CAPÍTULO III

TÉCNICAS DE SIMULACIÓN ESTADÍSTICA

3.1 Introducción

Los datos sintéticos son elementos de suma importancia en los sistemas de diseño en presas de almacenamiento, ya que se evalúa el propósito del sistema con sumo cuidado y con mayor precisión estadística. Para la generación de este tipo de datos es importante tomar en cuenta las propiedades estadísticas que en un principio los registros históricos deben tener, así como los parámetros calculados a partir de los ya mencionados registros. Debido a que la simulación hidrológica de estos datos es una descripción matemática de la respuesta hidrológica a una serie de eventos durante un periodo de tiempo seleccionado (Hoggan, 1989). Esto quiere decir que se pueden calcular simulaciones tanto diarias, como mensuales o anuales de una corriente superficial, basándose en los registros de lluvias o en la entrada de volumen por ríos cercanos a la presa.

A lo largo de este capítulo se mencionarán las técnicas que se usarán para poder generar las series de valores antes mencionadas, estos datos son la herramienta esencial para poner en marcha el algoritmo de simulación hidrológica y con el cual se estimará la capacidad útil de la presa de almacenamiento y su confiabilidad.

3.2 Método de simulación

El modelo de simulación hidrológica es la herramienta con la cual se generarán gastos a partir de registros históricos, lo cual proveerá de un número limitado de secuencias de datos sintéticos, todas con la misma probabilidad de que sucedan en un momento dado.

Para el buen funcionamiento de estas secuencias es importante el tipo de modelo que se utilice, para este caso en particular se empleará un modelo basado en las técnicas de generación de Markov, el cual indica que la síntesis de la secuencia de los gastos aleatorios generados no ignora la existencia de persistencias, es decir, que se presente una tendencia de que los gastos grandes sean seguidos por gastos grandes y que los gastos pequeños sean seguidos por gastos pequeños. Este tipo de modelo supone que cada gasto es dependiente sólo en uno o más de los gastos recientes (Fiering y Jackson, 1971), además de que los gastos deben tener una distribución normal.

El modelo mencionado, muestra los gastos como una serie de valores dependientes, en la cual cada valor tiene una parte definida y una parte aleatoria del gasto sintético. Esta generación de valores considera que los gastos son el resultado de un proceso aleatorio el cual va cambiando con el tiempo y que a la vez está envuelto en la probabilidad. Para la generación de simulaciones anuales de variables normales, se puede definir con la siguiente ecuación:

$$Q_{i+1} = Q_{j+1} + r_j \left(Q_i - \overline{Q}_j \right) + t_i \sigma_{j+1} \left(1 - r_j^2 \right)^{1/2} \qquad (3.1)$$

En donde:

Q_{i+1} : Gasto generado en el periodo i+1

Q_{j+1} : Media de los gastos observados en el periodo j + 1

r_j : Coeficiente de correlación para la relación de gastos en los periodos

Q_i : Gasto generado en un periodo i

$\overline{Q}_i$: Media observada en los gastos en el periodo j

t_i : Número aleatorio seleccionado de una distribución normal, teniendo como media cero y una varianza unitaria.

σ_{j+1} : Desviación estándar de los gastos observados para un periodo j + 1

Fiering y Jackson (1971) modificaron la ecuación anterior por la siguiente:

$$Q_{i+1} = \overline{Q} + r_1\left(Q_i - \overline{Q}\right) + t_i s\left(1 - r_j^2\right)^{1/2} \qquad (3.2)$$

En donde:

Q_{i+1} : Gasto generado en el periodo i + 1

Q_i : Gasto generado en el periodo i

$\overline{Q}$: Media anual de los gastos del registro histórico

r_1 : Coeficiente de autocorrelación de gastos anuales

t_i : Número aleatorio seleccionado de una distribución normal, teniendo como media cero y una varianza unitaria.

s : Desviación estándar de los gastos anuales

Las dos ecuaciones anteriores (3.1 y 3.2) son para gastos anuales, para gastos mensuales se desarrolla con la siguiente ecuación:

$$Q_{i,j} = \overline{Q}_j + b_j\left(Q_{,j-1i} - \overline{Q}_{j-1}\right) + t_i\sigma_j\left(1 - r_{(j)}^2\right)^{1/2} \qquad (3.3)$$

