

Generación de series sintéticas de caudales usando un Modelo Matalas con medias condicionadas

Julian Moreno Cadavid¹ & José Enrique Salazar²

1. Universidad Nacional de Colombia, Docente Escuela de Sistemas, Facultad de Minas

2. Empresas Públicas de Medellín, Profesional Tecnología Mercados de Energía

jmoreno1@unalmed.edu.co

Recibido para evaluación: 06 de Marzo de 2008 / Aceptación: 06 de Mayo de 2008 / Versión Final: 27 de Mayo de 2008

Resumen

En este artículo se presenta el uso de un modelo Matalas para la generación de series hidrológicas sintéticas. Dicho modelo emplea como datos de entrada las series históricas de caudales pero además considera el valor esperado de una variable exógena para la determinación de las medias. Tal variable es la Temperatura Superficial del Mar en la región del Niño 3.4 (TSM) la cual está relacionada, no sólo estadística si no físicamente, con los caudales. A manera de ejemplo se muestra el desarrollo del modelo para tres ríos específicos del sistema hídrico Colombiano para un horizonte de largo plazo, empleando el pronóstico de mediano plazo de la NOAA para la (TSM). La implementación del modelo se realizó utilizando las herramientas Matlab y @Risk para Excel.

Palabras Clave: Generación de Series Hidrológicas, Modelo de Thomas y Fiering, Modelo Matalas, SST.

Abstract

In this paper the use of a Matalas model for synthetic hydrologic series generation is presented. Such model uses streamflow historical series but it also considers the expected value of an exogenous variable to estimate statistical means. This variable is the Sea Surface Temperature at the Niño 3.4 región (SST) which is statistically and physically related to monthly streamflows. As an example, the development of the model using the Matlab and @Risk tools for three specific rivers of the Colombian hydropower system is presented for a long term forecasting horizon, using the NOAA's SST medium term forecast. The implementation of such model was made using the Matlab and @Risk for Excel tools

Key Words: Hydrological Series Generation, Thomas & Fiering Model, Matalas Model, SST.

1. Introducción

Dado el alto componente hidráulico del parque de generación de energía eléctrica en Colombia (cerca del 70% de la capacidad instalada corresponde a plantas de este tipo), la evolución de la hidrología desempeña un papel fundamental en la operación comercial de las plantas del sistema y por tanto en la formación del precio de la electricidad. Teniendo esto en mente, la modelación de las series hidrológicas resulta ser una tarea importante, no solo para las entidades encargadas de la planeación y operación del mercado (básicamente la UPME¹ y XM² respectivamente), si no también para las empresas involucradas que requieren bien sea evaluar la viabilidad de proyectos futuros, o proyectar los ingresos provenientes de sus plantas existentes. Este último caso con múltiples fines: elaboración de presupuesto, valoración de plantas, etc.

Un caso concreto que ejemplifica claramente la utilidad que para las empresas implica la modelación de dichas series es la reciente adopción en el mercado eléctrico colombiano del esquema de cargo por confiabilidad (CREG, 2006). De manera muy resumida, puede decirse que bajo este esquema una planta nueva debe participar en una subasta en la que compite con otras plantas por una remuneración que reciben a cambio de garantizar la entrega, bajo ciertas condiciones del mercado, de una cierta cantidad de energía conocida como obligación de energía firme. El resultado de dicha subasta es la determinación del valor de dicha remuneración o prima. Dentro de este contexto la pregunta que debe hacerse cada agente en este mercado es cuál debería ser el valor de esa prima para que la inversión en cada planta sea remunerada de manera adecuada, y de su respuesta depende la manera en la que participa en la subasta. Esa pregunta, sin embargo, no es fácil de responder pues cada agente debe considerar que sus ingresos provienen en una parte del cargo por confiabilidad, pero también por sus ingresos por ventas de energía en el mercado, los cuales a su vez están determinados por dos variables: cantidad de energía vendida y precio al que se vende. Y ese precio precisamente,

tal como se indicó al principio de la Introducción, está determinado en gran medida por la evolución de la hidrología.

