

Evaluación de la incertidumbre en la simulación de caudales en puntos no aforados con un modelo distribuido y mediante un procesador estocástico

Juan Camilo Múnera¹, Félix Francés¹, Ezio Todini², Gabriele Coccia²

⁽¹⁾Instituto de Ingeniería del Agua y el Medio Ambiente, Universidad Politécnica de Valencia

⁽²⁾Departamento de Ciencias de La Tierra y Geológico Ambiental, Universidad de Bolonia

juancmunera@yahoo.es

1 RESUMEN

En este trabajo se presenta una metodología orientada a la evaluación de incertidumbre asociada a la estimación de caudales en puntos no aforados de una cuenca mediante el procesamiento a posteriori de los resultados de simulación de un modelo hidrológico distribuido. Para ello se han probado dos estrategias diferentes:

La primera de ellas se basa en la adaptación de un pos-procesador estadístico de incertidumbre de tipo bayesiano denominado *Model Conditional Processor* (MCP) propuesto por Todini (2008), para su aplicación en las estaciones de aforo con información histórica. El MCP permite evaluar la incertidumbre predictiva (IP) asociada a las predicciones realizadas con uno o más modelos mediante la estimación de una función de densidad de probabilidad (fdp) de la variable de interés -el nivel o caudal observado en el río-, condicionada a los valores simulados con los modelos que hacen referencia a la misma variable. El MCP se ha aplicado a varias estaciones de aforo en la cuenca del Río Illinois, que fueron incluidas en el proyecto de intercomparación de modelos hidrológicos distribuidos (DMIP2) del NWS, utilizando el modelo TETIS como modelo hidrológico de previsión. En este trabajo, se propone una modificación del MCP, que consiste en ajustar fdp paramétricas a las series de caudal observadas y simuladas en lugar de las fdp empíricas propuestas en la formulación original, y su combinación con una técnica de regionalización, en la que se han explorado algunas relaciones funcionales que permitan extrapolar a cualquier punto no aforado de la cuenca, los parámetros de las fdp de los caudales observados y algunos parámetros adicionales requeridos para describir la fdp conjunta entre ambas variables. Adicionalmente, se han utilizado fdp conjuntas bivariadas truncadas, buscando mejorar la adaptación al separar los datos en dos muestras correspondientes a los caudales bajos y altos.

La segunda estrategia se basa en la utilización de dos filtros de Kalman acoplados en paralelo para representar el comportamiento dinámico de un sistema lineal discreto, mediante un algoritmo conocido como MISP (Todini, 1978). El primer filtro se ocupa de la estimación con mínima varianza del estado del sistema dado un conjunto de parámetros, y el segundo filtro se encarga de actualizar los parámetros de la matriz de transición de estado del primer filtro, también con mínima varianza. Las variables que se han incluido en el vector de estado son los caudales observados y simulados en una estación de aforo utilizada como pivote y los caudales simulados en el punto de interés que se supone sin aforo, que en este caso se ha hecho coincidir con una estación al interior de la cuenca, para poder comprobar la eficacia del algoritmo.

Los resultados obtenidos indican que ambas metodologías constituyen herramientas robustas y muy útiles para estimar la incertidumbre asociada a las estimaciones realizadas con el modelo hidrológico, y en algunos casos para mejorar tales estimaciones, tanto en los puntos de aforo, como en puntos no aforados de la cuenca.

2 Introducción

En la práctica hidrológica resulta fundamental asignar una medida de la incertidumbre a la estimación de caudales, o de cualquier otra variable de estado, efectuada con cualquier modelo hidrológico, o en otras palabras, establecer con qué probabilidad se puede verificar la ocurrencia real de un caudal de cierta magnitud, condicionado a la estimación que se ha realizado sobre éste con el modelo o modelos de previsión. Esta aproximación probabilística del problema deriva del hecho de que cualquier modelo hidrológico, independiente de su tipología (modelos conceptuales, físicamente basados, basados en los datos, etc.), puede tener mayor o menor grado de acierto sobre la variabilidad temporal de la serie histórica de caudales, y más específicamente, sobre las diferentes características de un evento de crecida, como la magnitud y el tiempo al pico del hidrograma, así como la forma de la rama de recesión al finalizar la tormenta causante del evento.

En efecto, cada modelo hidrológico requiere unos datos de entrada y tiene una estructura específica con un cierto número de parámetros agregados o distribuidos espacialmente para simular los diversos procesos hidrológicos en la cuenca. En función de lo anterior, cada modelo puede tener mayor o menor habilidad para representar los mecanismos de respuesta de la cuenca al ocurrir una tormenta, que en muchos casos presenta un comportamiento altamente no lineal.

La calidad de una simulación o predicción realizada con un modelo hidrológico dependerá de la información disponible en para representar la variabilidad espacio-temporal de las variables hidrometeorológicas, así como de la capacidad del modelo para propagar en el tiempo el estado de humedad en la cuenca, responsable en buena medida de la no linealidad, y del tipo de conceptualización y parametrización realizada de los procesos hidrológicos dominantes. Además de lo anterior, en muchos casos adquiere especial relevancia la propagación de la escorrentía en la red de cauces, que se puede modelar mediante relaciones simples basadas en hidrogramas unitarios lineales, o mediante formulaciones más sofisticadas como la que se ha implementado en el modelo TETIS (Francés et al, 2007), en la cual se solucionan las ecuaciones de la onda cinemática parametrizadas de acuerdo a una caracterización topológica y geomorfológica de la red de drenaje, que involucra conceptos de la teoría de Geometría Hidráulica (Leopold y Maddock, 1953).

