

MODELACION HIDROLOGICA ESTOCASTICA:
DESARROLLO DE UN MODELO DE GENERACION
SINTETICA DE SERIES TEMPORALES

POR:

Francisco Pablo García Gutiérrez.

Ingeniero Civil (Honores). Univ. de Lancaster (Inglaterra),
Master of Science en Ing. Hidráulica. Univ. de Newcastle upon Tyne (Inglaterra)
Doctor Ingeniero de Caminos, Canales y Puertos. Univ. de Valencia (España)

Académico de Número de la Academia Nacional de Ciencias de Bolivia.

Catedrático de Hidrología e Hidráulica e Investigador.

Universidad Autónoma Gabriel Rene Moreno.

Santa Cruz – Bolivia, Julio del 2010

DIRECCIÓN

: PO BOX 411

Teléfono: (591-3)-3583647

e-mail: franciscogarcia@cotas.com.bo

SANTA CRUZ - BOLIVIA

MODELACION HIDROLOGICA ESTOCASTICA: DESARROLLO DE UN MODELO DE GENERACION SINTETICA DE SERIES TEMPORALES

SUMARIO

En el presente trabajo se presenta un modelo de generación de series sintéticas temporales de variables hidrometeorológicas de aplicación a eventos de carácter mas bien hidrológicos o consecuencia de estos – como ser los caudales de un rio que se generan en una cuenca hidrográfica-, basado en principios de la teoría de probabilidades, lo que se llama un proceso estocástico o aleatorio aplicado a procesos hidrológicos. El modelo que se presenta es de carácter general en la literatura especializada.

Se ha desarrollado un programa de computación que facilita el cálculo de las variables que intervienen en el proceso y, se presenta una alternativa en la generación de las variables aleatorias (generación de números aleatorios) que se requiere en la fase de computo para garantizar una periodicidad alta, haciendo uso de las facilidades del programa visual basic 6.0.

El modelo desarrollado se ha aplicado en la determinación de la serie temporal de caudales mensuales y anuales para el rio Grande (Santa Cruz-Bolivia).

SUMMARY

A model for the synthetic generation of time series related to hidro-meteorological variables is presented, based upon the generally accepted theory of probability applied to hidrological processes, which are termed stochastic processes, and more specifically those related to hidrological events or those which are a consequence of these; such as the discharge of a river generated within its watershed.

A computer programme was developed in order to facilitate the calculation of the variables involved in the processes and an alternative method is proposed to generate random numbers which are required in the procedure. A facility available in Visual basic 6.0 has been used in order to guaranty a high periodicity on the generation of random numbers.

The developed methodology has been applied for the generation of the discharge series of the Rio Grande in Santa Cruz – Bolivia.

Palabras clave: Hidrología estocástica, Procesos estocásticos, Generación de números aleatorios, Distribuciones Normales.

MODELACION HIDROLOGICA ESTOCASTICA: DESARROLLO DE UN MODELO DE GENERACION SINTETICA DE SERIES TEMPORALES

HIDROLOGIA ESTOCASTICA Y LA GENERACION SINTETICA DE CAUDALES.-

1.- Introducción a la naturaleza estocástica o aleatoria de eventos hidrometeorológicos.-

Todos los fenómenos relacionados a la generación y distribución de la lluvia en nuestro planeta, y todas sus consecuencias directas como ser: el escurrimiento superficial por las cuencas, la infiltración a las porciones más profundas del suelo y de manera especial el caudal de los ríos -como lugar donde llegan las aguas dentro del ciclo hidrológico para finalmente ir a los océanos-, son procesos aleatorios, es decir que tienen un cierto nivel de incertidumbre o una probabilidad de que sean igualados o excedidos en un número de años, y por lo tanto están gobernados por leyes de la probabilidad. A los procesos gobernados por probabilidades se les llama **estocásticos**.

De manera que si como resultado de un proceso matemático que ha sido establecido siguiendo principios de la física clásica para explicar un fenómeno natural, se obtiene un valor numérico, este será y tiene carácter **determinístico** ya que ha seguido un proceso de ese tipo, pero se puede decir que **es esencialmente estocástico**, ya que el evento es de esa naturaleza.

Cualquier proceso de análisis hidrológico relacionado a un lugar determinado o una región geográfica supone el tener una serie de datos relacionados a la variable de interés – precipitación, evaporación, caudales, etc.,- para poder emprender estudios tendentes al entendimiento y explicación del fenómeno, por una parte y, a cuantificar los elementos que hacen al mismo y a los se desprenden de él, por la otra.

En muchas ocasiones los datos disponibles de precipitación en lo que supone series temporales que cubran un periodo de tiempo adecuado (y además que estos sean fiables) y la imposibilidad de obtener los mismos a través de modelación hidrológica determinística (Ver apartado 2 para una definición), que requiere de datos tanto de carácter fisiográfico como hidrometeorológicos en cuencas hidrográficas de extensión considerable y que presentan una red de drenaje compleja e importante dentro de las áreas de aporte, es que se ha estimado de gran utilidad y valor el utilizar un modelo de generación estocástica.

Existe muy poca tradición en Bolivia en el uso de modelación estocástica para la generación sintética de series de caudales y otras variables de carácter hidrometeorológicas. Sin embargo, esta técnica es utilizada en muchos países con resultados verdaderamente alentadores, no solo por los resultados en si, sino porque al estar intrínsecamente inmersos en el proceso las variables que definen el proceso físico, quedan muy pocos argumentos para rebatirlos.

MODELACION HIDROLOGICA ESTOCASTICA: DESARROLLO DE UN MODELO DE GENERACION SINTETICA DE SERIES TEMPORALES

Aunque el sentido de la palabra estocástica, puede dar a entender que se trata de un proceso puramente aleatorio; en hidrología, este no es el caso, ya que puede decirse que los procesos relacionados a las series temporales que se desarrollan con esta técnica son parcialmente aleatorios. Esto es, relacionan una parte determinística y otra probabilística (Ver Gráfica 2).

En hidrología determinística la variable temporal se asume que está explicada totalmente por otras variables que se desarrollan siguiendo un método adecuado, basado generalmente en leyes físicas de la naturaleza. **En hidrología probabilística, el interés no está centrado con la secuencia temporal, pero si en la probabilidad que de un evento sea igualado o excedido, es decir que está basado en la ley de las probabilidades o chances.** En hidrología estocástica, la secuencia temporal es todo lo que interesa. La representación estocástica preserva las propiedades de ocurrencia asociada con la secuencia de los eventos.

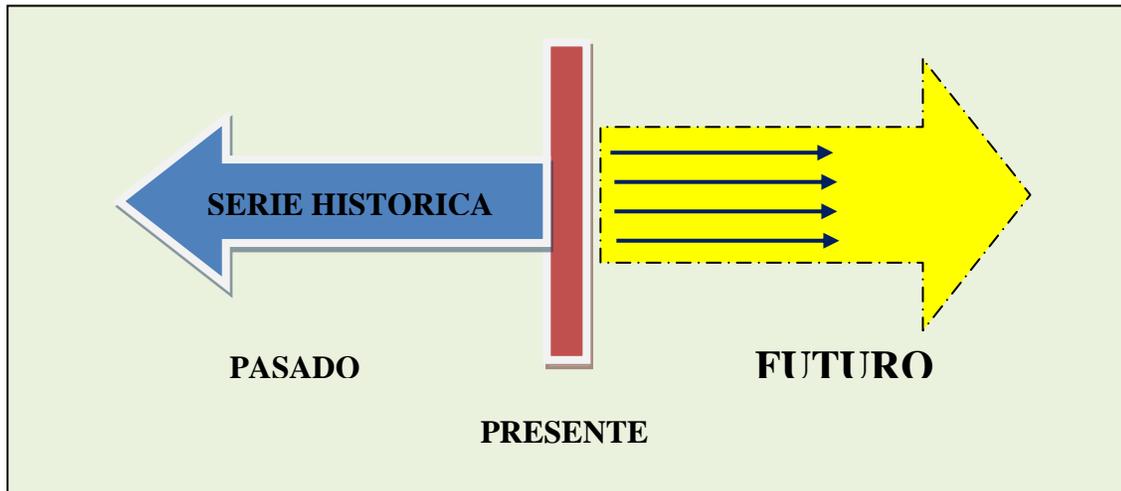
El problema perenne del hidrólogo es la insuficiencia de datos, sean de precipitación o, de manera más común, de caudales. En la mayoría de los casos, se asume –con mucha propiedad- que el futuro es estadísticamente similar al pasado. Este elemento es el que da sustento a la hidrología estocástica.

