

Predicción de caudales en la cabecera de la cuenca del Paute mediante el modelo DBM

A. Quichimbo¹, R.F. Vázquez^{1,2}

¹ Departamento de Recursos Hídricos y Ciencias Ambientales, Universidad de Cuenca, Av. 12 de abril S/N, Cuenca, Ecuador.

² Facultad de Ingeniería, Universidad de Cuenca, Av. 12 de abril S/N, Cuenca, Ecuador.

Autores para correspondencia: andresquichimbo@gmail.com, raulfvazquezz@yahoo.co.uk

Fecha de recepción: 21 de septiembre de 2014 - Fecha de aceptación: 17 de octubre de 2014

RESUMEN

El modelo Mecanicista Basado en Datos (DBM) se ha utilizado conjuntamente con el filtro de Kalman (como una técnica de asimilación de datos) para la predicción de caudales en una subcuenca ubicada en la parte alta de la cuenca del río Paute. Los resultados sugieren que el modelo DBM, conjuntamente con la técnica de asimilación de datos empleada, produce predicciones de mejor calidad en la subcuenca de estudio, en comparación al uso exclusivo del modelo DBM; de hecho, el filtro de Kalman provee una medida de la incertidumbre asociada al empleo del modelo DBM para efectuar pronósticos de caudales. Estos resultados, no solo que alientan el uso futuro de modelos basados en técnicas de minado de datos, sino que además alientan el uso de la herramienta actual tanto para realizar predicciones como para el pronóstico y alerta temprana en cuencas Andinas.

Palabras clave: Modelización numérica; DBM; asimilación de datos; función de transferencia (TF); parámetros dependientes de estado (SDP).

ABSTRACT

The Data-Based Mechanistic (DBM) model was used in conjunction with the Kalman filter (as a data assimilation technique), to predict the discharge from a sub-catchment located in the upper part of the Paute basin. The results showed that this conjunctive use of the DBM model and the Kalman filter produced better predictions of the discharge in the study site, as compared to the solely use of the DBM model; indeed, the use of the Kalman filter provided an estimate of the uncertainty associated to the use of the DBM model for forecasting purposes. These results not only motivate the future use of data mining techniques for discharge forecasting, but also encourage the use of the current tool for both, prediction as well as forecasting extreme events on Andean catchments.

Keywords: Numerical modelling; DBM; data assimilation; transfer function (TF); state dependent parameters (SDP).

1. INTRODUCCIÓN

El análisis, la predicción y el pronóstico de caudales es uno de los problemas de mayor interés en la hidrología moderna. Comprender los procesos hidrológicos y tener la capacidad de pronosticarlos a tiempo, es esencial para la gestión sostenible de cuencas y de sus recursos naturales, entre los cuales, el agua es quizás el de mayor relevancia dados sus usos potenciales en términos de demandas doméstica, industrial, producción agrícola y generación hidroeléctrica. En particular este último uso es de fundamental importancia para el desarrollo del Ecuador (CONELEC, 2009; CONELEC, 2011). Por lo tanto, si un modelo es capaz de reflejar la esencia de cómo funciona una cuenca hidrológica (para

toda la gama de sus estados posibles), entonces se podría extrapolar con cierta confianza, más allá de las condiciones observadas y llegar a predicciones y/o pronósticos confiables (Vázquez, 2003).

Por otro lado, en relación al pronóstico de caudales en tiempo real, la región Austral del Ecuador es una área donde la alta variabilidad típica de los Andes se constituye en una limitante natural para la aplicación trivial de modelos, por lo que se destaca aún más la necesidad de desarrollar técnicas más eficientes para dichos pronósticos, basadas fundamentalmente en el uso de herramientas numéricas de simulación alimentadas con insumos de un sistema de monitoreo en tiempo real (Vázquez, 2003; Young, 2013).

El presente artículo reporta la evaluación del modelo híbrido “Data Based Mechanistic” (DBM), que combina el aprendizaje a partir de los datos con la selección de una estructura con cierto significado físico (Box & Jenkins, 1976), bajo la hipótesis de que esta es una herramienta adecuada para la predicción (y potencialmente, pronóstico) de caudales en cuencas Andinas del Sur del país.