En modelos de tipo hidráulico serán importantes otros aspectos, como la definición de las condiciones de contorno, la geometría del cauce y la llanura inundable, y la caracterización de la resistencia al flujo, etc. En otras tipologías de modelos basados en los datos, como en aquellos que usan Redes Neuronales Artificiales, será muy importante definir la arquitectura óptima de la Red, así como las variables de entrada y sus rezagos temporales, las funciones de activación y los algoritmos de entrenamiento y verificación, etc.

Independiente del tipo de modelo o estrategia de modelación adoptada, existen varias técnicas para abordar el problema de la incertidumbre en la estimación de caudales en alguna sección de un Río o de cualquier otra variable de interés, como el nivel de agua, o el volumen en un embalse. En este sentido, algunos métodos como el GLUE (Beven and Binley, 1992; Beven and Freer, 2001), el cual es una extensión del General Sensitivity Analysis (Hornberger and Spear, 1980) analizan la sensibilidad del modelo ante todo el espectro de posibles valores que puede adoptar el espacio de parámetros, o ante las variaciones de los datos de entrada. Otras aproximaciones intentan descomponer las diversas fuentes de incertidumbre que aparecen en el proceso de modelación, para propagar cada una de ellas en el mismo mediante técnicas de simulación tipo Montecarlo, las cuales conducen a la generación de cuantiles de la variable simulada en cada paso de tiempo. No obstante, este tipo de aproximaciones hacen referencia a la Incertidumbre de validación del modelo (Todini, 2009) y no a la incertidumbre predictiva (IP), al no considerar explícitamente la fdp de los caudales observados condicionados a las simulaciones o predicciones realizadas con el modelo.

Krzysztofowicz (1999) introduce un procesador de incertidumbre hidrológica denominado Hydrologic Uncertainty Processor (HUP) orientado a la evaluación de la IP asociada a las predicciones hidrológicas, creando las bases para su estimación objetiva (Todini, 2008; Coccia y Todini, 2010). Más recientemente, se han hecho algunos intentos de combinar las predicciones realizadas con varios tipos de modelos mediante una técnica conocida como Bayesian Model Averaging (BMA) (Raftery et al, 2003; Raftery et al, 2005; Vrugt y Robinson, 2007). A partir de estas dos formulaciones, Todini (2008) propone una metodología Bayesiana, denominada Model Conditional Processor (MCP), basada en una aproximación multi-Normal para estimar la IP. Esta última metodología puede ser vista como una extensión de la aproximación Bayesiana HUP de Krzysztofowicz (1999), así como una generalización del método BMA de Raftery et al (2003).

3 Caso de estudio

Como caso de estudio se ha seleccionado las cuenca de los Ríos Baron Fork, Elk river e Illinois, localizadas en los estados de Arkansas y Oklahoma de EE.UU, y que hacen parte de la segunda fase del proyecto de intercomparación de modelos hidrológicos distribuidos (DMIP2) (Smith et al, 2004), el cual surgió del interés del National Weather Service (NOAA/NWS) en incorporar este tipo de modelos en la predicción de avenidas para la gestión de emergencias por inundación. La estación de Tahlequah en el Río Illinois es la de mayor superficie de drenaje entre estas cuencas cubriendo unos 2484 km² (Figura 1). Al interior de las cuencas hay algunos puntos de aforo adicionales que permiten estudiar la variabilidad espacio temporal de los caudales generados en esta región.

Se han elegido estas cuencas dado que para cumplir los objetivos del proyecto DMIP2, el Laboratorio de Hidrología de NOAA-NWS ha disponibilizado una gran cantidad de información cartográfica de características físicas y ambientales de las cuencas, así como las series históricas de las principales variables hidrometeorológicas, como son los caudales en algunas estaciones de aforo y los campos de precipitación en la región de estudio estimados a partir de Radar, ambos con resolución temporal horaria. La ETP y la Temperatura para el mismo período de tiempo (01/10/1995-30/09/2002) se han obtenido del Reanalysis (NCEP/NCAR) con resolución temporal de 3 horas. Las series de caudal de las estaciones restantes fueron obtenidas del archivo de datos instantáneos IDA del USGS.

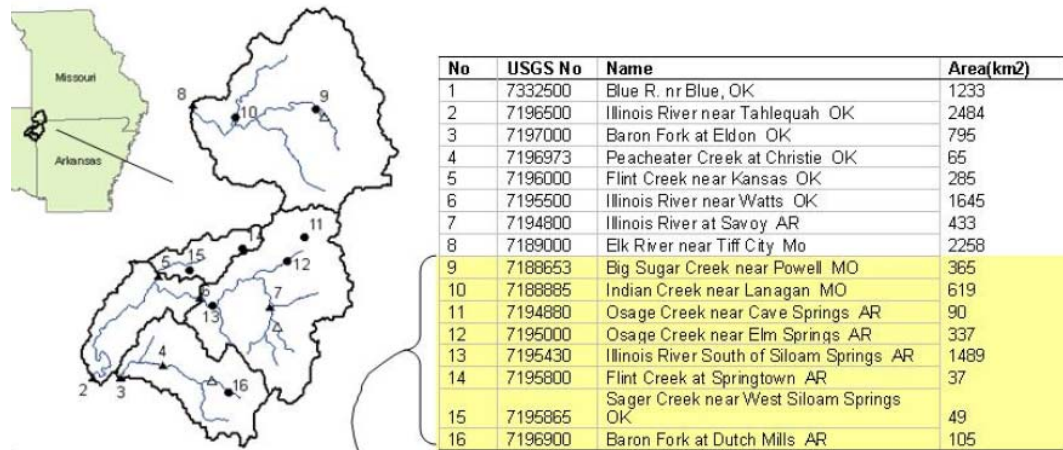


Figura 1 Cuencas de estudio del proyecto DMIP2 en la región de Oklahoma (Fuente: Proyecto DMIP2, NOAA)

El modelo hidrológico utilizado en el estudio es el modelo de tipo distribuido TETIS (Vélez et al, 2002; Francés et al, 2007), desarrollado en la Universidad Politécnica de Valencia.

