4), 201-206. Obtenido de <http://www.revistascca.unam.mx/rica/index.php/rica/article/viewFile/22530/21671>
- Campos Aranda, D. F. (1989). Estimación de los parámetros óptimos de la distribución Gumbel mixta por medio del algoritmo de Rosenbrock. *Ingeniería Hidráulica en México*.
- Chamizo, J. A. (2010). Una tipología de los modelos para la enseñanza de las ciencias. *Revista Eureka sobre Enseñanza y Divulgación de las Ciencias*, 7(1), 26-41.
- Chelouah, R., & Siarry, P. (2003). Genetic and Nelder-Mead algorithms hybridized for a more accurate global optimization of continuous multimodal functions. *ELSEVIER*, 148, 335-348. doi:10.1016/S0377-2217(02)00401-0
- Chow, V., Maidment, D. R., & Mays, L. W. (1994). *Hidrología aplicada*. Santafé de Bogotá: McGraw-Hill Interamericana, S.A.
- CONAGUA. (2017). *Banco Nacional de datos de aguas Superficiales (BANDAS)*. Obtenido de Subdirección general Técnica: <http://www.conagua.gob.mx/CONAGUA07/Contenido/Documentos/Portada%20BANDAS.htm>
- Darwin, C. (1859). *El origen de las especies*. <http://www.cervantesvirtual.com/>.

- Domínguez Mora, R., Fuentes Mariles, Ó., Arganis Juárez, M. L., & Mendoza R., A. (2009). Calculation of bivariate double gumbel probability density function via a genetic algorithm: application to Huites Dam Basin. *Journal of Flood Engineering*, 1(1), 41-54.
- Dominique, C., & Fuamba, M. (Octubre de 2013). Discussion of "Detection of Flood-Prone areas using digital elevation models" by salvatore Manfreda, Margherita Di Leo, and Aurelia Sole. *JOURNAL OF HYDROLOGIC ENGINEERING © ASCE*, 6(10), 781-790. doi:10.1061/(ASCE)HE.1943-5584.0000367
- Dueñas Puebla, D. (2016). *Estimación de las avenidas de diseño de la Presa Infiernillo, Michoacán. A partir de métodos estadísticos y espectrales.* (Tesis para obtener el grado de maestría ed.). Ciudad de México: UNAM, Facultad de ingeniería.
- Durand, N., & Alliot, J.-M. (1999). A combined Nelder-Mead simplex and Genetic Algorithm.
- Escalante Sandoval, C. (2007). Application of biavariate extreme value distribution to flood frequency analysis: a case study of Northwstern Mexico. *Nat Hazards*, 42, 37-46.
- Escalante Sandoval, C. A., & Reyes, C. L. (2008). *Técnicas Estadísticas en Hidrología.* CDMX: Departamento de Ingeniería Hidráulica. UNAM, Facultad de Ingeniería.
- Fan, S.-K., Liang, Y.-C., & Zahara , E. (2006). A genetic algorithm and a particle swarm optimizier hybridized with Nelder-Mead simplex search. *ELSEVIER. ScienceDirect*, 50, 401-425. doi:10.1016/j.cie.2005.01.022
- Fuentes Mariles, Ó. A., Morales Rodríguez, H. L., & Arganis Juárez, M. L. (2016). COMPARACIÓN DE DOS MÉTODOS PARA OBTENER HIDROGRAMAS PARA EL CÁLCULO DE INUNDACIONES PROLONGADAS. *Memorias de la Décima Quinta Conferencia Iberoamericana en Sistemas, Cibernética e Informática (CISCI 2016).*
- Goldberg, D. E. (1999). *Genetic Algorithms in Search, Optimization and Machine Learning* (20th ed.). Massachusetts: Addison-Wesley.
- Gómez Gordillo, F. (2015).). *Actuzalización de las avenidas de diseño para la operación de los vertedores de la presa El Infiernillo.* (UNAM, Ed.) CDMX: Tesis para obtener el grado de maestría.
- Gómez, J. F., Aparicio, J., & Patiño, C. (2010).). *Manual de análisis de frecuencias en hidrología* (Primera ed.). Morelos, México: Instituto Mexicano de Tecnología del Agua.
- Gómez, J. F., Aparicio, J., & Patiño, C. (2010). *Manual de análisis de Frecuencias en Hidrología.* Jiutepec, Morelos: Instituto Mexicano de Tecnología del Agua.
- Haan, C. (1977). *Statistical methods in Hydrology.* Iowa: Iowa Sate University press, Ames.
- Herrera Alanís, J. L., Arganis Juárez, L. M., & Domínguez Mora, R. (Agosto de 2014). Estimación de parámetros óptimos para funciones de distribución de dos poblaciones. (U. Instituto de Ingeniería, Ed.) *XXVI Congreso latinoamericano de hidráulica.*