La esencia básica para el análisis estocástico es que el proceso sea estacionario, es decir, que las propiedades estadísticas del proceso no varían en el tiempo. Así, las propiedades de los registros históricos se pueden utilizar para derivar series sintéticas largas.

Estas series deben mostrar propiedades estadísticas similares a aquéllas de la **serie histórica**, es decir de la serie que se conoce y, a partir de la cual se desea generar valores posteriores en el tiempo. Algunas propiedades de las series temporales hidrológicas pueden ser investigadas en un dominio temporal a través del análisis de correlogramas. Cuando ciertas “tendencias” (trends en Inglés), se hayan detectado estas deben ser quitadas de las series originales y, la serie de los residuales es la que se examina. Así, el interés se centra en la distribución de probabilidad de los elementos de la serie de residuales.

La fuente por lo tanto es “la serie histórica”, y a partir de sus propiedades, se generarán las series probables en el futuro. La Gráfica 1 permite ilustrar el concepto.

MODELACION HIDROLOGICA ESTOCASTICA: DESARROLLO DE UN MODELO DE GENERACION SINTETICA DE SERIES TEMPORALES



Grafica 1.- Concepto de generación estocástica de serie temporales.

La fuente por lo tanto es “la serie histórica”, y a partir de sus propiedades, se generarán las series probables en el futuro. La Gráfica 1 permite ilustrar el concepto.

Lo rescatable es que los métodos estocásticos fueron utilizados en hidrología para resolver el problema de diseño de reservorios y su operación. La capacidad de un reservorio o embalse depende de la secuencia de flujos, especialmente la secuencia de flujos bajos. Los métodos estocásticos proveen medios para estimar la probabilidad de secuencia de años secos durante cualquier periodo específico en el futuro.

Una serie temporal, como se explicará en el apartado 2, puede modelarse matemáticamente como una combinación de los componentes determinísticos y de residuales aleatorios. Así, una ecuación de generación estocástica puede ser simple, preservando la media, la varianza y la autocorrelación de retardo 1 (Lag-1 autocorrelation).

Aunque se necesitan muchas propiedades para describir completamente una serie histórica, el análisis estocástico necesita incluir solamente aquellas que son importantes para el sistema físico que se esté estudiando. Es por ello, que es condición necesaria el identificar el esquema más apropiado para el problema que se esté tratando.

Los métodos de generación artificial de series temporales se basan en el uso de los “**registros históricos**” como una muestra de la población total. Mientras que los métodos convencionales consideran a los registros como la “población total”. De ello, se deduce que el diseño estaría basado en estimaciones de lo que “podría haber pasado”, en vez de “lo que ha pasado”.

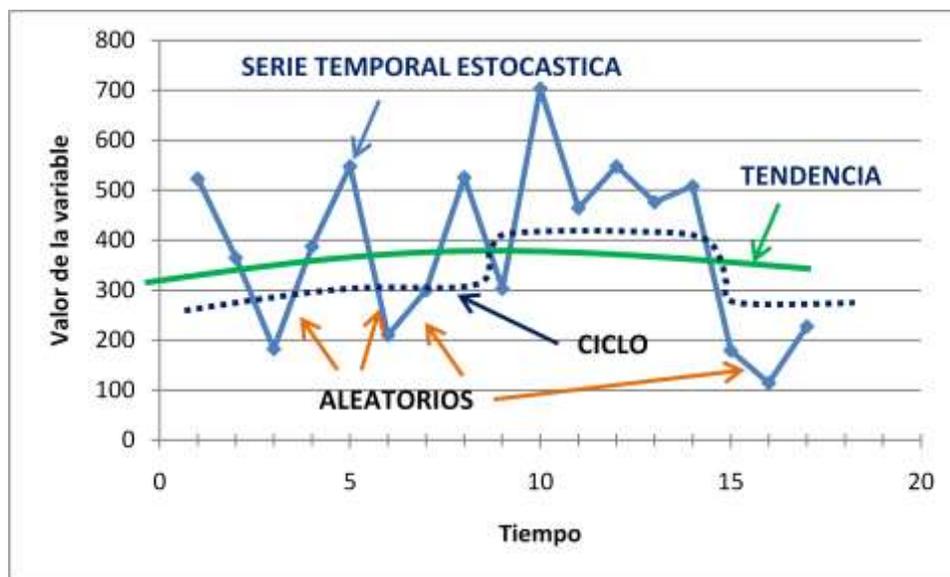
MODELACION HIDROLOGICA ESTOCASTICA: DESARROLLO DE UN MODELO DE GENERACION SINTETICA DE SERIES TEMPORALES

En líneas generales, la misma idea básica de generación estocástica es aplicable a caudales, precipitación, evapotranspiración y, como se dijo, a otras variables hidrometeorológicas.

2. Las series temporales hidrológicas y su naturaleza estocástica: el principio de la modelación estocástica.-

Cualquier serie temporal de valores observados de variables hidrológicas (ver Gráfica 2)– y que son de naturaleza estocástica - contiene los siguientes componentes:

- Una “tendencia” y, sumada a ella un componente “cíclico”; ambos son de carácter determinístico y no son independientes del tiempo a partir del cual la serie empieza, ni del tamaño de la misma; y,
- Un componente de carácter estocástico.



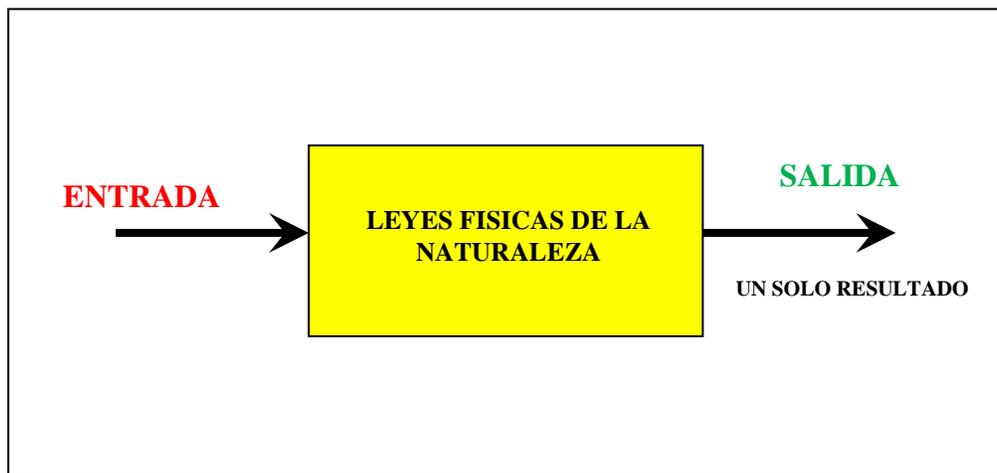
Gráfica 2.- Naturaleza de una serie temporal.-

El primer componente es el determinístico, describe una “tendencia general” y está basado en las propiedades estadísticas de la serie histórica conocida.