2. MATERIALES

La zona de estudio, pertenece a la parte alta de la cuenca del Paute. Comprende tres subcuencas: la del río Yanuncay, la del río Tarqui, y la del río Tomebamba (Fig. 1). La estación de medición de caudal, utilizada en el presente estudio, se encuentra ubicada en el río Tomebamba aguas abajo de la intersección con del río Yanuncay (Fig. 1). El área de estudio total es de aproximadamente 1275,4 km². Esta zona es accidentada con valles y llanuras, con un rango de elevación que va desde los 2440 hasta los 4420 metros sobre el nivel del mar (m s.n.m.). Para la presente modelización, se contó con series diarias de evapotranspiración, caudal y lluvia en el período (1 de julio de 1981 - 27 de marzo de 1985), en el que se ha registrado una lluvia media diaria de 2,39 mm (aproximadamente 778,7 mm por año).

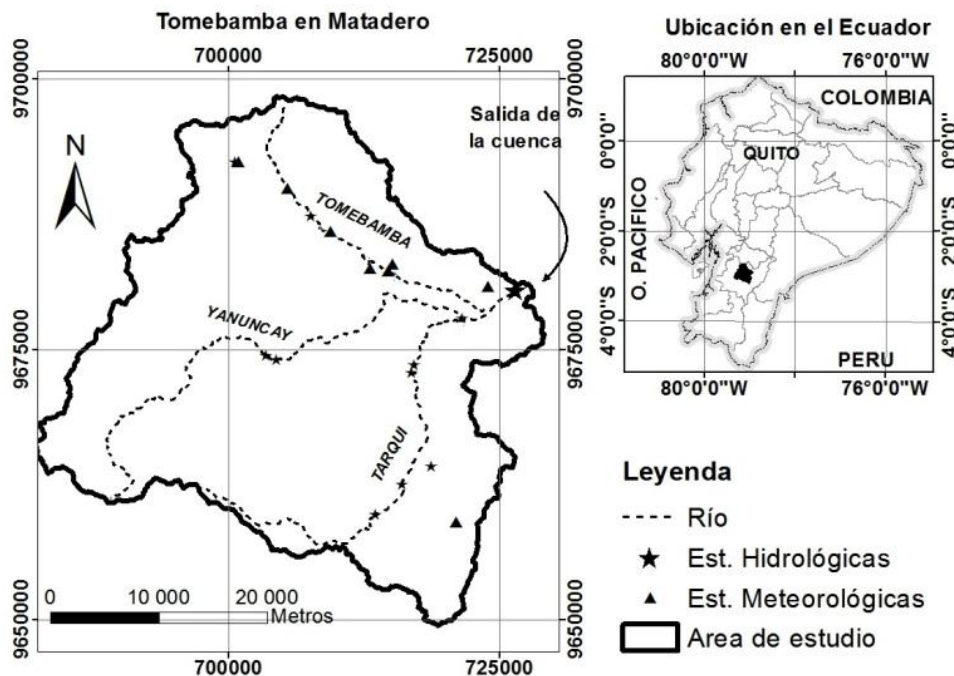


Figura 1. Ubicación de la zona de estudio y distribución espacial de las respectivas redes hidro-meteorológicas.

3. MÉTODOS

3.1. El sistema de modelización DBM

El modelo DBM (Young, 2001) combina por un lado las características de los modelos “métricos” que, basados en datos observados, caracterizan la respuesta del sistema modelado, usando para ello estimación u optimización estadística (Wheater *et al.*, 1993); y por otro lado, de los modelos conceptuales, cuya estructura se basa en las percepciones y experiencia del modelador.

La filosofía de la metodología DBM (Young, 1998; Lees, 2000a; Young, 2001; Young, 2002a; Lin, 2003) se puede representar de forma esquemática como se indica en la Fig. 2, en la que se distinguen dos etapas principales y bien diferenciadas que permiten observar la combinación de los modelos métrico, por un lado, y conceptual por el otro.

La modelización DBM de precipitación-escorrentía presenta una estructura no lineal (Lees, 2000b; Young, 2001; Young, 2002a), conformada por dos componentes: uno lineal, que describe el comportamiento del hidrograma unitario instantáneo; y otro no lineal, que describe la relación entre la precipitación medida y la precipitación efectiva, controlando de esta manera el aporte de la precipitación al hidrograma a través del tiempo.