Este artículo se encuentra organizado de la siguiente manera. En la siguiente sección se presenta la formulación matemática del modelo empleado, considerando la incorporación de una variable exógena (diferente a los caudales) como es la anomalía de las temperaturas de la superficie del mar (TSM) en la zona Niño 3.4. En la sección 3 se describe el preprocesamiento que debe realizarse a los datos de entrada para mejorar el desempeño del modelo, mientras que en la sección 4 se muestran los resultados obtenidos. Finalmente en la sección 5 se presentan las conclusiones de este trabajo.

2. Formulación del Modelo

El creciente número de aplicaciones en el área de los recursos hidráulicos ha exigido un estudio más detenido de los fenómenos hidrológicos, dando origen durante la década de los 70's a una disciplina conocida con el nombre de hidrología estocástica (Fiering & Jackson, 1971; Yevchevich, 1972, Shahin, 1993; McCuen, 1993) cuyo fundamento es contemplar las variaciones aleatorias que presentan dichos fenómenos, para que luego sea posible estimar la influencia que estas variaciones tienen sobre las consecuencias físicas y económicas de diversos procesos hidráulicos (Valencia, 1982). Dentro de esta disciplina, un modelo para la generación de caudales sintéticos fue desarrollado en 1962 por Thomas y Fiering (Thomas & Fiering, 1962). Este modelo contempla además de la media y la varianza, el coeficiente de correlación pues estos autores consideraron que los registros históricos indicaban la importancia de preservar este estadístico (consecuencia del fenómeno de persistencia observable en los procesos hidrológicos). Su forma es la siguiente:

$$X_{t+1} = \hat{\mu}_x + \hat{\rho}(X_t - \hat{\mu}_x) + \hat{\sigma}_x(1 - \hat{\rho}^2)^{1/2} \varphi \quad (1)$$

Donde φ es ruido aleatorio con distribución normal. La técnica de los valores esperados (empleada para definir los parámetros del modelo) dice que los estadísticos μ_x , σ_x y ρ son preservados como valores esperados. Es decir que los parámetros estadísticos de una serie sintética como las ya

1. Unidad de Planeación Minero Energética. <http://www.upme.gov.co/>

2. Expertos en Mercados. <http://www.xm.com.co/>

definidas tienden a los correspondientes valores poblacionales μ_x , σ_x y ρ , cuando el número de términos de la serie tienda a infinito. Por eso se dice que las series sintéticas son estadísticamente indistinguibles de la serie histórica en términos de la media, la varianza y el coeficiente de correlación. Igualmente, se espera que las series sintéticas proporcionen alguna idea sobre las variaciones muestrales del proceso, lo que equivale a decir que se espera que las series sintéticas guarden algún grado de coincidencia con las series reales que podrán ocurrir en el futuro.

Hasta aquí, se presenta un modelo para la generación de series sintéticas para una sola variable, es decir, para un solo río. Sin embargo, la necesidad de preservar propiedades de correlación espacial de procesos hidrológicos que tienen lugar en sitios diferentes (varios ríos), dio origen a modelos que generan varias series sintéticas a la vez, cada una correspondiente a uno de los sitios considerados. En 1967, Matalas (1967) propuso el siguiente modelo, desarrollado luego por Young y Pisano (Young y Pisano, 1968), que en realidad es una extensión del de Thomas y Fiering al caso de series múltiples:

$$X_{t+1} = AX_t + B\phi \quad (2)$$

En donde X_t es un vector columna de variables aleatorias, cada una de las cuales corresponde, por ejemplo, al caudal en el año t en uno de los z sitios diferentes; ϕ es un vector de z componentes aleatorias, independientes entre sí e independientes de las variables de X_t , cada una de las cuales tiene media cero y varianza unitaria (normalmente distribuida); A y B son matrices cuadradas de orden z . Para determinar A se tiene la siguiente ecuación:

$$A = S_{yx} S_{xx}^{-1} \quad (3)$$

En el caso de interés S_{xx} corresponde a la matriz de covarianza entre ríos (recordar que por definición la covarianza entre dos variables no es más que el promedio de los productos de las desviaciones de ambas variables respecto a sus medias respectivas), mientras que S_{yx} corresponde a la matriz de covarianza entre ríos pero con un rezago de un período (un mes). Por su parte B es una matriz que cumple con la siguiente propiedad:

$$BB^T = S_{yy} - S_{yx} S_{xx}^{-1} S_{yx}^T \quad (4)$$

Esta es una ecuación matricial en B ($BB^T = C$, siendo C dada), cuya solución puede encontrarse utilizando algunos conceptos del álgebra matricial. Como C es el producto de una matriz por su transpuesta, entonces C debe ser simétrica y definida positivamente (o al menos semidefinida). De acuerdo a esta restricción sobre C , B puede encontrarse de la siguiente manera:

$$B = PR^{1/2} \quad (5)$$

Donde P corresponde a la matriz de autovectores de C y R a la matriz diagonal de autovalores de C , calculada como aparece en (4).

Hasta aquí es la descripción del modelo de Matalas general. Ahora, si se quiere incorporar la información adicional que proviene de una variable exógena, como en el caso de interés puede ser información climática (pronóstico de la temperatura superficial del mar en la zona Niño 3.4 TSM) se puede proceder de la siguiente manera:

- 1) Se realiza el preprocesamiento al pronóstico que se tenga de la TSM según el procedimiento descrito en este documento (en la sección de Preprocesamiento de los datos).
- 2) Se generan trazas sintéticas de TSM con el modelo de Thomas y Fiering, calculando previamente los μ_x , σ_x y ρ respectivos.
- 3) Se emplean modelos de regresión lineal simple que relacionen cada uno de los caudales (ríos) con un valor rezagado de la TSM. El rezago utilizado para cada río es aquel que presente una mayor correlación. De esta manera se utiliza el pronóstico de la TSM (ya preprocesado) para determinar el valor esperado de los caudales de cada río en un horizonte determinado.
- 4) Se utiliza el valor esperado de los caudales como media condicionada para el modelo de Matalas, reescribiendo (2) como:

$$X_{t+1} = \mu_{t+1} + A(X_t - \mu_t) + B\phi \quad (6)$$

Donde μ es un vector de medias mensuales condicionadas para los z ríos del sistema.

3. National Oceanic & Atmospheric Administration. <http://www.cpc.ncep.noaa.gov/>

4. International Research Institute for Climate and Society. <http://iri.columbia.edu/>

Para obtener el pronóstico de la variable exógena (valor que se convierte en la media esperada en el modelo de Thomas y Fiering), en este caso de la anomalía TSM, se puede recurrir a alguna entidad especializada como la NOAA³ o el IRI⁴. Para hacer más claridad sobre esta variable cabe mencionar que la temperatura superficial del Mar se refiere a la temperatura media del océano que usualmente se mide mediante boyas fijas donde la profundidad a la que se realiza la medición es aproximadamente de un metro o mediante medidas satelitales que utilizan la radiación del océano en dos o más longitudes de onda en la parte infrarroja del espectro electromagnético que pueden estar relacionadas empíricamente con la TSM. Varias de estas mediciones realizadas a lo largo de la línea ecuatorial del océano pacifico se utilizan para monitorear y pronosticar el fenómeno del Niño y se dividen básicamente en tres regiones conocidas como Niño 1+2, Niño 3, Niño 4 y Niño 3.4, siendo esta última, localizada en 5° Norte - 5° Sur / 170° - 120° Oeste, Una región relevante con clima en Colombia. Dentro de las dos fuentes mencionadas que proveen un pronóstico de esta variable cabe agregar que la NOAA emplea un modelo global complejo denominado Climate Forecast System - CFS (Saha et al., 2006) que considera las interacciones entre los océanos, masas terrestres y la atmósfera.