TETIS es un modelo con base física de tipo conceptual y distribuido en el espacio, capaz de simular la respuesta hidrológica en cualquier punto del dominio espacial de la cuenca. El modelo tiene la capacidad de simular los procesos hidrológicos más relevantes, incluyendo la variabilidad espacial de cada componentes del balance, y se puede utilizar a cualquier escala temporal. Además, tiene una novedosa estructura de parámetros, haciendo una distinción entre la variabilidad espacial de las características físicas a escala de cuenca, representadas en mapas de parámetros estimados a priori a partir de la información medioambiental disponible, y en los parámetros efectivos que se deben emplear en el modelo a escala de celda para tener en cuenta las posibles fuentes de incertidumbres en su estimación y los efectos de escala espacial y temporal. La relación entre los mapas de parámetros estimados a priori y los mapas efectivos se considera como una función de corrección, que en su forma más simple, constituye un factor corrector (FC) que modifica globalmente los primeros. Esta estructura diferenciada reduce drásticamente el número de parámetros a calibrar al conjunto de FC posibilitando su calibración automática (Vélez, 2003; Vélez and Francés, 2004; Francés et al, 2007; Vélez et al, 2009).

Para la evaluación de la incertidumbre asociada a las simulaciones continuas realizadas a escala horaria con el modelo TETIS en puntos no aforados de la cuenca de estudio se proponen dos aproximaciones diferentes, las cuales se describen en los apartados a continuación.

3.1 Procesador de incertidumbre MCP en puntos de aforo

La primera propuesta metodológica está basada en la aplicación del pos-procesador bayesiano de incertidumbre, MCP (Todini, 2008). El MCP permite establecer en las estaciones de aforo, donde se conocen las series históricas del caudal o nivel de agua, la función de densidad de probabilidad (fdp) de alguna de estas variables condicionada a toda la información disponible, normalmente el caudal simulado por un modelo hidrológico. En esta aproximación, primero se hace una transformación no paramétrica de las observaciones y de las simulaciones del modelo a un espacio multi-Normal mediante la operación *Normal Quantil Transform* (NQT) (Kelly y Krzysztofowicz, 1997), con la intención de generar nuevas variables análogas en el campo transformado, cada una de ellas con fdp marginal Normal Estándar (Todini, 2008; Coccia y Todini, 2010).

En la transformación NQT, la probabilidad asignada a cada valor de una serie en el campo original será la misma en el campo gaussiano, siendo esta probabilidad el vínculo de cada variable considerada en los dos campos.

Para asignar esta probabilidad, en la versión actual del MCP se recurre a una técnica no paramétrica, que consiste en ajustar a la serie histórica de cada variable en el campo original una función de distribución de probabilidad (FDP) empírica del tipo Weibull Plotting Position. De este modo, las variables originales y y \hat{y} son transformadas a sus análogas η y $\hat{\eta}$, respectivamente, las cuales tienen distribución Normal Estándar $N(0,1)$. De este modo, la relación entre las variables originales y sus valores transformados es:

$$P(y \leq y_i) = \frac{i}{n+1} = P(\eta \leq \eta_i), \quad i = 1, 2, \dots, n$$

Donde n es el número de datos históricos disponible, e i es el posición ocupada por el valor y_i en la serie ordenada en orden creciente.

El uso de la distribución empírica Weibull plotting Position implica la necesidad de ajustar un modelo complementario en el campo original a las colas de la distribución de cada variable, es decir, tanto a las observaciones como a las predicciones del modelo, para cubrir los cuantiles con probabilidad mayor al máximo valor registrado en la muestra con probabilidad $\left(\frac{n}{n+1}\right)$, o menores al valor mínimo con probabilidad $\left(\frac{1}{n+1}\right)$.

Dado que en este caso de estudio se busca poder extrapolar la metodología para estimar la incertidumbre en puntos no aforados, en lugar de utilizar fdp empíricas, para asignar esta probabilidad se han probado algunas fdp paramétricas, normalmente de 3 parámetros, buscando obtener un buen ajuste y representar adecuadamente la componente asimétrica en la distribución de los caudales. Las distribuciones que han sido probadas son la *Normal Generalizada* (GNO), la *Logística Generalizada* (GLO), la *Generalizada de Pareto* (GPA), la *Pearson tipo III* (PE3) y la *General Extreme Value* (GEV). Tras algunas pruebas de ajuste realizadas se ha adoptado la distribución GEV, con parámetros estimados por el método de los *L-momentos* (Hosking, 1990; Hosking y Wallis, 1995; Hosking y Wallis, 1997). A continuación se presentan las ecuaciones para calcular la FDP de la GEV y su inversa para estimación de cuantiles:

$$F = \exp\left[-\left\{1 - k(x - \xi)/\alpha\right\}^{1/k}\right]$$

$$x = \xi + \frac{\alpha\left\{1 - (-\log F)^k\right\}}{k}$$

Los parámetros ξ , α y k se pueden obtener analíticamente a partir de los 3 primeros L-momentos (L1, L2 y T3), los cuales se estiman a partir de la muestra de datos históricos.

