

El uso de GLUE como algoritmo de aproximación al frente de Pareto: Aplicación al modelo hidrológico de Thomas (abcd) *The use of GLUE as an approximation algorithm for the Pareto front: Application to the abcd hydrological model of Thomas*

(Recepción 28/12/2016; Aceptación 09/03/2017)

Duque Gardeazábal N.^{1,2}; Vega Viviescas C.¹; Obregón Neira N.¹

¹Grupo de Investigación en Ingeniería de los Recursos Hídricos (GIREH)

Departamento de ingeniería civil y agrícola, Universidad Nacional de Colombia-Bogotá, Colombia

²Miembro de la IAHR Bogotá Young Professional Network

Teléfono: +57 1 3165000 ext 13406

Email: nduqueg@unal.edu.co

Resumen. *Diversos problemas hidrológicos requieren la calibración del modelo a partir de datos observados. En muchos casos, los modelos deben cumplir múltiples objetivos de calibración, encontrando un conjunto aceptable de soluciones, denominada Frente de Pareto. Este artículo presenta la identificación del frente, para el modelo abcd de Thomas, en el río Combeima (Colombia), aplicando el método de calibración Generalized Likelihood Uncertainty Estimation (GLUE) con tres funciones objetivo (FO). Se encontró que al calibrar con diferentes FO el conjunto de parámetros fue diferente y con un rango amplio. Además se observó que su influencia sobre la escorrentía simulada no fue sustancial.*

Palabras clave. *Calibración multiobjetivo; Frente de Pareto; GLUE; Modelo de Thomas.*

Abstract. *Several hydrological problems require to calibrate a hydrological model from the observed data. In many cases, models must include several metrics, finding an acceptable set of solutions called the Pareto Front. This paper presents the Pareto Front identification for the abcd Thomas model, applying the GLUE calibration method with three objective functions (FO) in the Combeima river (Colombia). It was found that the calibration results with different FO were different and its range were ample. Moreover, the influence in the simulated runoff was not significant.*

Keywords. *GLUE. Multiobjective calibration. Pareto front. Thomas model.*

1. Introducción

En hidrología, los modelos matemáticos sirven para comprender los procesos hidrológicos, e intentar predecir escenarios futuros dentro de los que se incluyen la predicción y el pronóstico (Dawson, Abrahart, & See, 2007; Silberstein, 2006). Sin embargo, no siempre se conocen los parámetros (ya que estos no necesariamente están asociados con mediciones que se puedan hacer en campo), por lo que deben obtenerse a partir de una muestra de datos observados de entrada y salida, solucionando el problema inverso.

Este problema inverso, es un proceso iterativo que tiene como objeto mejorar los resultados de la simulación a través de la variación de los parámetros del modelo. Los resultados son comparados mediante el cálculo de métricas, denominadas funciones objetivo (FO) que indican de forma cuantitativa la habilidad del modelo para reproducir ciertos fenómenos hidrológicos (Dawson et al., 2007). Sin embargo, se ha resaltado por varios autores que el enfoque de encontrar un único y mejor conjunto de parámetros es erróneo (Beven & Binley, 1992; Gupta, Sorooshian, & Yapo, 1998), debido a que tanto la estructura del modelo como las mediciones utilizadas para calibrarlo tienen errores.

Cada función objetivo tiene características propias, las cuales priorizan cierto conjunto de parámetros óptimos que no necesariamente son los mismos para

otra función objetivo. En muchos casos, los modelos deben satisfacer objetivos de desempeño que pueden ser contradictorios, por esta razón, se plantea la calibración de modelos de forma multiobjetiva.

Comparado con las soluciones monoobjetivo (SO, por sus siglas en Inglés), la calibración multiobjetivo (MO) generalmente no tiene una única solución, sino un conjunto aceptable de soluciones (Ngatchou, Zarei, & El-Sharkawi, 2005). Este conjunto de soluciones se denomina el frente de Pareto.

Aquí se presentan algunos tipos de algoritmos para la solución de problemas MO y por tanto, cómo hacen la identificación del frente. Además, se muestra un caso de estudio en la cuenca del río Combeima en Colombia, aplicando el método de calibración GLUE al modelo abcd de Thomas, para identificar el frente de Pareto con tres FO.

2. Problema inverso multiobjetivo

2.1. Frente de Pareto

El frente es un conjunto de soluciones del problema inverso que relaciona dos o más FO. Se dice que un conjunto de parámetros S pertenece al frente cuando no existe otra solución R tal que mejore un objetivo sin empeorar al menos uno de los otros definidos previamente, en lo que se conoce como un vector de parámetros no dominado.

El criterio de dominancia, permite la selección de soluciones o combinaciones de parámetros que



pertenecen al frente. Este criterio compara dos vectores de parámetros que posiblemente corresponden al frente.

El vector \vec{u} domina a \vec{v} si al menos una de las FO evaluadas $f(\vec{u})$ es mejor para \vec{u} . En consecuencia, aquellos vectores que no son dominados por ningún otro son los que pertenecen al frente de Pareto.