El modelo resultante es una función de transferencia (TF) de la siguiente forma:

$$y_t = \frac{B(z^{-1})}{A(z^{-1})} u_{t-\delta} + \xi_t \quad (1)$$

donde y_t es, en el presente caso, el caudal en el tiempo t ; $u_{t-\delta}$ la precipitación efectiva en el tiempo $t-\delta$; y ξ_t representa la incertidumbre. z^{-1} es un operador de cambio recursivo que se define como $z^{-1}(y_t) = y_{t-1}$. $A(z^{-1})$ y $B(z^{-1})$ se definen como:

$$\begin{aligned} A(z^{-1}) &= 1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_n z^{-n} \\ B(z^{-1}) &= b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_m z^{-m} \end{aligned} \quad (2)$$

cuya estructura queda determinada por la triada $[n \ m \ \delta]$, donde, n es el orden de $A(z^{-1})$, m es el orden de $B(z^{-1})$; y δ representa el tiempo de retraso.

La identificación del orden de la TF se realiza en base a criterios no solo estadísticos, sino de otros criterios cuyo objetivo es medir cuán bien se ajusta el modelo a los datos e indicar si existe sobre parametrización de los mismos. Los criterios utilizados en el presente artículos son: Coeficiente de Eficiencia o índice de Nash (3):

$$EF_2 = 1 - \frac{\sigma^2}{\sigma_y^2} \quad (3)$$

el criterio de identificación de Young:

$$YIC = \ln \left(\frac{\sigma^2}{\sigma_y^2} \right) + \ln \left(\frac{NEVN}{np} \right) \quad (4)$$

y el criterio de información de Akaike:

$$AIC(np) = N \ln \sigma^2 + 2 np \quad (5)$$

donde:

$$\begin{aligned} \sigma_y^2 &= \frac{1}{N} \sum_{k=1}^N [y_k - \bar{y}]^2 \\ \bar{y} &= \frac{1}{N} \sum_{k=1}^N y_k \end{aligned}$$

$$NEVN = \frac{1}{np} \sum_{i=1}^{np} \frac{\sigma^2 \hat{p}_{ii}}{\hat{a}_i^2}$$

$np=n+m+1$ es el número de parámetros estimados del vector $\hat{a}(N)$ y \hat{a}_i es el i -ésimo elemento de este vector (parámetros de la función de transferencia). \hat{p}_{ii} es el i -ésimo elemento de la diagonal de la matriz de covarianza $\hat{p}(N)$ obtenida del proceso de estimación de parámetros. σ_y^2 es la varianza de las observaciones disponible; σ^2 es la varianza de los residuos generados del modelo ($\sigma^2 \hat{p}_{ii}$ sería la varianza de la incertidumbre estimada del i -ésimo parámetro estimado); y N es el número de observaciones disponibles para la estimación de los parámetros.

EF_2 es una medida de cuán bien se ajusta el modelo a los datos. tiende a la unidad cuando σ^2 es baja comparada con σ_y^2 ; y tiende a cero cuando ambos tienen valores con magnitud similar (Vázquez, 2003).