En el caso simple de un único modelo de previsión, una vez realizada la transformación de las variables correspondientes a los caudales observados y simulados al campo gaussiano, se hipotiza que en éste existe una relación lineal entre las variables, es decir, se asume que su distribución conjunta $f(\eta, \hat{\eta})$ es de tipo Normal Bivariada. La media y la varianza de esta fdp conjunta se puede expresar como:

$$\mu_{\eta, \hat{\eta}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Sigma_{\eta, \hat{\eta}} = \begin{bmatrix} 1 & \sigma_{\eta, \hat{\eta}} \\ \sigma_{\eta, \hat{\eta}} & 1 \end{bmatrix}$$

Puesto que las marginales de las variables transformadas η y $\hat{\eta}$ son Normales estándar $N(0,1)$, la covarianza entre las mismas es equivalente al coeficiente de correlación, y la matriz de covarianza se puede expresar como:

$$\Sigma_{\eta, \hat{\eta}} = \begin{bmatrix} 1 & \rho_{\eta, \hat{\eta}} \\ \rho_{\eta, \hat{\eta}} & 1 \end{bmatrix}$$

Aplicando el teorema de Bayes en este campo gaussiano, a partir de la fdp conjunta Normal bivariada y de la fdp marginal del caudal simulado $N(0,1)$, es posible deducir la fdp del caudal observado condicionada al valor del caudal simulado por el modelo en cualquier paso de tiempo $f(\eta/\hat{\eta} = \hat{\eta}^*)$, es decir, la fdp predictiva en el campo gaussiano; esta fdp condicionada también resulta ser una fdp $N(\mu, \sigma)$. Los dos momentos de esta fdp dependen únicamente del coeficiente de correlación calculado entre las dos variables en el campo gaussiano:

$$\begin{aligned} \mu_{(\eta/\hat{\eta}=\hat{\eta}^*)} &= \rho_{\eta, \hat{\eta}} \cdot \hat{\eta}^* \\ \sigma_{(\eta/\hat{\eta}=\hat{\eta}^*)}^2 &= 1 - \rho_{\eta, \hat{\eta}}^2 \end{aligned}$$

Dada la no linealidad del proceso de transformación NQT, para convertir esta fdp condicionada en el campo gaussiano a su equivalente en el campo original, es necesario discretizarla en un número suficientemente grande de intervalos de probabilidad, y posteriormente convertir los cuantiles muestrales resultantes al campo original mediante el proceso de transformación inversa NQT^{-1} . Una vez realizada la transformación inversa, es posible calcular el valor esperado y las bandas de incertidumbre para el nivel de significancia requerido. Un desarrollo más detallado de la metodología y sus ecuaciones se puede consultar en Todini (2008), Coccia y Todini (2010), Múnera et al (2010).

El proceso descrito anteriormente no tiene en cuenta el hecho de que muchos modelos presentan un comportamiento diferente para reproducir los caudales bajos y medios, de los caudales altos asociados a las crecidas. Adicionalmente, es normal que los primeros se presenten con una frecuencia mucho mayor que los segundos, razón por la cual tienen un peso mucho mayor en la muestra, afectando el cálculo del coeficiente de correlación, al considerar el conjunto completo de datos. Por esta razón, en muchos casos, la parte superior de la muestra, que tiene gran interés al concentrar los máximos caudales de los eventos registrados en la serie histórica, no está bien representada al ajustar el modelo estadístico a toda la muestra de valores.

Una solución que se ha adoptado para estos casos es dividir la serie histórica en dos muestras representativas de ambos conjuntos y ajustar a cada una de ellas una fdp conjunta bivariada truncada, buscando mejorar la adaptación al separar los datos en dos muestras correspondientes a los caudales bajos y altos, respectivamente. En la Figura 2 se ilustra el efecto de la separación en muestras en la estación de aforo Baron Fork para el período de calibración (01/10/1995-30/09/2002).

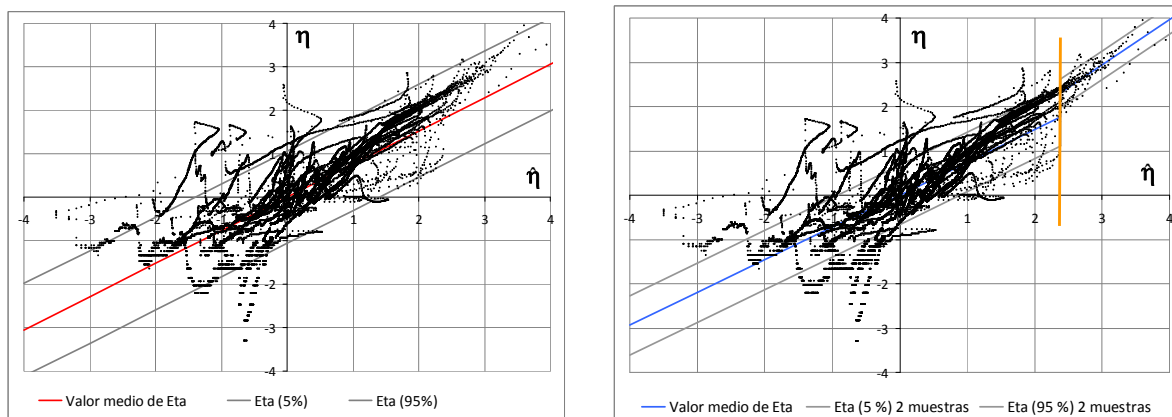


Figura 2 Diagrama de dispersión en el campo Normal de los datos observados y simulados con el modelo TETIS en la cuenca Baron Fork. (Izq.) Valor esperado y cuantiles del 5 y 95 % utilizando todos los datos. (Der.) Valor esperado y los mismos cuantiles al separar los datos en dos muestras asignadas a distribuciones bivariadas truncadas.

Con la separación en dos muestras se consigue hacer una descripción más realista de la incertidumbre predictiva en ambos conjuntos de datos. Las ecuaciones y parámetros de la distribución conjunta Normal truncada se describen en detalle en Coccia y Todini (2010).