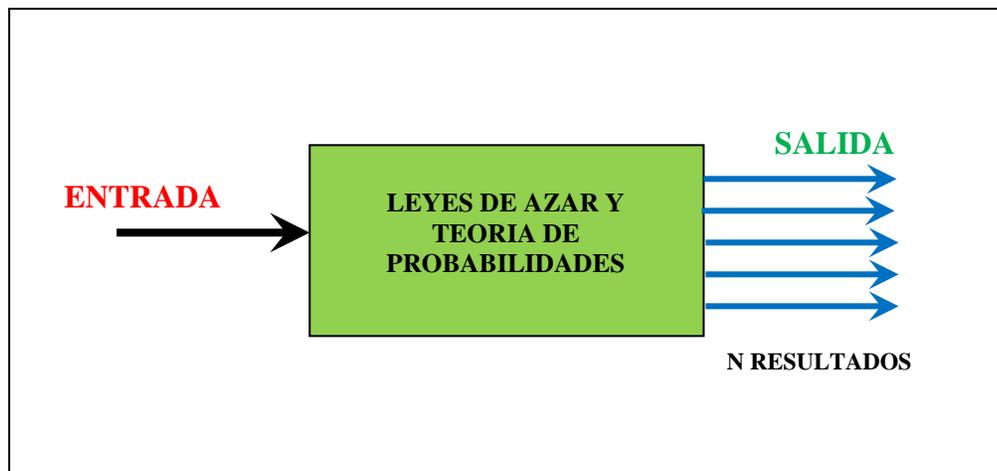
El componente estocástico, es el que define la incertidumbre o la probabilidad de ocurrencia de un evento. Este es estacionario, es decir, que los estadísticos de la muestra no difieren de los estadísticos de la población, excepto de aquellas que se produzcan por la variabilidad de la muestra y, sean independientes del tiempo.

MODELACION HIDROLOGICA ESTOCASTICA: DESARROLLO DE UN MODELO DE GENERACION SINTETICA DE SERIES TEMPORALES

Como los modelos determinísticos asocian a cada proceso una o varias relaciones basadas en leyes de la física, este dará un solo resultado y nada más (Ver Grafica 3). Mientras que aquéllos de naturaleza estocásticas (Ver Gráfica 4) y cuyos resultados están marcados por la aleatoriedad de los eventos, pueden tener varios resultados, marcados a su vez, por las probabilidades de ocurrencia de los eventos. **Los *N* resultados o series generadas tienen igual probabilidad de presentarse en el futuro, y una de ellas, no se sabe cuál, será probablemente parecida a la serie real futura.**



Grafica 3.- Modelo determinístico.



Grafica 4.- Modelo estocástico.

Si se quitan los componentes de tendencia y cíclicos de la serie, queda el componente estocástico estacionario (ver Gráfica 2). Este contiene un componente aleatorio, o de azar, y puede o no tener elemento de correlación. La **correlación** describe *como cada término de una serie es afectado por lo que ha ocurrido antes*; por ejemplo: un verano

MODELACION HIDROLOGICA ESTOCASTICA: DESARROLLO DE UN MODELO DE GENERACION SINTETICA DE SERIES TEMPORALES

(equivalente a un periodo húmedo en Bolivia) mas húmedo que lo que se conoce como “normal”, puede inducir a tener un otoño (periodo generalmente seco) con caudales mayores que el promedio de esa temporada. Por ello, el elemento de azar y la estructura de correlación del componente estocástico debe ser aislado y cuantificado. Si la serie temporal se ha descompuesto, y las partes que la componen han sido tratadas y examinadas; cada una de las partes puede reproducirse mediante simulación matemática utilizando números que ocurren al azar, series de Markov, coeficientes de correlación serial, etc., incluyendo la reintroducción de los componentes de periodicidad y tendencia. El “modelo” así generado, puede utilizarse para “generar” datos sintéticos en las cantidades que se deseen, y las series que así se producen pueden utilizarse para estimar cualquier evento correspondiente a un determinado N-año de periodo de retorno como si los “datos hubieran sido observados”.

Quizas es importante aclarar el concepto de lo que “periodicidad” implica cuando está asociada a eventos hidrológicos. Esta implicación de periodicidad es la ***posibilidad de que ciertos valores de eventos al azar pueden ser mayores en ciertos momentos (tiempos) que en otros.*** En otras palabras, podría decirse que hay un cambio cíclico de probabilidades, en vez de eventos. Es decir, que hay que reconocer que ciertos procesos que pueden ser los causantes de estos eventos, pueden estar sujetos a probabilidades cíclicas, aunque estas no sean perfectas.

3.- La Generación estocástica de variables.-

Una de las condiciones básicas para que se pueda utilizar con éxito un proceso de generación estocástica es que la serie de origen debe tener las siguientes características:

- Que la distribución temporal de origen siga una distribución normal y,
- que la respectiva distribución tenga por media cero y varianza uno.

Generalmente, cuando el intervalo de tiempo es anual o mensual, estos modelos producen buenos resultados y, a partir de los mismos se pueden desagregar a tiempos menores. Estos han sido utilizados para la generación de caudales anuales y mensuales.

Como se indicó anteriormente, si la distribución de origen tiene una distribución con media cero y varianza uno, se está hablando de un proceso Normal de media (μ) cero y varianza (τ^2) uno, que se denota por:

$$N(\mu; \tau^2) = N(0,1) \quad (1)$$

4.- Modelos Autoregresivos de medias móviles (ARMA).-

Existen lo que se llaman modelos o procesos Autoregresivos, denotados por sus siglas en inglés (Auto Regressive) AR(p), o modelos “autoregresivos con retardo p”. Como su nombre lo indica, estos generan el presente o futuro en función de lo que ha ocurrido en el pasado; de ahí el nombre de “autoregresivos”. Por otro lado, el retardo, hace mención a las etapas necesarias para que se produzca la variación de un estado a otro. Generalmente, los modelos AR(1), implican que el flujo en el periodo “i” es regresado a través del flujo en el periodo “i-1”. De manera análoga, un proceso de media móvil, es denotado por sus siglas en inglés (Moving Average) MA(q), o proceso de media móvil de orden “q”.

Uno de los modelos más sencillos es el que se utiliza para determinar valores en intervalos anuales, este tipo de modelos genera, pues, flujos anuales y toma la siguiente forma:

$$Q_i = \bar{Q} + \rho(Q_{i-1} - \bar{Q}) + t_i \sigma \sqrt{(1 - \rho^2)} \quad (2)$$

Donde: t_i = es un valor aleatorio que proviene de una distribución apropiada con media cero y varianza uno (este debe seguir el proceso $N(0,1)$), por lo que debe proceder de un proceso de generación de número aleatorio normalmente distribuido $N(0,1)$.

σ = Es la desviación estándar o típica de Q.

ρ = Es el coeficiente de autocorrelación serial con retardo “k”. Calculado a partir de la ecuación (14).

\bar{Q} = Media de Q.

De manera que cuando los parámetros del modelo, es decir \bar{Q} , σ y ρ han sido determinados a partir de los datos, una secuencia de tamaño “n” puede generarse utilizando el algoritmo dado por la relación (2).

El valor de Q_i , puede ser calculado por muestreo de Monte Carlo siguiendo una distribución de probabilidad “t”. Sin embargo, se ha demostrado que la generación utilizando la distribución normal -ya mencionada- es la más efectiva.

MODELACION HIDROLOGICA ESTOCASTICA: DESARROLLO DE UN MODELO DE GENERACION SINTETICA DE SERIES TEMPORALES

Básicamente la ecuación (2), quiere decir que **el flujo en el periodo “i”, es el valor promedio de una regresión lineal de “ Q_i ” en “ Q_{i-1} ”, más un componente aleatorio para preservar la varianza σ^2 .**

La estructura de la ecuación (2) facilita su programación en ordenador, para simplificar los procesos de cálculo y generación de números aleatorios, haciendo que la variable Q_i , sea un proceso X de media cero y varianza unitaria. Es decir $X(0,1)$. Para el caso de variable normal, este se reduce a:

$$X_i = \frac{(Q_i - \bar{Q})}{\sigma} \quad (3)$$

De esa manera, y en base a la ecuación (2) y teniendo en cuenta un proceso $X(0,1)$, el algoritmo generador, sería:

$$X_i = \rho X_{i-1} + \epsilon_i \quad (4)$$

Y, saliendo del sistema normalizado (proceso $X(0,1)$), se tiene que:

$$Q_i = \sigma X_i + \bar{Q} \quad (5)$$

Los valores de ϵ_i se seleccionan de una población apropiada que tenga media cero y varianza:

$$\sigma_\epsilon^2 = 1 - \rho^2 \quad (6)$$

De esa manera, el valor de ϵ_i , está dado por:

$$\epsilon_i = t_i * \sigma_\epsilon \quad (7)$$

Por inspección, se observa que en la ecuación (4), tanto la varianza y la correlación serial con retardo 1 del proceso, se conservan. Por otro lado, y dado que (4) es de carácter lineal, σ^2 y \bar{Q} también se preservan. Es decir, que las características de origen a través de la media y de la varianza de la serie histórica se preservan a través del proceso de generación.