Varios de los elementos que son parte de esta ecuación ya fueron definidos, los elementos que aún no han sido definidos y son parte de la ecuación 3.3 se mencionarán a continuación:

$Q_{i,j}$: El gasto generado en el intervalo i, el cual ocurre en un periodo j

b_j : El producto del coeficiente de correlación multiplicado por la relación de las desviaciones estándar en los periodos j y j -1

$Q_{i,j-1}$: El gasto generado en el intervalo en un periodo j - 1

$\overline{Q}_{j-1}$: La media de los gastos observador en el periodo j – 1

σ_j : Desviación estándar en un periodo j

Al igual que en la generación de gastos anuales, en la simulación mensual se tienen algunas modificaciones, siendo las variables normales trambién.

$$Q_{i,j} = \overline{Q}_j + b_j\left(Q_{i,j-1} - \overline{Q}_{j-1}\right) + t_i s_j \left(1 - r_{(j)}^2\right)^{1/2} \tag{3.4}$$

$$b_j = r_{(j)}\left(\frac{s_j}{s_{j-1}}\right) \tag{3.5}$$

En donde:

$Q_{i,j}$: El gasto generado en el intervalo i, el cual ocurre en un periodo j

$\overline{Q_j}$: Media de los gastos para el periodo j

b_j : El producto del coeficiente de correlación multiplicado por la relación de la desviación estándar en los periodos j y j - 1

$r_{(j)}$: El coeficiente de correlación entre los gastos de los periodos j y j-1

$Q_{i,j-1}$: El gasto generado en el intervalo i en un periodo j-1

s_j, s_{j-1} : Desviación estándar de los gastos durante los periodos j y j-1

$\overline{Q}_{j-1}$: La media de los gastos observados en el periodo j-1

En donde $Q_{1,1} = \overline{Q}_j$ para inicializar en generación de simulaciones mensuales, de la misma manera $Q_1 = \overline{Q}_1$ en simulaciones anuales.

El uso de números aleatorios representa la parte probabilística de la ecuación mencionada anteriormente. Estos números deben pertenecer a una distribución uniforme en el intervalo [0,1]. Estos, números aleatorios deben transformarse a números aleatorios normales con varianza unitaria y media con valor cero.

En ocasiones, este modelo puede arrojar valores negativos, pero si eso pasara, entonces los gastos generados negativos se tomarán con valor de cero. El modelo es del tipo estocástico, por que se basa en correlaciones seriadas, es decir, cada valor depende del anterior uno a uno (Hoggan, 1989). Otra característica importante es que intervienen variables aleatorias, por lo tanto pueden ser generadas y analizadas el número de trayectorias que se quieran. En este caso en particular, se generarán mil trayectorias con registros variables, es decir, con tamaños de muestras distintos.

Sin embargo, cabe destacar que estos registros sintéticos no equivalen a crear registros futuros basándose en los registros históricos. Aunque los primeros tengan características similares a los segundos, esto se realiza con la finalidad de crear una herramienta que ayude a visualizar el estudio bajo diferentes posibilidades de registros pero basados en patrones ya establecidos (registros históricos).

3.3 Transformación Box-Cox

Para poder utilizar correctamente el modelo de simulación, es necesario que la muestra o la serie de registros sintéticos tengan distribución normal. Si esto no sucediera, este método arrojaría valores erróneos con respecto a los registros históricos. Es por ello que antes de generar los registros históricos se utilizará la transformación de Box – Cox para normalizar las muestras. La transformación de Box - Cox utiliza las siguientes relaciones:

$$Y_i = \frac{X_i^{\lambda} - 1}{\lambda} \qquad \text{para} \quad \lambda \neq 0 \qquad (3.6)$$

y:

$$Yi = \ln(Xi) \qquad \text{para} \quad \lambda = 0 \qquad (3.7)$$

Donde:

Yi : Volumen transformado

Xi : Volumen original sesgado

λ : Parámetro tal que Yi tenga un coeficiente de asimetría estadísticamente igual a cero

El valor de λ podrá ser determinado por prueba y error de tal manera que el coeficiente de asimetría en la muestra sea estadísticamente igual a cero (McMahon y Míen,1986).