Esta formulación se puede extender a simulaciones realizadas con más de un modelo hidrológico o de cualquier tipología, tanto para el caso original sin división en muestras, como para el caso de dividir en dos fdp truncadas; en estos casos, se asume que la fdp conjunta en el campo multinormal es de tipo Normal multivariada o Meta-gaussiana (Todini, 2008).

3.2 Regionalización de parámetros para aplicación del MCP en puntos no aforados

Por otro lado, en este trabajo se ha pretendido extender esta formulación a puntos de la cuenca no aforados. Para esto es necesario conocer la distribución de los caudales observados, cuyos parámetros se pueden inferir a partir de los primeros 3 L-momentos. Además de lo anterior, se requiere estimar el coeficiente de correlación entre los caudales observados y simulados en el campo gaussiano, único parámetro que caracteriza la distribución conjunta entre las dos variables cuando no se realiza la separación en muestras. En este caso, se ha decidido hacer la separación en dos muestras, por lo que es necesario en lugar del coeficiente de correlación, utilizar las covarianzas entre observados y simulados y las medias y varianzas muestrales en los dos conjuntos.

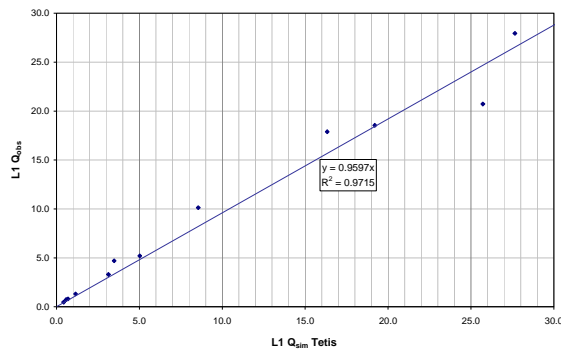
Para estimar los 3 parámetros de la distribución GEV para la serie observada en los puntos no aforados y los 6 parámetros que permiten representar las fdp bivariadas truncadas en las cuencas se ha recurrido a una regionalización de estos 9 parámetros mediante la cual se han construido diferentes expresiones funcionales, utilizando como variables independientes el área de drenaje de las subcuencas y la media de los caudales simulados. Para hacer la división en muestras se ha utilizado un umbral común para todas las cuencas $\hat{\eta} = 2.5$.

En este análisis regional se han utilizado los datos históricos y de las simulaciones realizadas con TETIS correspondientes a las 12 estaciones de aforo con información disponible en la cuenca de estudio. En la Figura 3 se presentan las expresiones regionales ajustadas para las cuencas de estudio a los parámetros ya mencionados.

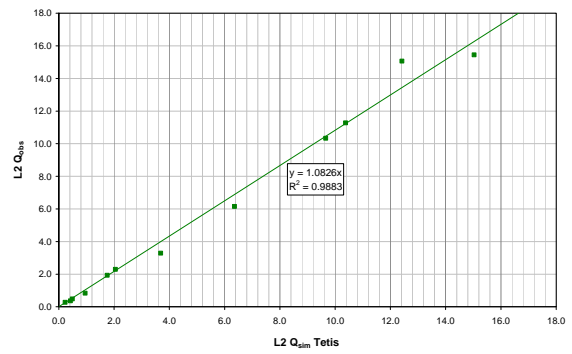
En la Tabla 1 se presentan los resultados de dos índices estadísticos que han sido evaluados en las 12 estaciones de aforo para el período de calibración: el primero es el valor esperado del error, y el segundo es el índice de eficiencia de Nash-Sutcliffe. Ambos índices se han calculado para la estimación realizada con el modelo Tetis, para el valor esperado de la estimación al aplicar el MCP a los datos históricos registrados, y finalmente, el valor obtenido con los parámetros que resultan de las expresiones regionales. Los resultados obtenidos en el período de validación son muy similares a los reportados en esta tabla.

Tabla 1 Valor esperado del error de estimación e Índice de Eficiencia de Nash – Sutcliffe (NSE) obtenidos para las cuencas de estudio en el período de calibración

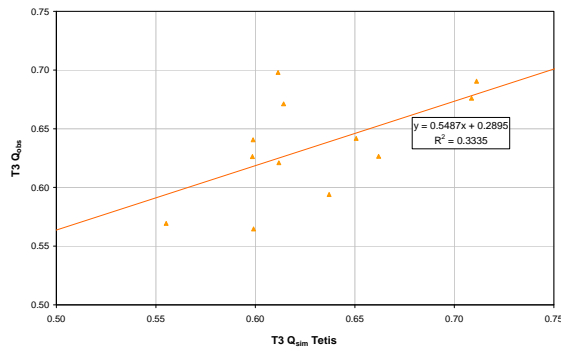
Estación de Aforo	Área (km ²)	E(error) TETIS	E(error) TETIS + MCP	E(error) TETIS+MCP (REG)	NSE TETIS	NSE TETIS + MCP	NSE TETIS + MCP (REG)
(2) Ill. Tahlequah	2413	-0.29	0.39	0.15	0.84	0.84	0.79
(3) B.F. Eldon	830	-1.58	0.70	0.34	0.81	0.82	0.83
(4) Peachheater Cr.	65.1	-0.07	0.03	0.05	0.70	0.73	0.66
(5) Flint Cr. (K)	284.1	-0.17	0.34	0.07	0.62	0.71	0.68
(6) Ill. Watts	1541	0.66	0.00	-0.16	0.80	0.80	0.81
(7) Ill. Savoy	418.4	-0.18	0.00	0.15	0.66	0.80	0.67
(8) Elk river	2258	5.00	-0.02	0.31	0.72	0.78	0.66
(12) Osage Cr.	303.5	-1.22	0.23	0.12	0.56	0.55	0.69
(13) Ill. Siloam Sp.	1452	-1.56	0.40	0.23	0.75	0.74	0.80
(14) Flint Cr. (AR)	36.8	-0.04	0.07	0.05	0.73	0.66	0.68
(15) Sager Cr.	48	-0.15	0.06	0.07	0.50	0.43	0.39
(16) B.F. Dutch	103.4	-0.15	0.03	0.21	0.34	0.36	0.18