Volviendo al modelo, considérense a manera de ejemplo los siguientes 3 ríos del sistema hídrico Colombiano: San Lorenzo, Guadalupe y San Carlos. Sus aportes mensuales medidos en m³/s para el periodo 2002 - 2006 se ilustran en la Figura 1. En esta Figura también se puede observar la variable anomalía TSM escalada en el eje y del lado derecho.

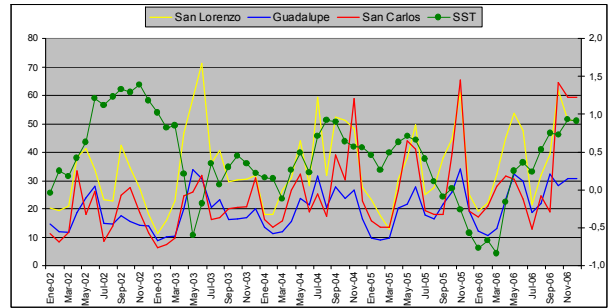


Figura 1. Aportes mensuales de los ríos San Lorenzo, Guadalupe y San Carlos.

Tabla 1. Coeficiente de correlación

	San Lorenzo	Guadalupe	San Carlos	SST 3.4
San Lorenzo	1	0.6206	0.4697	-0.4578
Guadalupe	0.6206	1	0.3939	-0.38
San Carlos	0.4697	0.3939	1	-0.3104
SST 3.4	-0.4578	-0.38	-0.3104	1

(Poveda 2004) En esta tabla puede observarse que la relación más marcada se da entre los caudales para un mes *t* de los ríos Guadalupe y San Lorenzo, mientras que la relación de ambos con el río San Carlos es más baja, particularmente la de Guadalupe.

En la parte superior de la Tabla 2 se muestran las correspondientes matrices *Sxx* y *Syx*, mientras que en la parte inferior se muestran las matrices *A* y *B* resultantes.

Tabla 2. Matrices del modelo

Matriz *Sxx*

	San Lorenzo	Guadalupe	San Carlos
San Lorenzo	0.9791	0.6066	0.3854
Guadalupe	0.6066	0.9791	0.46
San Carlos	0.3854	0.46	0.9791

Matriz *A*

	San Lorenzo	Guadalupe	San Carlos
San Lorenzo	0.6989	-0.0485	-0.0444
Guadalupe	0.1063	0.5895	-0.0953
San Carlos	0.1168	-0.1329	0.5325

Matriz *Syx*

	San Lorenzo	Guadalupe	San Carlos
San Lorenzo	0.6378	0.356	0.2035
Guadalupe	0.4249	0.5978	0.2188
San Carlos	0.239	0.1857	0.5053

Matriz *B*

	San Lorenzo	Guadalupe	San Carlos
San Lorenzo	-0.573	-0.4096	0.2521
Guadalupe	-0.681	-0.1155	-0.354
San Carlos	-0.6965	0.4499	0.1387

3. Preprocesamiento de los Datos

Para garantizar el desempeño del modelo descrito en las secciones anteriores es importante anotar que los datos utilizados, tanto históricos como esperados, requieren de un preprocesamiento que facilita el cálculo. En este caso, tal como ya se expuso, las variables consideradas son básicamente los caudales de los ríos que alimentan a las diferentes plantas de generación del país y la TSM para la zona Niño 3.4. Para los caudales se toma la serie histórica disponible, la cual para esta investigación caso contó con los datos mensuales de los ríos que se muestran en la Tabla 3 (21 en total) desde 1963 hasta 2006. Luego a dichos datos se les saca el logaritmo natural para garantizar que al generar las series sintéticas no se produzcan valores negativos de los caudales.

Tabla 3. Series de Caudales medios mensuales considerados.

Nare	San Carlos	Cauca
Guatapé	Batá	Calima
Guadalupe	Guavio	Alto Anchicayá
Concepción	Chuzá	Digua
Tenche	Bogotá	Porce
Grande	Prado	Urza
San Lorenzo	Magdalena	Miel

Posteriormente para cada uno de los doce meses se obtienen las medias y las desviaciones estándar, con las cuales se "estandarizan" los datos resultantes. A manera de resumen se puede ver este proceso como:

$$X_{m,a}^* = \log(x_{m,a}) \tag{7}$$

Siendo $X_{m,a}$ el dato de los caudales de los ríos en el mes m del año a .