- Herrera Alanís, J. L., Domínguez Mora, R., & Arganis Juárez, L. M. (Noviembre de 2012). Estimación de parámetros óptimos de la distribución doble Gumbel usando el Método Nelder-Mead (método Simplex). (U. Instituto de Ingeniería, Ed.) *XXII Congreso Nacional de Hidráulica*.
- IIMAS-UNAM. (2005). *Introducción a los algoritmos genéticos. Taller de introducción a los Algoritmos Genéticos*. CDMX: Grupo de Computación bioinspirada.
- Jain, A., & Srinivasulu, S. (Abril de 22 de 2004). Development of effective rainfall-runoff models using integration of deterministic, real-coded genetic algorithms and artificial neural network techniques. *WATER RESOURCES RESEARCH*, 40, 1-12. doi:10.1029/2003WR002355
- Kite, G. W. (1988). *Frequency and risk analyses in hydrology*. Collins: Water Resources Publications, Colo. St. University Fort Collins.
- Mathews, J. H., & Fink, K. D. (2003). *Métodos numéricos con Matlab* (Tercera edición ed.). Madrid: Prentice Hall.
- Mathews, H. J., & Kurtis, D. F. (2000). *Métodos Numéricos con Matlab*. (T. Edición, Ed.) Madrid: Prentice Hall.
- Menchaca Méndez, A. (2008). *Algoritmo híbrido para resolver problemas de optimización con restricciones*. (C. d. Computación, Ed.) México D.F.: Tesis Doctoral.
- Nelder, J., & Mead, R. (01 de January de 1965). A simplex Method for function minimization. *The computer Journal*, 8(1)(27), 308-313. doi:https://doi.org/10.1093/comjnl/7.4.308
- Nicklow, J., Reed, P., Savic, D., Dessalegne, T., Harrell, L., Chan-Hilton, A., . . . Zechman, E. (Julio-Agosto de 2010). State of the art for Genetic Algorithms and beyond in water resources planning and management. *JOURNAL OF WATER RESOURCES PLANNING AND MANAGEMENT*, 136(4), 412-432.
- Pérez López, J. R. (2005). *Contribución a los métodos de optimización basados en procesos naturales y su aplicación a la medida de antes en campo próximo*. Santander, España: Tesis doctoral. Universidad de Cantabria. Departamento de Ingeniería de comunicaciones.
- Quintana Hernández, P. A., Villalobos Oliver, E. B., & Cornejo Serrano, M. C. (2005). *Métodos numéricos con aplicaciones en Excel* (Primera ed.). Barcelona, España: Reverté S.A.
- Ramírez Miranda, J. C. (2016). *Simulación con registros sintéticos de la política de operación de la presa "El Infiernillo Mich." con propósito de generación de electricidad*. (UNAM, Ed.) CDMX: Tesis de grado de licenciatura.
- Ramírez, A. I. (Abril-Junio de 2010). La seguridad de las presas desde la perspectiva hidrológica. *Red de Revistas Científicas de América Latina, el Caribe, España y Portugal*, 2, 157-173. Obtenido de <http://www.redalyc.org/articulo.oa?id=353531973010>
- Ramírez, A. I., & Aldama, A. A. (2000). Análisis de frecuencias conjunto para la estimación de avenidas de diseño. *Asociación Mexicana de Hidráulica*.

- Ratón Almansa, M. (2002). Optimización Continua Basada en Algoritmos de Estimación de Regresión.
- Restrepo Londoño, J. (2016).). *ESTUDIO HIDROLÓGICO Y DE SEDIMENTOS FINOS PARA EL PROYECTO HIDROELÉCTRICO LAS CRUCES*. (UNAM, Ed.) CDMX: Tesis para otorgar el grado de maestría en ingeniería.
- Rivas Carpio, J. J. (2016). *Simulación continua del comportamiento de dos vasos que trabajan en paralelo. Caso del siemte de presas Temascal-Cerro de Oro*. CDMX: Tesis para otorgar el grado de maestría en ingeniería. UNAM.
- Rosenbrock, H. H. (Julio de 2012). An Automatic Method for finding the Greatest or Least Value of a Function. *Oxford Journals*, 175-184.
- Saavedra Barrera, P. (2012). *Introducción a la optimización no lineal*. CDMX.: Universidad Autónoma Metropolitana, Departamento de matemáticas.
- Saavedra Barrera, P. (2012). *Introducción a la optimización no lineal*. (U. A. Metropolitana, Ed.) Iztapalapa, México: Departamento de Matemáticas.
- Secretaría de medio ambiente y recursos naturales. (2007). *Acuerdo por el que se dan a conocer los estudios técnicos de aguas nacionales superficiales de la Región Hidrológica número 18*. Obtenido de http://dof.gob.mx/nota_detalle_popup.php?codigo=5175730
- Wang, Q. J. (Septiembre de 1991). The Genetic Algorithm and its application to calibrating conceptual rainfall-runoff models. *WATER RESOURCES RESEARCH*, 27(9), 2467-2471.
- Zahara, E., & Kao, Y.-T. (2009). Hybrid Nelder-Mead simplex search and particle swarm optimization for constrained engineering design problems. *ELSEVIER Expert Systems with Applications*, 36, 3880-3886. doi:10.1016/j.eswa.2008.02.03
- Zhang , X., Srinivasan, R., & Bosch, D. (2009). Calibration and uncertainty analysis of the SWAT model using Genetic Algorithms and Bayesian Model Averaging. *Journal of Hydrology*, 374, 307-317.