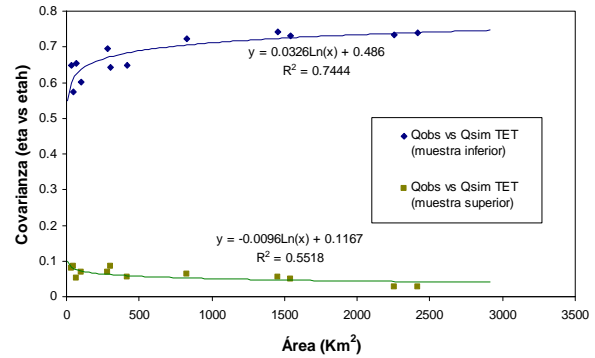
(a)



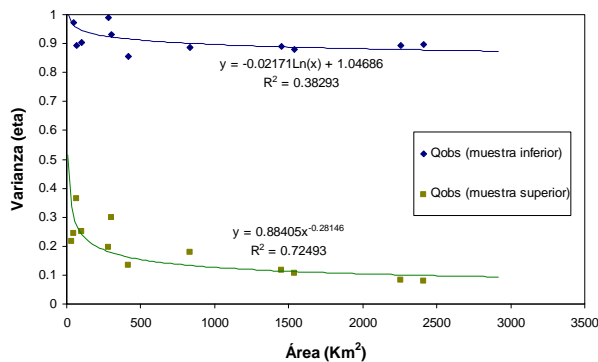
(b)



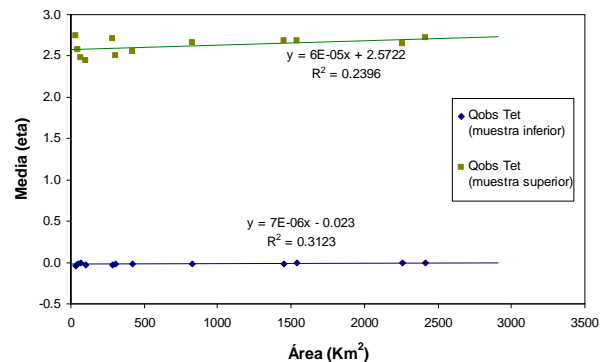
(c)



(d)



(e)



(f)

Figura 3 Expresiones regionales para aplicación indirecta del MCP en puntos no aforados de las cuencas de estudio. En los paneles (a), (b), y (c) se presentan las funciones de regresión ajustadas para estimar los 3 primeros L-momentos ajustados a las series observadas. En los paneles (d), (e) y (f) se presentan las funciones correspondientes para estimar los parámetros de las dos distribuciones conjuntas bivariadas truncadas en el campo gaussiano.

En la actualidad se están explorando otras relaciones de este tipo involucrando otros modelos de previsión en una aproximación multi-modelos, con la que se busca obtener unas relaciones funcionales más robustas y reducir la incertidumbre en la estimación de los parámetros de regresión.

En la Figura 4 se presentan los resultados del MCP con los parámetros estimados con las relaciones regionales en la cuenca de tamaño medio Osage Creek para dos eventos ocurridos en junio de 2000. Se muestran los caudales observados en la estación, así como los simulados con el modelo TETIS, el valor esperado obtenido con el MCP y las bandas de incertidumbre obtenidas con la fdp predictiva. En estas últimas series se aprecia una ligera discontinuidad producto de la separación en dos muestras. En ambos eventos se observa una ligera mejora en el desfase temporal entre los hidrogramas observados y simulados; no obstante, en el primero de ellos hay una desmejora en la magnitud del pico respecto al modelo, mientras que en el segundo se aprecia una mejora substancial. Por otro lado, se evidencia que la mayor parte de los caudales observados caen dentro de la banda.

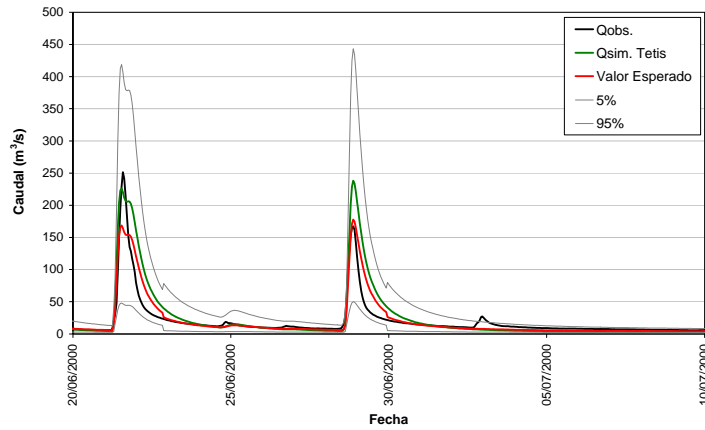


Figura 4 Resultados de la formulación del MCP con parámetros a partir de las funciones regionales ajustadas aplicado a la cuenca Osage Creek, afluente del Río Illinois. Caudal observado, caudal simulado con TETIS, valor esperado calculado con el MCP y Bandas de incertidumbre del 90 %.