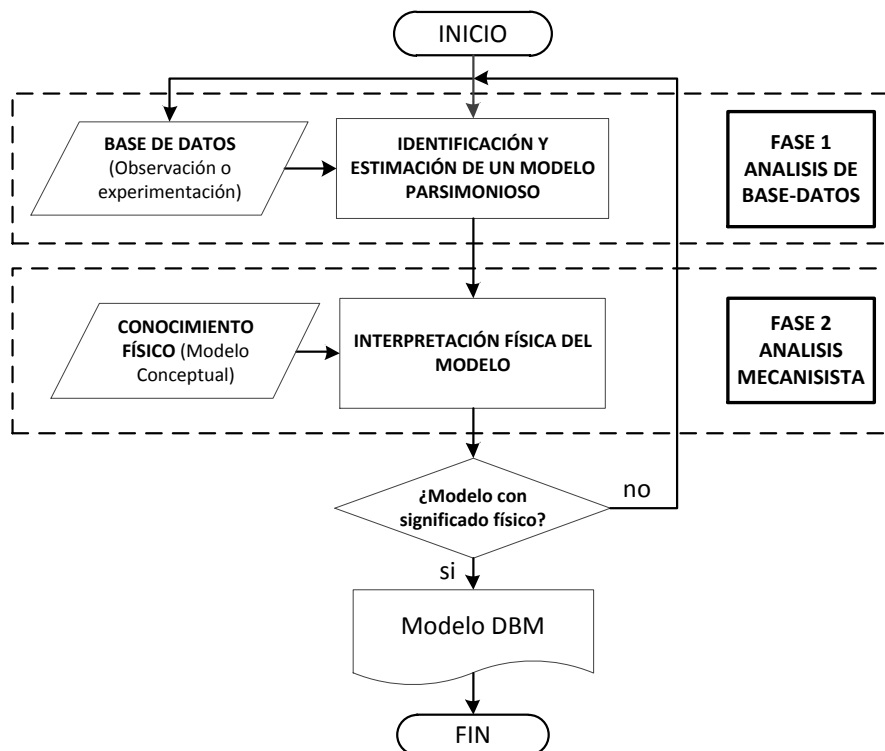


Figura 2. Representación esquemática de la estructura de modelo DBM mediante un diagrama de flujo de procesos.

El YIC consta de dos partes: la primera parte nos indica, con un criterio estadístico, cuan bien se ajusta el modelo a los datos y la segunda parte nos indica si el modelo esta sobre-parametrizado, en cuyo caso, el segundo término tiende a dominar la función tomando valores cada vez mayores; por lo tanto el valor que debe tomar la función para determinar un modelo óptimo debería ser negativo con su valor absoluto cada vez mayor.

Por último, el AIC se utiliza para identificar el orden de modelos Auto Regresivos (AR). Sin embargo, puede además utilizarse para identificar el orden del modelo de TF. Es parecido al YIC pues busca una relación entre el ajuste del modelo y su complejidad, por lo que ayuda a asegurar el carácter parsimonioso (sin sobre-parametrización) de la TF.

Por otro lado, el componente no lineal del modelo se considera como la precipitación efectiva, que es la causante de las variaciones del caudal observado. En el contexto de la modelización DBM, éste puede ser cualquier función (Young & Tomlin, 2000; Mwakalila *et al.*, 2001; Young, 2001; Young, 2002a; Young, 2002b; Young, 2013) como $u_t = f(y_t, r_t)$. Esto no quiere decir que la precipitación efectiva es una función del caudal, sino más bien que el caudal está actuando como un

sustituto objetivamente identificado (sobretudo conveniente) del almacenamiento de la cuenca. Desde el punto de vista hidrológico, el caudal es una función del almacenamiento de la cuenca y su patrón de cambio temporal es probablemente similar (Young & Tomlin, 2000; Young, 2001). Por lo tanto, esta función intenta, de la manera más simple, la representación de los diferentes procesos de evapotranspiración y variación de la humedad de suelo en el sistema. Para ello la precipitación es modificada de tal manera que refleje los efectos de la precipitación antecedente en la cuenca, metodología que se relaciona con el conocido Índice de Precipitación Antecedente (API) (Young, 2002a; Beven, 2004).

Los parámetros de $A(z^{-1})$ y $B(z^{-1})$ se estimaron en el presente estudio mediante la herramienta CAPTAIN (Pedregal *et al.*, 2007) disponible para Matlab®.

Uno de los aspectos más importantes de la modelización DBM es el significado físico que se le puede dar al modelo resultante; de hecho, el modelo no se considera como válido si no tiene una interpretación física adecuada.

La etapa final de la modelización DBM es la validación, proceso mediante el cual se evalúa el potencial de predicción del modelo (Refsgaard & Knudsen, 1996).

3.2. Identificación, estimación y validación

Para los procesos de identificación, calibración y validación del modelo se utilizó un ensayo "split-sample" (Refsgaard & Knudsen, 1996). De esta forma, y en función de la disponibilidad de datos (precipitación y escurrentía), para la subcuenca del Tomebamba en Matadero, se definieron los siguientes intervalos: (1 de julio de 1981 - 3 de octubre de 1983) para identificación y calibración; y (4 de octubre de 1983 - 27 de marzo de 1985) para validación.

En la fase de identificación y calibración, la elección de la estructura más adecuada se basó en la evaluación de las distintas estructuras ensayadas y en particular mediante la función SRIV (Simplified Refined Instrumental Variable) para la estimación de los parámetros de la TF lineal y la función SDP (State-Dependent Parameter) para la parte no lineal. Estas funciones conjuntamente con otras funciones estadísticas utilizadas en el presente estudio se encuentran disponibles en la herramienta CAPTAIN.

La etapa final de la modelización DBM es la validación, proceso mediante el cual se evalúa el potencial de predicción del modelo; ésta se realiza por lo general con un grupo de datos diferente al utilizado en las etapas de identificación y estimación (split-sample; Refsgaard & Knudsen, 1996).