Antes de entrar a desarrollar el modelo que se ha de aplicar, es de interés recalcar los siguientes aspectos fundamentales, sobre los que se basan los procesos de generación estocástica:

- La serie histórica debe seguir una distribución Normal, con media cero y varianza unitaria.

MODELACION HIDROLOGICA ESTOCASTICA: DESARROLLO DE UN MODELO DE GENERACION SINTETICA DE SERIES TEMPORALES

- En el proceso de generación se debe preservar la media y la varianza. Ecuación (4).
- En la generación de la variable aleatoria “ti”, debe preservarse tanto en la distribución que se utilice como en la media y la varianza de la misma, es decir que debe generarse dentro de un proceso $N(0,1)$.

5.- Modelo de flujo estacionales.-

Hay que tener en cuenta que el modelo de flujo presentado líneas arriba y que responde de manera general a la relación (2), es estacionario en lo que al tiempo se refiere para cada etapa de cálculo, pero no es adecuado para intervalos de tiempo menores al anual. De esa manera, se plantea un modelo de carácter “estacional”, haciendo mención a las características del flujo de mes a mes, o estación a estación (se refiere a estaciones referidas al clima), donde esta característica también debe ser representada en el modelo.

En este caso, la intención es generar caudales en periodos mensuales. A partir de este, se puede, como se dijo con anterioridad, desagregar en tiempo menores, sean semanales o diarios.

Todo lo que ha dicho en la sección 4 del presente y, que es aplicable al modelo representado por la ecuación (2), es también aplicable a los flujos estacionales, que se describen a continuación.

El modelo que se plantea sigue la siguiente relación:

$$Q_{i,j} = \overline{Q}_j + \rho_j \frac{\sigma_j}{\sigma_{j-1}} (Q_{i-1,j-1} - \overline{Q}_{j-1}) + t_{i,j} \sigma_j \sqrt{(1 - \rho_j^2)} \quad (8)$$

Donde:

“j” = Se refiere a una estación o mes, j varía por lo tanto de 1 a 12 a largo del año, para el caso de intervalos mensuales.

“i” = Se utiliza para designar el año, desde 1 hasta n, de manera similar que la ecuación (2).

ρ_j = es el coeficiente de autocorrelación serial entre las series de los meses Q_j y Q_{j-1} , y que se establece por el procedimiento establecido en el apartado 6.1 y las relaciones (12), (13) y (14).

Los demás símbolos son similares a los establecidos para la relación (2).

MODELACION HIDROLOGICA ESTOCASTICA: DESARROLLO DE UN MODELO DE GENERACION SINTETICA DE SERIES TEMPORALES

La relación (8) supone la determinación de $\overline{Q_j}$, σ y ρ para cada uno de los meses y estos valores deben ser actualizados antes de pasar a una siguiente etapa de cómputo, siempre y cuando esto sea necesario, es decir que se vaya a generar más de un año. Nuevamente, de la consideración de la relación (8) es clara la condición de media cero y varianza unitaria.

Es importante resaltar, la transmisión de estacionalidad que se efectúa a través de la variable " $Q_{i-1,j-1}$ ", ya que básicamente se deben ordenar los datos respetando el año hidrológico es decir, para el caso de Bolivia de Octubre de un año –inicio de la época húmeda- a Septiembre del siguiente –fin de la época seca-. Siendo la época húmeda la que abarca de Octubre a Marzo y la seca de Abril a Septiembre. Se tiene, por lo tanto que tener especial cuidado al inicio de los periodos toda vez que el mes de Octubre, al estar influenciado por la historia del mes de Septiembre, puede transmitir estacionalidades bajas o altas dependiendo de cual haya sido el régimen de caudales del año hidrológico anterior al que se considera.

Por otro lado, cabe recalcar que la "estacionalidad" también puede establecerse en base a dos meses, es decir a los "j-1" y "j+1", introduciendo el promedio de estos en vez del término " $Q_{i-1,j-1}$ " en la ecuación (8). Como se explicará en la parte relacionada a la modelación, dicha configuración no aporta mejoras sustanciales a los resultados que da el modelo estacional establecido con la relación (8).

La ecuación (8) indica que básicamente el caudal en el mes "j" del año en consideración es una regresión lineal entre dicho valor y el valor del caudal del mes anterior en el año anterior. Esto es importante, ya que debe tenerse cuidado en no confundir estaciones, ya que de otra manera se estarían induciendo caudales bajos después de la época seca dentro de la época húmeda y, viceversa, caudales altos dentro de la época seca, después de la época húmeda.

Continuando con el desarrollo del algoritmo necesario para poder aplicar la relación (8), de manera análoga a la que se utilizó en el apartado 4 arriba para la relación (2) del modelo simple anual, se tiene que;

$$Y_{i,j} = \frac{(Q_{i,j} - \overline{Q_j})}{\sigma_j} \quad (9)$$

Donde; $Y_{i,j}$ = Es el flujo residual en el mes "j" del año "i".

Así llevando dicho sistema al proceso $N(0,1)$, se tiene que en relación a la ecuación (8), esta última queda en la siguiente forma:

MODELACION HIDROLOGICA ESTOCASTICA: DESARROLLO DE UN MODELO DE GENERACION SINTETICA DE SERIES TEMPORALES

$$Y_{i,j} = \rho_j * Y_{i-1,j-1} + (1 - \rho_j^2)^{1/2} * t_{i,j} \quad (10)$$

Y, finalmente se tiene que:

$$Q_{i,j} = \sigma_j * Y_{i,j} + \overline{Q}_j \quad (11)$$

En relación a las relaciones (2) y (8), en las que una determina los flujos anuales y la otra flujos estacionales de forma mensual, cabe destacar que la primera resulta adecuada para grandes reservorios donde las variaciones estacionales en los flujos no afecta el requerimiento de almacenamiento o, en casos en los cuales el flujo estacional es alto, cuando las estaciones de flujos altos precede a la de alto uso para satisfacer la demanda. En los demás casos es aconsejable el uso del modelo mensual, por otro lado, y con buenos resultados se ha utilizado este último para rellenar datos faltantes –cuando estos son pocos comparado con los existentes- en una serie histórica. De esa manera se puede tener una serie histórica completa, y a partir de la misma generar la serie de años deseada.

La generación de datos de los años faltantes sobre una base mensual, se ha desarrollado en base al modelo estacional descrito y dado por la relación (8). Dicha modelación se describe en el apartado 7.

Antes de entrar a aplicar la modelación, es preciso dar a conocer los parámetros y variables necesarios que intervienen para poder aplicar el proceso dado ppor la ecuación (8).

6.- Definición de parámetros.-

Hay que tener en cuenta que cualquiera que sea el modelo a desarrollar, se debe tener una idea clara de los parámetros necesarios que corresponden al mismo y, poder evaluarlos en su justa dimensión.

6.1.- La autocorrelación con retardo k.-

El coeficiente de autocorrelación con retardo generalizado “k”, se define como sigue:

Supongase que se dispone de una serie temporal denotada por $X_1, X_2, X_3, \dots, X_n$, y que dicha serie se desplaza “k” unidades de tiempo, entonces se pueden formar las parejas: $(X_1, X_{1+k}), (X_2, X_{2+k}), \dots, (X_{n-k}, X_n)$. Es decir que se tendrían dos series que salen de la misma serie original.

MODELACION HIDROLOGICA ESTOCASTICA: DESARROLLO DE UN MODELO DE GENERACION SINTETICA DE SERIES TEMPORALES

El coeficiente de correlación entre las parejas citadas o ambas series, es decir, la serie original hasta X_{n-k} y la que se desplaza "k" de la original hasta X_n , es el que se llama coeficiente de autocorrelación de orden "k", y al desplazamiento "k", también se le llama retardo o rezago de orden "k". Dicho coeficiente se denota por " ρ_k ". Si se representa ρ_k en función del retardo "k", se tiene el autocorrelograma de la serie.