3.3 Filtros de Kalman acoplados en paralelo (*Mutually Interactive State Parameter Estimation, MISP*)

La segunda propuesta para evaluar la incertidumbre en la estimación de caudales con el modelo hidrológico en puntos no aforados contempla la utilización de filtros de Kalman mediante la técnica conocida como MISP (Todini, 1978). MISP hace uso de dos filtros acoplados en paralelo para representar el comportamiento dinámico de un sistema lineal discreto. El primer filtro se ocupa de la estimación con mínima varianza del estado del sistema x_t dado un conjunto de parámetros θ_t ; el segundo filtro se encarga de actualizar los parámetros de la matriz de transición de estado del primer filtro, también con mínima varianza, utilizando el estado estimado en el paso de tiempo actual y en el anterior. En cada momento se tiene en cuenta la interacción entre ambos filtros para conseguir una estimación óptima del estado del sistema y de los parámetros, que en esencia equivale a resolver un problema de tipo no lineal.

La formulación matemática completa del modelo MISP se puede consultar en el trabajo de Todini (1978). Las ecuaciones del sistema dinámico y de medición para aplicación del filtro son, respectivamente:

$$\begin{aligned} x_t &= \Phi_{t/t-1} x_{t-1} + \Gamma_{t-1} w_{t-1} \\ z_t &= H_t x_t + v_t \end{aligned}$$

Las ecuaciones en el espacio de los parámetros son:

$$\begin{aligned} \theta_t &= \theta_{t-1} + \Gamma_{t-1}^* w_{t-1}^* \\ z_t^* &= H_t^* \theta_t + v_t^* \end{aligned}$$

Donde los elementos del vector θ_t son los elementos no nulos y no constantes incluidos en la matriz de transición de estado $\Phi_{t/t-1}$.

En la configuración del vector de estado se deben relacionar cantidades que sean medibles a intervalos discretos de tiempo. Las variables que se han incluido en el vector de estado son los caudales observados y simulados en una estación de aforo utilizada como pivote y los caudales simulados en el punto de interés sin aforo. Adicionalmente, también se incluyen las mismas variables en el paso de tiempo anterior, para tener en cuenta la estructura de autocorrelación en el proceso. Previamente a la ejecución del filtro se ha hecho una transformación logarítmica de los datos, buscando conseguir una mejor aproximación de la hipótesis de ruido blanco, tanto en los errores del sistema como en los de medición. Como resultado del filtro se obtiene una estimación de mínima varianza del caudal en el punto de interés y la varianza asociada a la estimación.

Una vez obtenidos los resultados del filtro, se lleva a cabo un pos proceso análogo al descrito del MCP para evaluar la incertidumbre de los caudales estimados con el filtro en relación a los simulados con el modelo; en este caso no es necesario pasar por la transformación NQT porque las variables fueron previamente transformadas al campo logarítmico. En este pos-proceso se han hipotizado fdp Lognormales para las

marginales, y una fdp Lognormal bivariada para la conjunta,. Aplicando una vez más el teorema de Bayes se obtiene la fdp condicionada, que también resulta ser de tipo Lognormal. En este caso se ha hecho una superposición de la desviación típica resultante de la fdp condicionada y de la desviación típica de la estimación realizada con el filtro para tener una descripción completa de la incertidumbre. A partir de la fdp resultante se obtiene el valor esperado y las bandas de incertidumbre para el nivel de significancia establecido.

En la Figura 5 se presentan los resultados obtenidos con el Filtro de Kalman aplicando la técnica MISP en la cuenca Peacheater Creek, afluente del Baron Fork, para el primer semestre del año 1999; en la figura se muestran los caudales observados y simulados con el modelo TETIS, la serie estimada con el algoritmo MISP y las bandas de incertidumbre del 90 %. En este caso, la estación de aforo Baron Fork se ha utilizado como pivote al aplicar el filtro en la cuenca interior; como resultado, se ha incrementado el valor del índice de eficiencia de Nash en el periodo de simulación de 6 años a un valor de 0.74 respecto al obtenido con el modelo TETIS (0.67).

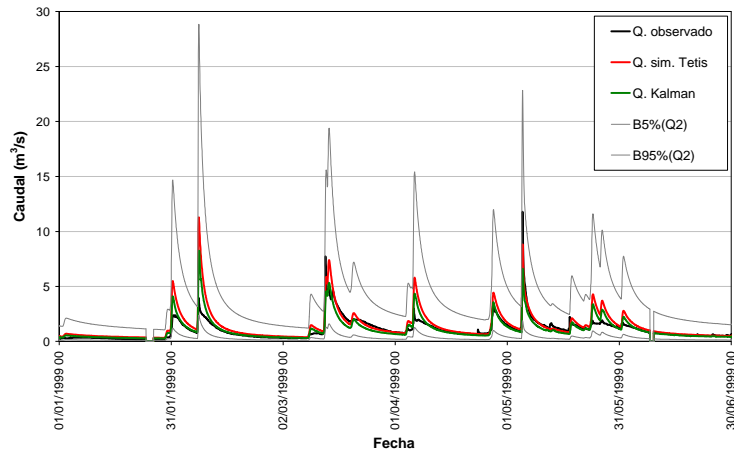


Figura 5 Aplicación del algoritmo MISP a la cuenca Peacheater Creek, afluente de la cuenca Baron Fork. Caudales observados, simulados con el modelo TETIS y con el filtro de Kalman. Período 01/01/1999-30/06/1999.

4 Conclusiones

Los resultados preliminares obtenidos con las dos formulaciones son satisfactorios, mejorando la estimación de los caudales observados respecto a los resultados obtenidos con el modelo hidrológico. Por otro lado, el tamaño de las bandas de incertidumbre está relacionado con el desempeño del modelo hidrológico en las cuencas de estudio, que se manifiesta en una mayor o menor correlación entre los caudales observados y simulados en el campo transformado.