Luego

$$X_{m,a}^{**} = \frac{X_{m,a}^* - \overline{X_m^*}}{\sigma_m^*} \tag{8}$$

En el caso de la TSM se procede de la misma manera para el mismo período de tiempo pero sin sacar el logaritmo natural de los datos (debido a que la estabilidad y rango de esta variable no lo justifican).

4. Resultados

El modelo descrito en la sección 2 considerando el preprocesamiento de los datos presentado en la sección 3 fue codificado en Excel con ayuda de la herramienta @Risk5, calculando previamente las matrices A y B con la ayuda de Matlab6 y los datos históricos desde 1963 hasta 2006. Los resultados obtenidos se presentan en la Figura 2 para 100 iteraciones, donde pueden observarse las ocurrencias en valores normalizados para 3 de los 21 ríos analizados en un horizonte de 42 meses (de Agosto de 2007 a Diciembre de 2010). En este caso el valor esperado de la SST se obtuvo del pronóstico de la NOAA del mes de Julio de 2007. Los 3 ríos que se presentan son los mismos utilizados en el ejemplo de la sección 2.

En la Figura 2 se pueden distinguir varios aspectos de interés para los tres ríos presentados. La línea interna (en amarillo) representa la media de las 100 ocurrencias, la franja interior (en marrón) representa una desviación estándar y por último la franja externa (en verde) representa los percentiles del 5 y el 95%. Particularmente, respecto a la media es posible notar la tendencia a condiciones húmedas respecto a las medias para los primeros meses en los tres ríos debido a relación inversa con los valores correspondientes de TSM, mientras que de allí en adelante se presentan valores que oscilan alrededor del cero respetándose las correlaciones cruzadas y las autocorrelaciones. Se observa también en esta figura que en algunas instancias se obtienen valores pico, los cuales son de gran interés en los análisis de riesgo en los que los agentes de los mercados eléctricos pueden estar interesados. Este punto es de particular importancia al analizar situaciones extremas del sistema eléctrico Colombiano las cuales son precisamente el foco de atención de mecanismos como el cargo por confiabilidad que determina cual es la remuneración que los agentes generadores reciben gracias a la prima o seguro que los usuarios pagamos (véase Jiménez, 2008).

Por su parte en la Figura 3 se presenta los resultados para las mismas 100 iteraciones pero para el valor de la anomalía SST. En este caso la línea interna que representa la media esperada corresponde a los valores que en su momento pronosticó la NOAA en un horizonte de 15 meses. A partir del mes 16 asume una media igual a cero.