Hay que tener que cuenta que se tienen dos series de tamaño "n-k" cada una, la primera (Que se denota por X_i) toma valores que van de la posición "1" a "n-k" y, la segunda (que se denota por X_{i+k}) toman valores que van desde la posición "1+k" hasta "n".

Analíticamente el Coeficiente de autocorrelación " ρ_k ", se define como, la covarianza entre la serie X_i y la serie X_{i+k} , dividida por la varianza de la serie original σ_x^2 , es decir:

Sea,

$$\text{Cov}(X_i, X_{i+k}) = \gamma_k = \frac{\sum_{i=1}^{n-k} (X_i - \bar{X})(X_{i+k} - \bar{X})}{N-1}, \quad \text{para } k=1, 2, \dots, n-1 \quad (12)$$

Donde: γ_k = covarianza entre X_i y X_{i+k} , y esta se divide por "n-1" para evitar sesgos, y "n" es el tamaño de la serie original.

Por otra parte, la varianza de la serie original está dado por:

$$\sigma_x^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} \quad (13)$$

Donde: σ_x^2 = Varianza de la serie original.

Una mejora del modelo se obtiene al introducir, en vez de la varianza de la serie original, el producto de las desviaciones estandares de las dos series X_i y X_{i+k} . Aunque esto requiere de computo adicional, vale la pena en aras de la calidad y bondad de resultados. De esa manera, se tendrían las dos desviaciones estandares " σ_{X_i} " y " $\sigma_{X_{i+k}}$ ", que se deben introducir en la ecuación (14).

Finalmente, se tiene que el coeficiente de autocorrelación con retardo "k", está dado por:

$$\rho_k = \frac{\gamma_k}{\sigma_x^2} = \frac{\gamma_k}{\sigma_{X_i} \sigma_{X_{i+k}}} \quad (14)$$

MODELACION HIDROLOGICA ESTOCASTICA: DESARROLLO DE UN MODELO DE GENERACION SINTETICA DE SERIES TEMPORALES

6.2.- Generación de la variable aleatoria.-

Es importante resaltar que al menos que el valor de la variable aleatoria “t” pueda ser calculada a partir de una distribución apropiada, las ecuaciones (2) con (3) y la (8) con (9), no podrán reproducir la distribución histórica, aun cuando se pueda replicar la media y la varianza.

La metodología utilizada es la de generación de la variable aleatoria por muestreo de transformación inversa (Inverse transform sampling, en ingles).

La gran ventaja, es que si la distribución es normal, el problema se convierte de fácil solución y, además simple, ya que se supone que se deben generar números aleatorios que sigan la distribución normal. Por otro lado, si la variable de interés (sea caudales o volúmenes mensuales) se distribuyen según una distribución “log-normal”, la generación es afectada por el uso de variables transformadas que son normalmente distribuidas, y la exponenciación que conlleva a estas series.

En la parte aleatoria de la relación (2) o la (8), se debe pues generar una variable aleatoria “t_i”, que además sea normalmente distribuida, es decir dentro del marco del proceso N(0,1). Esta variable o número; aparte de la aleatoriedad, en la cual hay que descartar una alta periodicidad en el algoritmo generador que se proponga para evitar que estos se repitan de manera rápida en el proceso, este debe seguir el proceso de media cero y varianza unitaria.

El algoritmo que se propone, es del tipo de “generación de congruencia lineal” (Linear Congruential Generator LCG por sus siglas en Ingles), ya que se basa en series de recurrencia lineales y, se ha visto por necesario el asegurarse que el periodo en la generación sea lo bastante alto como para evitar repeticiones de baja frecuencia. En este sentido se ha utilizado el generador del compilador de Visual Basic 6.0 que tiene una periodicidad de 2^{24} (es decir del orden de 17.000). Por su parte, el algoritmo lo que hace es utilizar los residuos que se obtienen al dividir la serie de origen por un valor (en este caso por 2^{24}) que sería la recurrencia; en el caso que nos ocupa el algoritmo LCG es:

$$X_i = (1140671485 * X_{i-1} + 12820163) \text{mod} 2^{24} \quad (15)A$$

Donde, finalmente el valor que dá como resultado es:

$$u_i = \frac{X_i}{2^{24}} \quad (15)B$$

Aparte de la generación de la variable aleatoria, debe asegurarse que su valor, en la aplicación dentro del sistema normal mediante el proceso N(0,1) de la ecuación (8), se

MODELACION HIDROLOGICA ESTOCASTICA: DESARROLLO DE UN MODELO DE GENERACION SINTETICA DE SERIES TEMPORALES

encuentre dentro de los valores de la distribución normal. Esta situación se alcanza, mediante la adecuación del valor generado, para que este esté dentro de valores mayores a cero y menores o iguales a 0.3989, y a partir de este valor se evalúa el correspondiente valor de la distribución. Este valor además puede ser positivo o negativo, tal cual se presenta en una distribución normal.

Por otra parte, y para poder verificar la bondad de ejecución del programa de ordenador, se ha utilizado el generador LCG descrito, ya que de esa manera se asegura que cada vez que se corra el modelo de nuevo para los mismos datos, por lo menos se tendrá la misma serie inicial de valores aleatorios y esto permite verificar y controlar si el programa está funcionando correctamente y que las variables de interés son generadas de forma adecuada. De otra manera, cada vez que se corre el modelo de nuevo, se tendrían diferentes valores de variables aleatorias y, por ende, resultados diferentes (aunque comparables).

6.3.- Persistencia de datos.-

Una de las medidas para valorar si la extensión de los datos es adecuada, sobre todo en lo que a la persistencia de caudales altos o bajos se refiere, es mediante el concepto presentado por Hurst (1951), el cual introduce una especie de medida para detectar elementos no periódicos de baja frecuencia (es decir de rara ocurrencia) dentro de las series. Dicha relación, mide el rango o diferencia (R_n) entre los excesos acumulados mas grandes por encima de la media y los déficits acumulados mas grandes por debajo de la media, en función de la desviación estándar de la serie (σ_n), el tamaño de la serie (n) y un coeficiente "h" (que se denomina coeficiente de Hurst). Esta obedece a la siguiente relación:

$$R_n = \sigma_n \left(\frac{n}{2}\right)^h \quad (9)$$

Para que los procesos sigan el fenómeno Hurstiano, estos deben ajustarse a valores de "h" entre 0.5 y 1, con una media de 0.73, pero por otro lado se tiene que para procesos con retardo 1, el valor de "h" debe ser igual a 0.5, lo cual indica que para regímenes quasi-estacionarios el modelo de autoregresión simple, es el que reproduce el valor designado por Hurst. Cuando el valor de "h" es mayor de 0.5, este indica que existe una persistencia de largo plazo en la serie. De todas maneras, aunque este elemento favorece el uso de estructuras más sencillas, hoy por hoy es el único elemento de medida de persistencia de las series, aunque no se ha demostrado que, en toda la extensión esta pueda ser tomada como una ley, por ello en el presente estudio debe, y se toma, solo como un indicador. Para el caso de aplicación que nos ocupa, en general los valores son algo inferiores a 0.5 hecho que indica la variabilidad de los valores de los caudales, aun dentro de un mismo mes, y que, por otro lado, no se tienen series con suficiente tamaño como para valorar la periodicidad y permanencia de los caudales.

MODELACION HIDROLOGICA ESTOCASTICA: DESARROLLO DE UN MODELO DE GENERACION SINTETICA DE SERIES TEMPORALES

7.- Modelo de Ordenador y proceso de computo de datos faltantes.

Con el fin de facilitar el proceso de computo siguiendo los procedimientos explicados para la generación de caudales mensuales, en los puntos 3 al 4, se desarrolló un programa de ordenador en lenguaje Visual Basic 6.0, denominado "GARESTOC -2.0", la versión "GARESTOC-1.0" hace mención al programa desarrollado para datos anuales.

En este punto hay que mencionar que también se hizo uso del algoritmo alternativo que toma dos meses de la serie histórica, como se lo explicó en el punto 4.; sin embargo, este presenta mayores desviaciones (aunque aún bajas) que el modelo establecido con la relación (8) que solo toma un mes.