3.4 Generación Monte Carlo

El método de Monte Carlo es muy usado es los lenguajes de programación ya que se usa para hallar la probabilidad de un suceso y tiene como objetivo el generar números aleatorios que posean una distribución uniforme en el intervalo (0,1), es decir, que no cualquier número generado por la computadora podrá ser utilizado en el caso específico de este análisis. Lo anterior se realiza bajo el empleo de una técnica sencilla que ayude a modificar los números aleatorios uniformes en el intervalo (0,1), creados por la máquina, a números aleatorios con distribución normal N(0,1) a través de la transformación asintótica, dada por una aproximación polinómica establecida por Abramowitz y Stegun (1965).

Por ejemplo, se tienen los siguientes valores constantes:

$a_0 = 2.36542789$	$b_1 = 1.245368978$
$a_1 = 0.90136578$	$b_2 = 0.219245712$
$a_3 = 0.03451578$	$b_3 = 0.006971325$

Después se identifica a F(z) como un número aleatorio uniforme en el intervalo (0,1), entonces se revisa el rango en el cual se encuentra F(z). Si $0 < F(z) \leq 0.5$, entonces se realiza la siguiente operación

$$v = [-2 \ln F(z)]^{½} \tag{3.8}$$

y

$$z = v - \left[\frac{a_0 + a_1 v + a_2 v^2}{1 + b_1 v + b_2 v + b_3 v^3}\right] \tag{3.9}$$

y $z = -z$ por lo tanto z esta dentro de N(0,1).

En el caso en que el número aleatorio uniforme se ubicara fuera del rango siguiente: 0.5 < F(z) ≤ 1.

Por lo tanto: $$F(z) = 1 - F(z) \tag{3.10}$$

donde z se calcula con las ecuaciones (3.8) y (3.9).

3.5 Parámetros estadísticos

Para la generación de datos sintéticos es imprescindible la utilización de parámetros estadísticos, los cuales son la esencia de la parte definida que conforma la ecuación de Fiering y Jackson (1971). Estos parámetros son calculados con base en los registros históricos y serán descritos brevemente a continuación, así como sus fórmulas correspondientes. Estos parámetros son de suma importancia, ya que las muestras sintéticas que se generen deben de tener una similitud estadística con los registros históricos.

3.5.1 Media

Se llama así a la suma de todos los valores dividida por el número total de los mismos. Entonces para una serie de datos registrados en un periodo de años n, la media $\overline{x}$ de los gastos x_i, se expresa con la siguiente ecuación:

Media anual

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3.11}$$

Media periódica

$$\bar{x}_j = \frac{1}{p}\sum_{k=1}^{p} x_{k,j} \qquad (3.12)$$

Donde j indica el mes sobre el cual se calcula y p expresa el número de años.

3.5.2 Varianza

Una característica importante que poseen los datos históricos es su variabilidad o cómo es que ellos se distribuyen, y ello puede medirse por medio del cálculo de la varianza. La varianza se representa por s^2 y se calcula de la siguiente manera:

Varianza anual:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \qquad (3.13)$$

Varianza poblacional:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N} \qquad (3.14)$$

Varianza periódica:

$$s_j^2 = \frac{1}{p}\sum_{k=1}^{p} x_{k,j}{}^2 - \frac{1}{p(p-1)}\left(\sum_{k=j}^{p} x_{k,j}\right)^2 \qquad (3.15)$$

Tanto en la varianza anual como periódica, el subíndice j indica el mes sobre el cual se calcula el parámetro, la letra p son el número de años.

3.5.3 Coeficiente de asimetría

El coeficiente de asimetría mide el grado de simetría de la muestra alrededor de la media. La media de la población es el centro de la muestra. En caso de que exista o no la simetría en una distribución se le llama asimetría y la definición más conocida es la siguiente (McMahon y Mein,1986):

$$C_s = \frac{a}{s^3} \tag{3.17}$$

$$a = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(x_i - \overline{x}\right)^3 \tag{3.18}$$

Donde:

n : Número de años del registro

x_i : Volumen del registro histórico en el periodo i

$\overline{x}$: Media del registro histórico

s^3 : Varianza al cubo

Los datos con asimetría positiva están sesgados hacia la derecha y los datos con asimetría negativa están sesgados hacia la izquierda. El grado de sesgado baja generalmente cuando el número de datos se eleva. Por lo tanto, la distribución de flujos anuales son generalmente menos sesgados que las distribuciones de flujos mensuales, que pertenecen al mismo río.