5. PALISADE. <http://www.palisade.com/risk/>

6. MATHWORKS. <http://www.mathworks.com/>

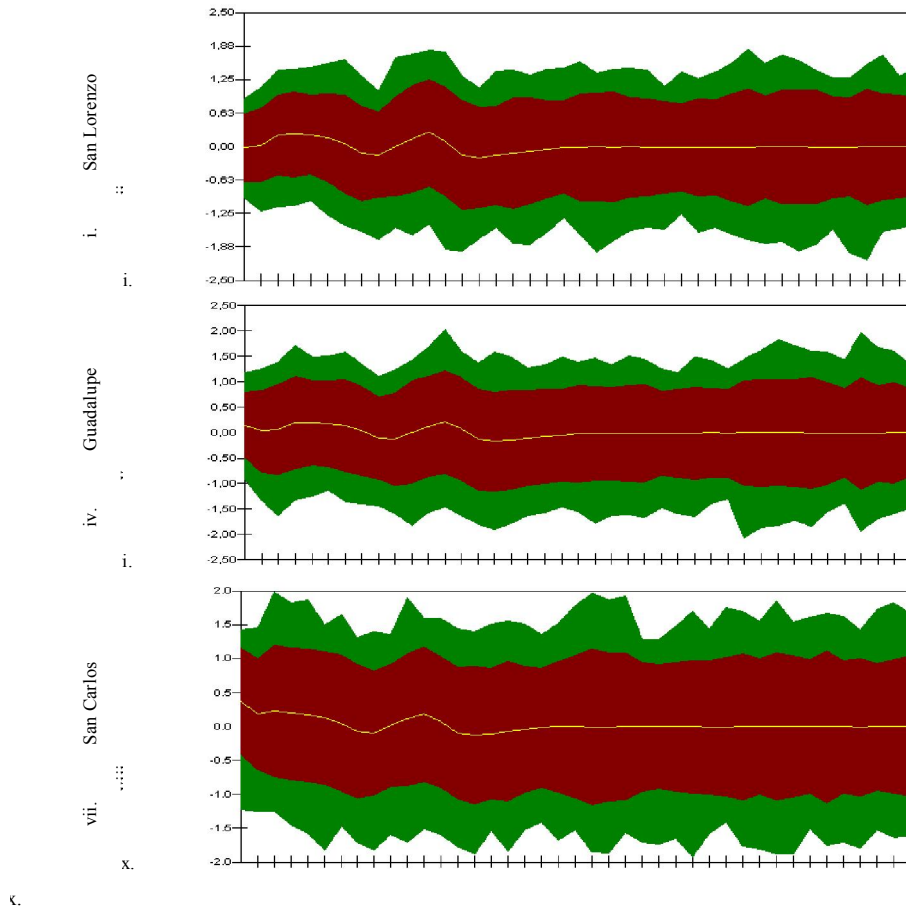


Figura 2. Generación de trazas de caudales (valores estandarizados)

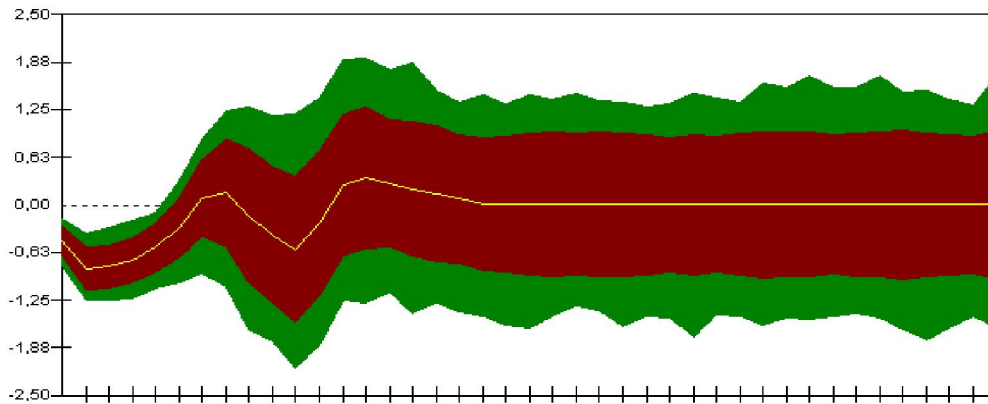


Figura 3. Generación de trazas de la anomalía TSM

Finalmente, para ilustrar aún más los resultados del modelo, en la Figura 4 se presenta una sola de las 100 las iteraciones del modelo. Aquí se observa que se preservan la magnitud de

correlaciones entre los valores (esta vez no normalizados si no en valores reales) para los tres ríos, al tiempo que se observa su relación inversa con los valores correspondientes de TSM.