8.- Aplicación práctica a la serie de caudales históricos mensuales en Río Grande.-

Como se dijo anteriormente, uno de los problemas -mejor dicho, la situación más frecuente- que se encuentra en ingeniería es la falta de datos, en este caso de caudales. El camino a seguir, por lo tanto, tiene generalmente dos líneas:

- Modelo determinístico.- El cual está basado en leyes de la naturaleza y depende por lo tanto de la información disponible tanto de carácter geográfica-física que envuelve; el tipo de suelo, cobertura vegetal, uso de suelo, etc., y de variables Hidrometeorológicas de la cual la más importante es la precipitación y sobre todo, la distribución de los pluviómetros en la cuenca y los periodos o años de registro.
- Modelo estocástico.- Como ya se indicó, dichos modelos se aplican en base a las leyes de probabilidad y basados en la serie histórica disponible. Por lo que, requieren de algún tipo de dato del pasado, para que en el momento presente pueda predecirse el futuro.

La serie disponible para el análisis corresponde a caudales mensuales medidos por SENAHMI (Servicio Nacional de Hidrología y meteorología) en la estación de Abapo – Río Grande del Departamento de Santa Cruz desde Enero del año 1975 a Diciembre del año 1991 (17 años). Se ha procedido –para tener la serie completa- a rellenar dentro de ella algunos valores mensuales faltantes entre 1985 a 1988 y 1991 (Ver apart. 8.1), siguiendo el proceso descrito en el apartado 5.

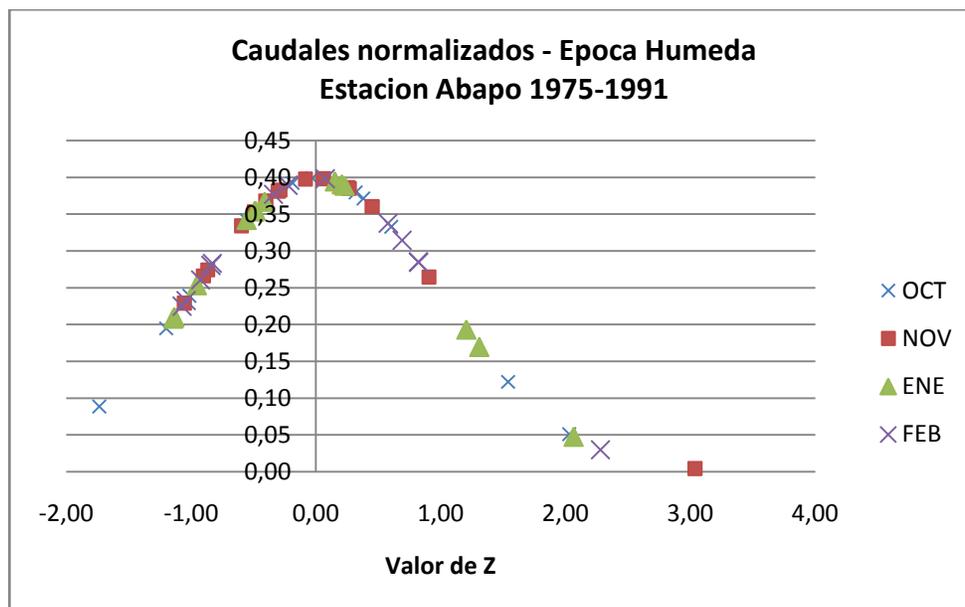
Como se indicó, una de las primeras y tareas es la de verificar si la serie histórica disponible es candidata para el análisis estocástico. El lector puede suponer que previamente a ello, también se ha verificado la consistencia de dichos datos por

MODELACION HIDROLOGICA ESTOCASTICA: DESARROLLO DE UN MODELO DE GENERACION SINTETICA DE SERIES TEMPORALES

procedimientos habituales aplicados en hidrología relacionando la serie con la de estaciones dentro de la cuenca aguas arriba de la misma.

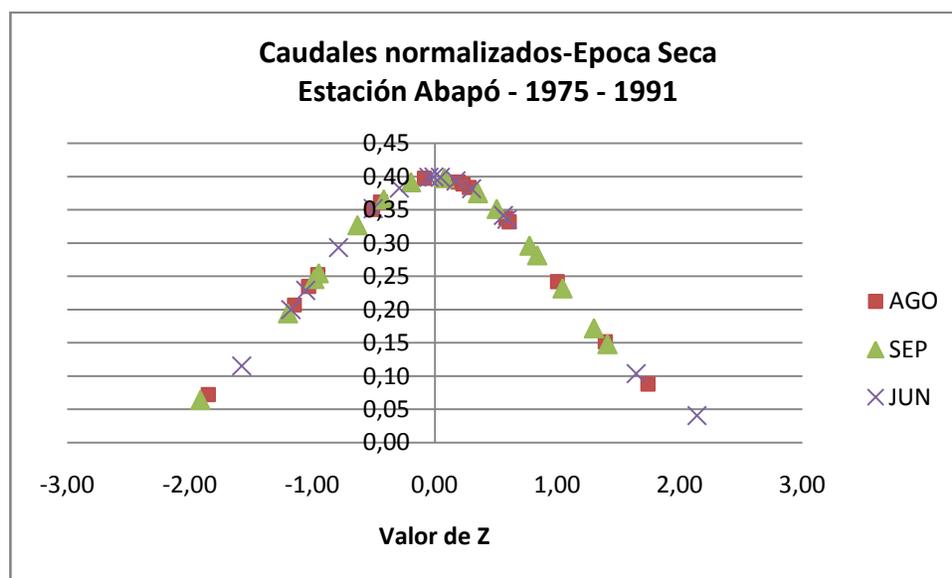
8.1.- Verificacion de las series de origen o históricas.-

Del análisis de las series de caudales de Senahmi en Abapo que abarcan desde 1975 hasta 1991 todas las series mensuales tienen una distribución Normal, con media cero y varianza uno, en ambas estaciones climáticas, es decir la húmeda (de Octubre a Marzo) y la seca (de Abril a Septiembre). Como muestra, se incluye en forma gráfica la distribución de los meses más salientes –por ser los de mayor o menor caudal- en época húmeda y seca. De la observación de las gráficas 5.- se puede inferir que estas siguen una distribución normal; y además, todas tienen media cero y varianza uno, hecho que convierte a los datos de esta estación en candidatos para ser sometidos a análisis estocástico.



Gráfica 5.A.- Caudales normalizados época húmeda. Abapo – Rio Grande.
Fuente FP Garcia Gutierrez 2009.

MODELACION HIDROLOGICA ESTOCASTICA: DESARROLLO DE UN MODELO DE GENERACION SINTETICA DE SERIES TEMPORALES

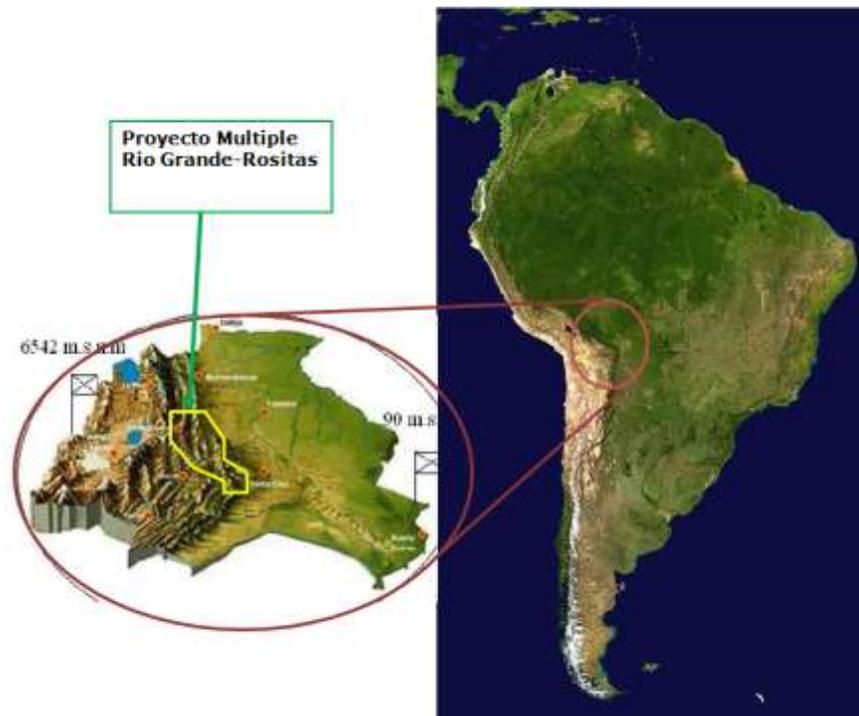


Grafica 5.B.- Caudales normalizados época seca Abapo – Rio Grande.
Fuente FP Garcia Gutierrez 2009.