3.3. Filtro de Kalman

En el contexto de asimilación y pronóstico de datos, la ecuación obtenida de la descomposición de la ecuación (1) puede expresarse como un sistema estocástico estado-espacio; de esta manera, se puede utilizar el filtro de Kalman (Kalman, 1960) para actualizar los estados del sistema (en el presente estudio, los componentes de la escurrentía) (Lees, 2000a; Young & Tomlin, 2000; Wheater *et al.*, 2008).

El filtro de Kalman aplicado en este estudio se resume de la siguiente manera: (a) predicción,

$$\begin{aligned}x_{t|t-1} &= Fx_{t-1} + Gu_{t-\delta} \\ P_{t|t-1} &= FP_{t-1}F' + Qr \\ y_{t|t-1} &= H^T x_{t|t-1}\end{aligned}\tag{6}$$

y (b) corrección:

$$\begin{aligned}x_t &= x_{t|t-1} + P_{t|t-1}H^T[1 + HP_{t|t-1}H^T](y_t - y_{t|t-1}) \\ P_t &= P_{t|t-1} + P_{t|t-1}H^T[1 + HP_{t|t-1}H^T]HP_{t|t-1} \\ y_t &= H^T x_t\end{aligned}\tag{7}$$

donde: x_t es un vector de $n \times 1$ de estados del sistema (componentes no observados del sistema, en el caso hidrológico los subflujos); F es una matriz de $n \times n$ de los parámetros de estado del sistema (eigenvalores del denominador de la TF); G es un vector $n \times 1$ que contiene los parámetros de la variable de control o variable exógena u_t (numeradores de las TFs resultantes luego de la descomposición de la ecuación 1), $H=[1 \ 1 \ 1]$ es el vector observación; P es la matriz de covarianza de $n \times n$ elementos asociada a los de los estados del sistema; y Q_r es una matriz de $n \times n$ que contiene los NVRs (Noise Variance Ratio) o hiperparámetros del modelo (Kalman, 1960), generalmente diagonal (Lees, 2000a).

Los NVRs permiten que las estimaciones de los estados sean ajustados de tal manera que tengan en cuenta la influencia de perturbaciones estocásticas de entrada que no pueden ser medidas. De esta manera definen el nivel de incertidumbre en la evolución de cada estado con respecto a la incertidumbre de la medición (Kalman, 1960).

4. RESULTADOS

El estudio mostró que la estructura [2 3 0] es la más efectiva paraméricamente (estructura parsimoniosa) para la subcuenca de estudio, lo cual coincide con (Young, 2002b; Romanowicz *et al.*, 2008) para una base de datos diaria (Tabla 1). Los parámetros obtenidos para la TF de la estructura seleccionada para Tomebamba en Matadero se muestran en la siguiente ecuación:

$$y_t = \frac{0.113(0.0018) - 0.016(0.0016)z^{-1} - 0.089(0.0001)z^{-2}}{1 - 1.454(0.0004)z^{-1} + 0.464(0.0002)z^{-2}} u_{t-\delta} \quad (8)$$

donde, la varianza de los parámetros estimados es el factor que se encuentra entre paréntesis.

Tabla 1. Estructuras identificadas para el modelo DBM.

[n m δ]	EF ₂	YIC	AIC
[3 3 0]	0,803	3,687	3,298
[2 3 0]	0,800	-0,926	3,307
[3 2 0]	0,791	-4,566	3,354
[1 3 0]	0,785	-4,365	3,377
[2 3 1]	0,775	-4,339	3,427

La parte no-lineal de la estructura DBM consiste en una función potencial de la forma $f(y_t) = cy_t^\beta$, la misma que se optimizó empleando el método Simplex (Nelder & Mead, 1965). El proceso de optimización de los parámetros de la función potencial arrojó los valores $c = 1.33$ y $\beta = 0.97$.