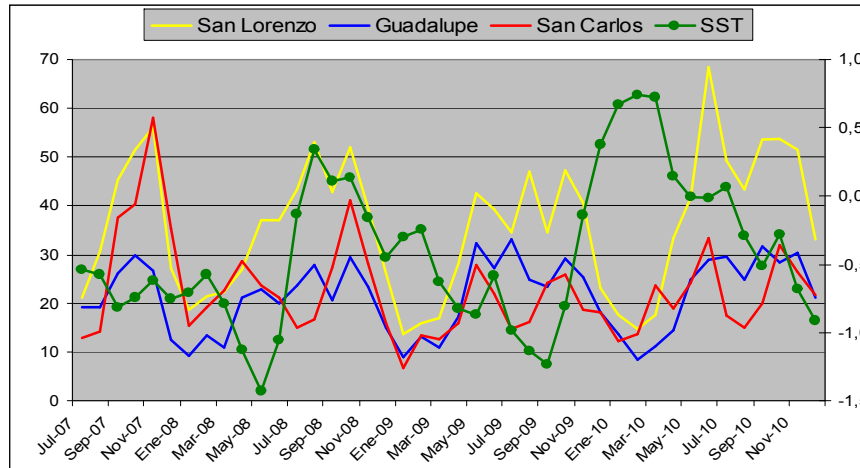


Figura 4. Generación de una trazas de caudales y de anomalía TSM

Conclusiones

Si bien existen diversos modelos provenientes de la hidrología estocástica con los que es posible modelar series de caudales para varios ríos, el presentado en este artículo exhibe un enfoque novedoso en el sentido que incorpora una variable exógena como es la SST la cual condiciona las series sintéticas obtenidas sin dejar de preservar ciertos estadísticos de interés como las varianzas y covarianzas. Este aspecto es especialmente importante si dichas series son utilizadas para el pronóstico del precio de la electricidad en mercados con características como el colombiano, en el que no solo las condiciones reales si no también la hidrología esperada, juega un papel fundamental en la determinación de dicho precio.

Es necesario mencionar sin embargo que la relación entre los caudales y la SST fue abordada en este trabajo por medio de una regresión lineal simple considerando un valor rezagado de la SST. La limitación de este enfoque es que asume linealidad entre estas variables lo que no necesariamente es cierto, pero si es válido como una primera aproximación.

Por último es importante comentar que una ventaja del modelo propuesto es su robustez matemática la cual no representa necesariamente dificultades para su implementación gracias al uso de herramientas de programación como Matlab, Excel y @Risk. La primera pues permite realizar operaciones

matriciales con relativa facilidad, mientras que las otras dos permiten un manejo fácil de los datos tanto de entrada como de salida.

Referencias

- CREG. Comisión de Regulación de Energía y Gas, 2006. Resolución 071.
- Fiering, M. y Jackson, B., 1971. Synthetic Streamflows. Water Resources Monograph 1. American Geophysical Union.
- Jimenez, G., 2008. Seguro contra apagones, eso es la subasta de energía. Reportaje a Pablo Corredor, Presidente de XM (filial de ISA). El Colombiano, 5 de Mayo.
- Mccuen, R., 1993. Microcomputer Applications in Statistical Hydrology. Prentice Hall.
- Matalas, N., 1967. Mathematical assessment of synthetic hydrology, Water Resources Research 3 (4), pp. 937-945.
- Saha, S.; et al., 2006. The NCEP Climate Forecast System. Journal of Climate, vol. 19, (15,) pp. 3483 3517.
- Shahin, M., 1993. Statistical Analysis in Water Resources engineering. Applied Hydrology Monographs. A. Balkema.
- Thomas, H. y Fiering, M., 1962. Mathematical synthesis of streamflow sequences for the analysis of river basins by

simulation. En: Maass A, Humfschmidt MM, Dorfman R, Thomas Jr HA, Marglin SA, Fair GM (eds.) Design of water resource systems. Harvard University Press, Cambridge.

Valencia, D., 1982. Optimización y Simulación en Sistemas de Recursos Hidráulicos. Universidad Nacional de Colombia.

Yevchevich, V., 1972. Stochastic Processes in hHydrology. Water Resource Publications. Fort Collins Co.

Young, G. y Pisano, W., 1968. Operational hydrology using residuals. Journal of Hydraulics Division, ASCE, 94(HY4), pp. 909-923.

En Internet:

http://www.elcolombiano.com.co/BancoConocimiento/S/seguro_contra_apagones_eso_es_la_subasta_de_energia/seguro_contra_apagones_eso_es_la_subasta_de_energia.asp