8.2.- Breve descripción de la cuenca alta del Rio Grande.-

La Cuenca Alta del río Grande hasta la estación de Abapó en el departamento de Santa Cruz, es bastante grande y tiene una extensión de 59000 km². El curso principal, que nace en la cordillera oriental de los Andes en el Departamento de Cochabamba, recorre una longitud de 662 km hasta el punto de ubicación de la futura hidro-eléctrica. La Ilustración 1 muestra dicha cuenca en relación a la extensión de Bolivia.

MODELACION HIDROLOGICA ESTOCASTICA: DESARROLLO DE UN MODELO DE GENERACION SINтетICA DE SERIES TEMPORALES



Ilustracion 1.- La cuenca Alta del Rio Grande

8.3.- Datos faltantes para completar la serie histórica disponible.-

Los datos faltantes de caudales mensuales de la serie histórica que se muestran en la Tabla 1, fueron calculados año por año de forma secuencial, tomando como base la serie histórica completa de 9 años continuos que abarca desde 1976 a 1984 (serie completa de 9 años) y, en la serie correspondiente se mantuvieron los valores medidos; es decir, que para el caso de la serie de 1985, se tomó la serie histórica del 76 al 84 como dato y se generó solo la del 85; de ella se mantuvieron los datos correspondientes a los caudales medidos, es decir los de Marzo, Abril y Mayo y se incorporaron los calculados para los meses de Enero, Febrero y de Junio a Diciembre. La serie así completada –incluyendo ya la del 85-; es decir de 1976 a 1985, formó la base para calcular los datos de 1986, y así, se procedió año por año hasta completar la serie de 1991. Con ello se logra tener una serie completa desde 1975 a 1991, con 17 años completos, para determinar los caudales afluentes al área de la presa en el periodo de 1992 a 2004, y aquéllos que faltan para completar la serie histórica.

AÑO	MESES SIN DATOS	TOTAL MESES
1985	Enero, Febrero y Junio a Diciembre	9
1986	Enero a Marzo	3

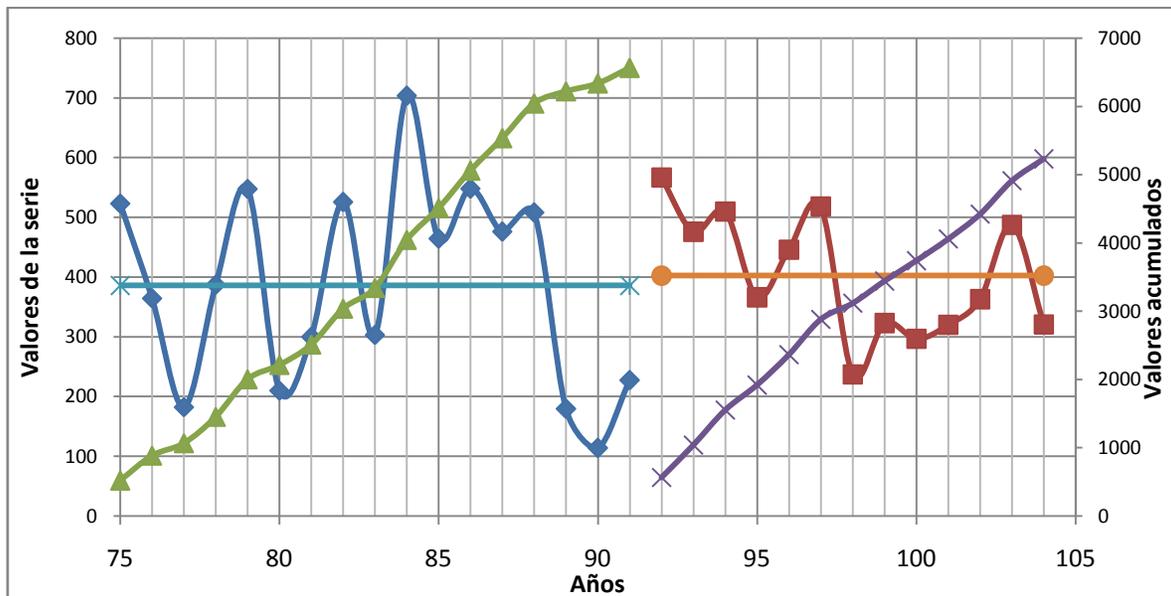
MODELACION HIDROLOGICA ESTOCASTICA: DESARROLLO DE UN MODELO DE GENERACION SINTETICA DE SERIES TEMPORALES

1987	Febrero a Mayo	4
1988	Enero a Abril	4
1991	Abril a Diciembre	9
TOTAL MESES SIN DATOS		29

Tabla 1.- Meses sin datos – Estación Abapo

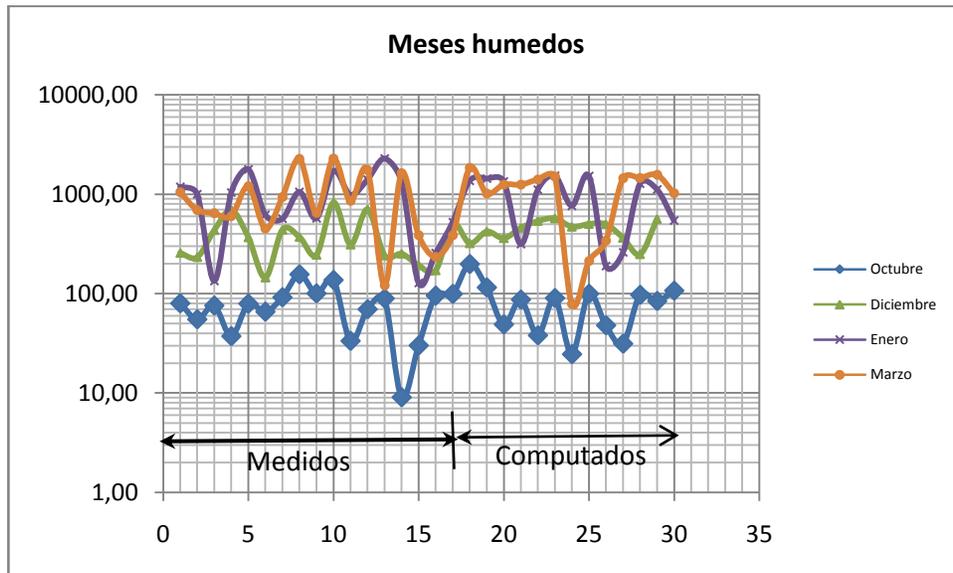
6.2.- Datos generados de 1992 a 2004.-

Con la serie completa de series de caudales mensuales que abarca desde 1975 a 1991 (17 años), se procedió a generar la serie que va desde 1992 hasta 2004, es decir de 13 años. Las variables de salida del modelo para valores anuales se muestran en la gráfica No. 2, en ella se puede ver que los valores acumulados tienen en ambos casos la misma pendiente. Las salidas del modelo para una selección de los meses húmedos (Octubre, Diciembre, Enero y marzo) y para los meses secos (Abril, Junio, y Agosto) se muestran en la gráfica No 3, y a partir de ella se puede ver que tanto las medias como las desviaciones estandards de todo el proceso son comparables y consistentes. Hay que tener en cuenta que el modelo al calcular los caudales de un año, vuelve a recalculer todas las variables incluyendo los datos generados para que estos sean la base en la generación de los nuevos caudales del año siguiente. Se puede apreciar también, que la estacionalidad producida en los meses secos y húmedos, son reproducidas y mantenidas por el modelo y que estas son persistentes con los datos.