Por otro lado, como parte fundamental del proceso de modelización DBM, la interpretación física, la ecuación (6) fue descompuesta de la siguiente manera:

$$y_t = 0.113u_{t-\delta} + \frac{0.982}{1-0.009z^{-1}}u_{t-\delta} + \frac{0.472}{1-0.140z^{-1}}u_{t-\delta} \quad (9)$$

La ecuación 7 sugiere que la precipitación efectiva u_t aporta al caudal con tres sub-flujos diferentes: uno muy rápido (instantáneo) que representa aproximadamente el 12% del caudal total y que se debe probablemente a procesos superficiales; un componente de flujo rápido con un tiempo de residencia de 1,33 días, que aporta aproximadamente el 30% al hidrograma y que probablemente se debe a procesos de poca profundidad; y un componente de flujo lento con un tiempo de residencia de 58,32 días y que aporta aproximadamente el 58% al cauce total del río siendo además su componente principal, el mismo que se asocia a procesos sub-superficiales y subterráneos.

En cuanto al rendimiento del modelo, evaluado mediante el índice de Nash (EF₂), se obtuvo un valor de 0,81 para el período de calibración y de 0,75 para el período de validación. En relación al

balance hídrico, el modelo tiene la capacidad de representar adecuadamente la producción de escorrentía, pues alcanza valores por sobre 97% del volumen observado para los períodos tanto de calibración como de validación.

Sin embargo, a pesar de que el modelo anterior presenta un buen rendimiento, además de proporcionar una interpretación física, el modelo no puede ser utilizado para fines de pronóstico debido a que el retraso es igual a cero ($\delta = 0$). Sin embargo, para dotarle al modelo de una capacidad predictiva fue conveniente definir un retraso artificial de un día ($\delta = 1$), teniendo en cuenta que el mismo puede afectar la descripción física del sistema que el modelo proporciona.

Debido al cambio de δ fue necesario una reestimación de los parámetros de la TF, así como también, una nueva descomposición. La nueva TF y su respectiva descomposición de muestran en las ecuaciones 10 y 11, respectivamente.

$$y_t = \frac{0.2159(0.0001) - 0.3016(0.0006)z^{-1} - 0.0894(0.0001)z^{-2}}{1 - 1.6909(0.0001)z^{-1} + 0.6949(0.0003)z^{-2}} u_{t-1} \quad (10)$$

$$y_t = 0,2159u_{t-\delta} + \frac{0,0067}{1-0,9865z^{-1}} u_{t-\delta} + \frac{0.0657}{1-0.7044z^{-1}} u_{t-\delta} \quad (11)$$

El cambio de δ provocó una reducción del rendimiento del modelo en el período de calibración de 0,80 a 0,77 medido con el índice EF_2 , tal como se muestra en la Tabla 1. La tabla también muestra, mediante el índice YIC, que el modelo [2 3 1] reduce la incertidumbre en la estimación de los parámetros de la TF, tal como lo muestra la reducción del índice de -0,92 a -4,33.

Una vez realizada la descomposición, la ecuación 11 fue expresada en forma estado-espacio e introducida dentro del filtro de Kalman. La utilización del filtro mejoró el rendimiento del modelo tanto de los caudales picos como de los caudales bajos. Los valores de EF_2 mejoraron de 0,77 a 0,83 para el período de calibración y de 0,72 a 0,81 para el período de validación. Además, la utilización de filtro de Kalman permite la estimación de la incertidumbre del pronóstico del modelo.

La Fig. 3 presenta el hidrograma resultante en un período de tiempo seleccionado de la etapa de calibración, además, en la misma figura muestra la franja de incertidumbre del modelo. Es importante observar en la figura que la mayoría de las observaciones caen dentro de la franja de incertidumbre, especialmente los caudales bajos. Sin embargo, también se observa que en los caudales bajos, la franja de incertidumbre cae por debajo de 0, lo cual indica que el modelo presenta serias inconsistencias en la descripción del sistema precipitación-escorrentía

Por otro lado, la Fig. 4 muestra el diagrama de dispersión para el período de validación, en la misma se puede observar que el modelo pronostica los caudales bajos con mayor eficiencia que los caudales picos.

Considerando tanto los resultados estadísticos como los gráficos, el modelo DBM presenta un rendimiento aceptable tanto en el período de calibración como en el de validación, con una reducción general de la calidad de las predicciones en el período de validación, lo cual es muy normal en el ámbito de la modelización hidrológica.

Finalmente, es importante mencionar que el caudal medio diario (observado) de la zona de estudio difiere significativamente en los períodos de calibración y validación, lo cual sugiere un potencial cambio del régimen de caudales en función del tiempo, que justificaría el bajo rendimiento en el período de validación. Sin embargo, con el uso del filtro de Kalman se reduce este efecto al permitirle al modelo incorporar las nuevas características de las observaciones más recientes.