En el caso de estudio analizado se ha encontrado que en general, se tiene mayor incertidumbre en las cuencas interiores con menor área de drenaje, lo cual tiene que ver en parte con el hecho de que las calibraciones se han realizado en las cuencas mayores, aparte de una mayor no linealidad en la respuesta.

La metodología indirecta del MCP basada en la regionalización de parámetros tiene la desventaja de que la incertidumbre inherente al proceso de regionalización no se incluye explícitamente en la evaluación de la incertidumbre de la variable a predecir, y por tanto, la incertidumbre real será mayor a la estimada. No obstante, proporciona una primera aproximación en sitios sin información histórica.

La metodología basada en filtro de Kalman ha permitido mejorar la estimación del modelo en un punto interior de la cuenca y ha permitido obtener las bandas de incertidumbre para el nivel de significancia requerido del 90%.

Finalmente, ambas formulaciones tienen aplicabilidad para el análisis de incertidumbre tanto en el caso de estimación histórica como en modo de predicción. En este último caso, se deberá incluir en el filtro de Kalman en lugar de los caudales observados (desconocidos en modo predicción) en la estación pivote, el valor esperado del caudal observado calculado previamente con el MCP, y adicionalmente, se deberá incorporar en el filtro la componente de incertidumbre asociada a la predicción de este caudal.

Agradecimientos: El presente trabajo ha sido subvencionado por el Ministerio español de Ciencia e Innovación a través de los proyectos “FloodMed” (CGL2008-06474-C02-02/BTE) y Consolider-Ingenio “SCARCE” (2010-CSD2009-00065).

Referencias bibliográficas

- Beven, K.J., Binley, A., (1992).** The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Process* 6(3), 279-298.
- Coccia G., and Todini E. (2010).** Application of a Bayesian Processor for Predictive Uncertainty Estimation in Real Time Flood Forecasting. *BHS Third International Symposium: Role of Hydrology in Managing Consequences of a Changing Global Environment*. Newcastle University, 19-23 July 2010.
- Francés F., Vélez J. I. and Vélez J. J. (2007).** Split-parameter structure for the automatic calibration of distributed hydrological models. *Journal of Hydrology*, 332, pp226– 240.
- Hornberger, G.M., Spear, R.C. (1980).** Eutrophication in Peel Inlet, I, The problem-defining behaviour and a mathematical model for the phosphorus scenario, *Water Res.* 14, 29-42.
- Hosking, J. R. M. (1990).** L-moments: analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B*, 52, 105–124.
- Hosking, J. R. M., and Wallis, J. R. (1995).** A comparison of unbiased and plotting position estimators of L-moments. *Water Resources Research*, 31, 2019–2025.
- Hosking, J. R. M., and Wallis, J. R. (1997).** Regional frequency analysis: an approach based on L-moments. Cambridge, England: Cambridge University Press.
- Kelly, K. S., and R. Krzysztofowicz. (1997),** A bivariate meta-Gaussian density for use in hydrology. *Stochastic Hydrol. Hydraul.*, 11, 17– 31.
- Krzysztofowicz, R. 1999.** Bayesian theory of probabilistic forecasting via deterministic hydrologic model. *Water Resour. Res.*, 35: 2739–2750.
- Leopold, L. B. y T. Maddock. (1953).** The hydraulic geometry of stream channels and some physiographic implications. U.S. Geol. Surv. Prof. Paper no. 252.
- Múnera J.C., Coccia G., Francés F., Todini E. (2010).** Aplicación de un procesador bayesiano para la evaluación de la incertidumbre predictiva mediante combinación de modelos hidrológicos en la predicción de crecidas en tiempo real. XXIV Congreso Latinoamericano de Hidráulica. Punta del Este, Uruguay.
- Raftery, A. E. (1993).** Bayesian model selection in structural equation models, *Testing Structural Equation Models*, Bollen, K.A., and Long, J. S. (Eds.), Newbury Park, CA, 163-180.
- Raftery, A. E., Balabdaoui, F., Gneiting, T., and Polakowski, M. (1995).** Using Bayesian model averaging to calibrate forecast ensembles, *Monthly Weather Review*, 133, 1155-1174.
- Smith, M.B., Seo, D.-J., Koren, V.I., et al. (2004).** The distributed model intercomparison project (DMIP): an overview. *Journal of Hydrology* 298(1–4), 4–26.
- Todini, E. (1978).** Mutually interactive state-parameter (MISP) estimation. Application of Kalman filter to hydrology, hydraulics and water resources. *Proc. of AGU Chapman Conference*. Univ. of Pittsburg.
- Todini, E. (2008).** A model conditional processor to assess predictive uncertainty in flood forecasting. *Intl. J. River Basin Management*. Vol. 6 (2), 123-137.
- Todini, E. (2009).** Predictive uncertainty assessment in real time flood forecasting, Ph. C. Baveye, M. Laba and J. Mysiak (eds.), *Uncertainties in Environmental Modelling and Consequences for Policy Making*. NATO Science for Peace and Security Series C: Environmental Security. Springer Netherlands, Amsterdam, NL, DOI: 10.1007/978-90-481-2636-1 9, ISBN 978-90- 481-2635-4, pp. 205-228.
- Vélez, J. I.; Vélez J. J. and Francés, F. (2002).** Modelo distribuido para la simulación hidrológica de crecidas en grandes cuencas. En congreso: XX Congreso Latinoamericano de Hidráulica. La Habana, Cuba. Publicado en CD. ISBN 959-7160-17-X.
- Vrugt, J. A. and Robinson, B. A. (2007).** Treatment of uncertainty using ensemble methods: Comparison of sequential data assimilation and Bayesian model averaging, *Water Resour. Res.*, 43, W01411, doi:10.1029/2005WR004838.