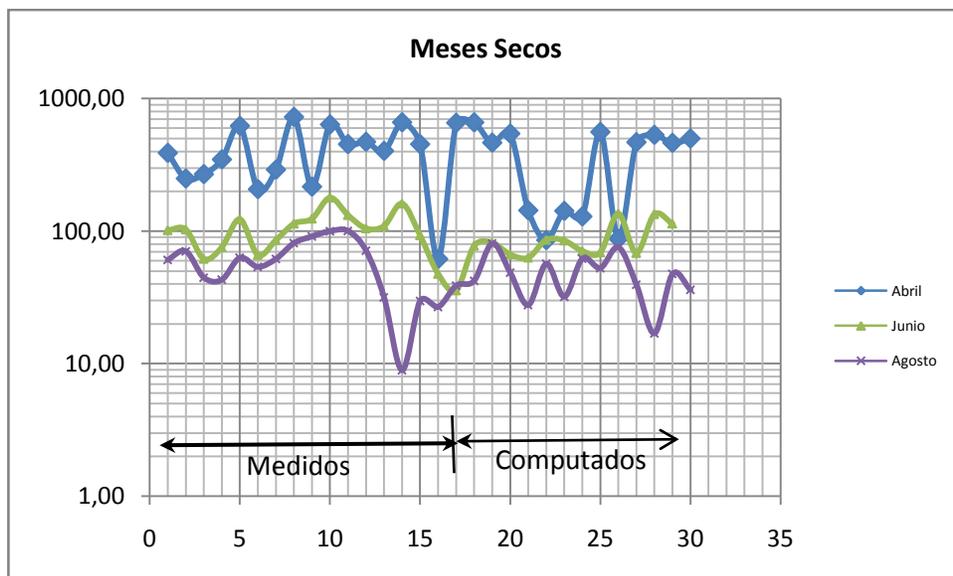


Gráfica 2.- Valores medios anuales y acumulados medidos (75 a 91) y Computados (1992 a 2004)

MODELACION HIDROLOGICA ESTOCASTICA: DESARROLLO DE UN MODELO DE GENERACION SINтетICA DE SERIES TEMPORALES



Grafica 3(a) Meses húmedos medidos y computados



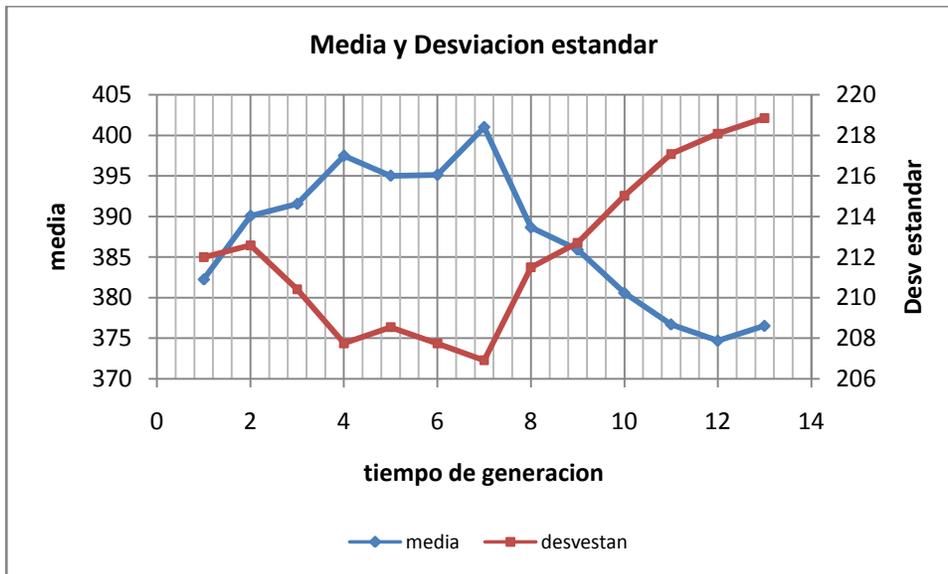
Grafica 3(b) Meses Secos medidos y computados

8.4.- Variables del proceso.-

Para ilustrar mejor los resultados de aplicación del proceso se presentan los resultados de las variables representados por sus valores medios anuales.

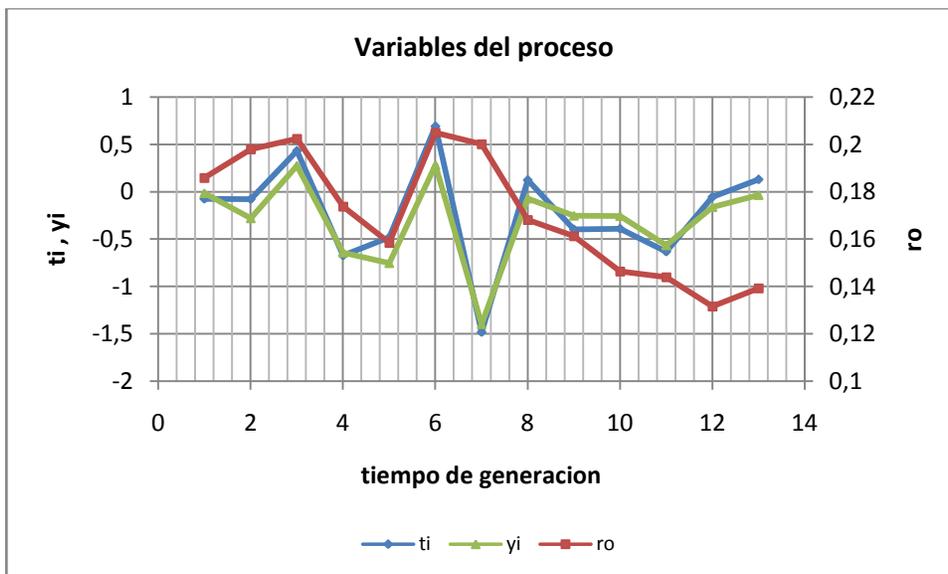
MODELACION HIDROLOGICA ESTOCASTICA: DESARROLLO DE UN MODELO DE GENERACION SINTETICA DE SERIES TEMPORALES

La gráfica 4 muestra los valores de la media y de las desviaciones estándar anuales de caudales generados.



Gráfica 4.- Valores medios anuales y desviaciones estándar anuales.

La gráfica 5 muestra los valores medios anuales de la variable aleatoria generada en el proceso $N(0,1)$ denotada por "ti"; el valor de los flujos residuales (Y_i) y del coeficiente de autocorrelación (ρ) del proceso. Observe el lector que estos últimos tienen valores bajos entre 0,12 y 0,207.



Gráfica 5.- Valores de medios anuales de variables del proceso estocástico.

MODELACION HIDROLOGICA ESTOCASTICA: DESARROLLO DE UN MODELO DE GENERACION SINTETICA DE SERIES TEMPORALES

9.- CONCLUSIONES.-

Se ha presentado la bondad de la aplicación de un modelo estocástico para la generación sintética de caudales mensuales. Si bien el proceso parece, en principio, tedioso y de compleja estructura, los resultados bien valen la pena el esfuerzo, toda vez que al ser fruto de un proceso estocástico, tal cual lo son todos los eventos naturales, quedan, - como se dijo-, pocos argumentos que puedan rebatirlos. El único cuidado que hay que tener es que los datos de origen sigan un proceso aleatorio cuya distribución tenga media cero y varianza uno.

REFERENCIAS.-

1. Linsley, R., Kohler, M; & Paulhus, J.
"HYDROLOGY FOR ENGINEERS"
McGraw-Hill Int. Book Company. Londres, Inglaterra, 1982.
2. Yevjevich, V.
"PROBABILITY AND STATISTICS IN HYDROLOGY".
Water Resources Publications. Fort Collins, Colorado, USA. 1972.
3. Yevchevich, V.
"STOCHASTIC PROCESSES IN HYDROLOGY".
Water Resource Publications. Fort Collins, Colorado, USA. 1972.
4. Garcia Gutierrez, F. P.
"ESTUDIOS DE ACTUALIZACIÓN DE LA FACTIBILIDAD DEL PROYECTO MULTIPLE RIO GRANDE - ROSITAS".
INFORME No. 1. HIDRAULICA.
Prefectura del Departamento de Santa Cruz. Santa Cruz – Bolivia, Abril de 2009.