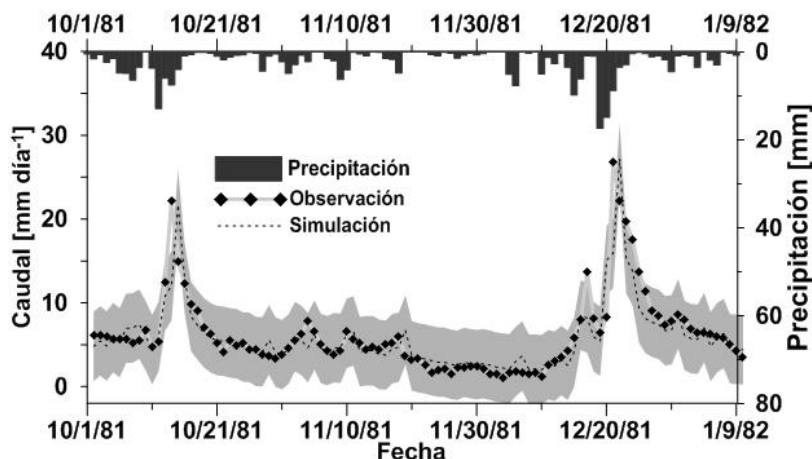


Figura 3. Predicciones para la cuenca del Tomebamba en Matadero: hidrograma de caudales observados y simulados (calibración).

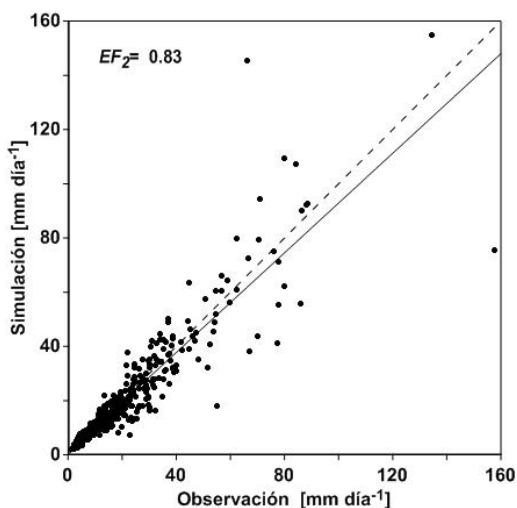


Figura 4. Diagrama de dispersión para el período de validación.

5. CONCLUSIONES

El sistema de modelización DBM se ha sometido a un proceso de evaluación de su capacidad de predicción considerando datos de una subcuenca altoandina. El modelo DBM combina características de un modelo conceptual con las de un modelo caja negra de minado de datos.

Las predicciones obtenidas con el modelo DBM sugieren que la estructura DBM, por sus características intrínsecas, es capaz de adaptarse a los datos de la subcuenca de estudio permitiendo inclusive una interpretación física en función de la descomposición de la TF.

Así, el presente estudio reveló que el rendimiento del modelo DBM fue aceptable con los registros disponibles para el presente estudio; se ajustó aceptablemente al caudal observado en el período de calibración sin reducir severamente su rendimiento durante el período de validación. En principio, esto alienta a pensar que si se dispone de un período adecuado de registros de unos pocos años, el sistema de modelización DBM se podría utilizar como una herramienta confiable para diferentes análisis tales como el relleno de vacíos en los registros observados y el pronóstico de caudales en sistemas en tiempo real.

Finalmente, el uso del filtro de Kalman, como técnica de asimilación de datos, mejora la capacidad predictiva del modelo DBM además de proporcionar la incertidumbre de las predicciones,

lo cual constituye una herramienta trascendental en la toma de decisiones en el caso de eventos extremos tales como inundaciones.

AGRADECIMIENTOS

El presente artículo se desarrolló en el contexto de la tesis de grado del primer autor (Ingeniería Civil, UC), y de los proyectos dirigidos por el segundo autor: (i) "Establecimiento de modelos numéricos para casos selectos de la gestión de recursos hídricos", financiado por la UC; y (ii) "Gestión de datos y modelización hidrológica para Soporte al Pronóstico de Alerta Temprana del Sistema Paute Integral", financiado por la Corporación Eléctrica del Ecuador (CELEC EP) y la UC. Los autores quisieran agradecer a las instituciones que facilitaron la información hidrometeorológica para llevar adelante el presente estudio, incluyendo a la Corporación Eléctrica del Ecuador (CELEC EP) y al Instituto Nacional de Meteorología e Hidrología (INAMHI).

REFERENCIAS

- Beven, K., 2004. *Rainfall - Runoff Modelling: The Primer*. John Wiley & Sons, Ltd., 457 pp.
- Box, G.E.P., G.M. Jenkins, 1976. *Time Series Analysis, Forecasting And Control* (2nd ed.). San Francisco: Holden Day.
- CONELEC. 2009. Estadística del sector eléctrico ecuatoriano. Quito-Ecuador, 388 pp. Disponible en http://www.conelec.gob.ec/images/documentos/doc_10048_Boletin%202009.pdf.
- CONELEC. 2011. Estadística del sector eléctrico ecuatoriano. Quito-Ecuador, 189 pp. Disponible en http://www.conelec.gob.ec/images/documentos/doc_10104_Be2011.pdf.
- Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. *Transactions of the ASME-Journal of Basic Engineering*, 82, 35-45.
- Lees, M.J., 2000a. Data-based mechanistic modelling and forecasting of hydrological systems. *Journal of Hydroinformatics*, 2(1), 15-34.
- Lees, M., 2000b. *Advances in transfer function based flood forecasting*. In Flood forecasting: What does current research offer the practitioner? In: Lees, M.J., P. Walsh (Eds.). BHS Occasional Paper No. 12, 7, 40-54. Produced by the Centre for Ecology and Hydrology on behalf of the British Hydrological Society.
- Lin, Z., 2003. *Modeling environmental systems under uncertainty: Towards a synthesis of data-based and theory-based models*. PhD-Thesis, Athens-Georgia: The University of Georgia.
- Mwakalila, S., P. Campling, J. Feyen, G. Wyseure, K. Beven, 2001. Application of a data-based mechanistic modelling (DBM) approach for predicting runoff generation in semi-arid regions. *Hydrological Processes*, 15, 2281-2295.
- Nelder, J., R. Mead, 1965. A simplex method for function minimization. *The Computer Journal*, 7, 308-313.
- Pedregal, D., C.J. Taylor, P.C. Young, 2007. *System identification, time series analysis and forecasting*. CAPTAIN handbook: The Captain Toolbox.
- Refsgaard, J., J. Knudsen, 1996. Operational validation and intercomparison of different types of hydrological models. *Water Resources Research*, 32, 2189-2202.

- Romanowicz, R., P.C. Young, K. Beven, F. Pappenberger, 2008. A data based mechanistic approach to nonlinear flood routing and adaptive flood level forecasting. *Advances in Water Resources*, 31(8), 1048-1056.
- Vázquez, R., 2003. *Assessment of the performance of physically based distributed codes simulating medium size hydrological systems*. PhD-Thesis, Katolieke Universiteit Leuven, Belgium.
- Wheater, H., A.J. Jakeman, K. Beven, 1993. *Modelling change in environmental systems*. Edited by A.J. Jakeman, M.B. Beck, J.M. McAleer: 101-132.
- Wheater, H., S. Sorooshian, I.K.D. Sharma, 2008. *Hydrological modelling in arid and semi-arid areas*. Cambridge: University Press.
- Young, P., 1998. Data-based mechanistic modelling of environmental, ecological, economic and engineering systems. *Environmental Modelling and Software*, 13, 1867-1874.
- Young, P., 2002a. *Data-based mechanistic and top-down modelling*. In: Proceedings International Conference on Integrated Assessment and Decision Support (iEMSs2002), 24-27.
- Young, P., 2001. *Data-based Mechanistic modelling and validation of rainfall-flow processes*. In: Model Validation: Perspectives in Hydrological Science. Eds. Malcolm G.A., P.D. Bates, 117-161. Chichester: John Wiley & Sons, Ltd.
- Young, P., 2002b. Advances in real-time flood forecasting. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 360, 1433-1450.
- Young, P., 2013. Hypothetico-inductive data-based mechanistic modeling of hydrological systems. *Water Resources Research*, 1-67.
- Young, P., C.M. Tomlin, 2000. Data-based mechanistic modelling and adaptive flow forecasting. In: Flood forecasting: What does current research offer the practitioner? (Eds. Lees, M.J., P. Walsh) BHS Occasional Paper No. 12, 25-39. Produced by the Centre for Ecology and Hydrology on behalf of the British Hydrological Society.