La optimización que ofrece el frente, reduce la incertidumbre por la selección subjetiva de las FO y brinda resultados óptimos desde un punto de vista “global” (Ngatchou et al., 2005).

2.2. Algoritmos de calibración

Muchas metodologías se han propuesto para calibrar modelos de forma SO, y algunas han sido adaptadas para la calibración MO. Así, el problema inverso MO, se puede clasificar en métodos clásicos y técnicas inteligentes diseñadas para encontrar el conjunto de posibles y aceptables soluciones conocidas como el frente de Pareto (Ngatchou et al., 2005).

Los métodos clásicos buscan convertir el problema MO en una sola función objetivo a través de diversos procedimientos. El más común de ellos es la ponderación de las funciones, que consiste en crear una nueva función con pesos que son aplicados a cada una de las funciones del vector objetivo (Ngatchou et al., 2005). Estos métodos de evaluación pueden ser usados con técnicas de calibración SO como el método SIMPLEX, búsqueda por incrementos, métodos de gradiente, etc. así como por métodos estocásticos como GLUE (Beven & Binley, 1992) y métodos de metaheurísticas basados en poblaciones. Sin embargo, la transformación a una forma SO tiene dos problemas: están orientados a encontrar un único conjunto que solucione el problema inverso; y mediante la variación de la transformación, solamente pueden identificar frentes localmente dominantes pero no globalmente. Cabe resaltar que estos métodos no tienen en cuenta el concepto de dominancia para la identificación del frente.

Por otra parte, las técnicas inteligentes se enfocan en optimizar simultáneamente cada una de las funciones individuales. Para hacer la calibración simultánea, estos métodos se basan principalmente en poblaciones, debido a que pueden evaluar varios posibles conjuntos de parámetros que tienen posibilidad de pertenecer al frente de Pareto (Ngatchou et al., 2005). La identificación del frente se hace mediante la evaluación de las FO en cada uno de los individuos pertenecientes a la población y al uso del criterio de dominancia que permite la selección de los mejores individuos. Sin embargo, estos algoritmos son complejos de programar y requieren datos adicionales como la cantidad de individuos, así como parámetros asociados al apareamiento entre ellos y el número de generaciones.

Evitando la complejidad de un algoritmo genético pero manteniendo el mismo esquema de poblaciones, una forma sencilla que permite encontrar el frente es

crear una población que pueda ser analizada para identificar los individuos que no son dominados por ningún otro, y que en consecuencia son parte del frente de Pareto. Esto se hace con un muestreo del espacio parametral mediante búsqueda por incrementos o a través de simulaciones de Monte Carlo (GLUE). Este procedimiento puede ser costoso computacionalmente, debido a la cantidad de individuos que requiere la población.

3. Caso de estudio

El presente ejercicio se realizó sobre la cuenca alta del río Combeima (Tolima, Colombia) con un área aproximada de 132 km². Se utilizaron los registros de estaciones meteorológicas e hidrológicas del IDEAM (Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia) para: 4 estaciones de lluvia, registros de temperatura media, mínima y máxima, y registros de caudal afluente en la estación San Vicente del Combeima (LM-21217220) definida como cierre de la cuenca.

4. Metodología

Primero se realiza un pretratamiento de los datos meteorológicos. Para representar la hidrología de la zona de estudio, se escogió el modelo abcd, que es un modelo hidrológico de tipo agregado y que corre en formato mensual. Descrito por Thomas en 1981, permite determinar mediante 4 parámetros el caudal promedio mensual afluente de una cuenca. Este realiza un balance de masas en el suelo representado por dos zonas (Al-Lafta, Al-Tawash, & Al-Baldawi, 2013), la primera donde se dan los almacenamientos superficiales y la zona saturada donde hay un almacenamiento permanente. Al estar basado únicamente en balance de masa y tener una parametrización enfocada en la disponibilidad del agua para evaporación y recarga, se acepta que el modelo pueda ser aplicado a cualquier tipo de cuenca alta, media o baja.

Basados en experimentos piloto, se utiliza el muestreo aleatorio de los parámetros a, b, c y d del modelo en intervalos cercanos a los puntos que generan los menores valores de una función objetivo (RMSE). Posteriormente se define un periodo de calentamiento del modelo de 36 meses para asegurar que el valor de las condiciones iniciales no tenga influencia sobre la evaluación de las FO definidas (se definen para la evaluación las funciones RMSE, MAE y el 1-NSE (se hace la inversión del NSE para garantizar un problema de minimización), (Dawson et al., 2007). Con lo anterior, se procede a realizar 10 millones de simulaciones de Monte Carlo (GLUE) para generar la población, y así, evitar algoritmos menos eficientes o más complejos.

Por último se aplica el criterio de dominancia a los individuos de la población muestreada, para identificar los puntos que corresponden al frente de Pareto. Este es un proceso iterativo que puede optimizarse descartando aquellos puntos que son dominados por un punto en particular.

5. Resultados

Puede verse en la Fig. 1 que efectivamente no existe un conjunto único de parámetros que minimice las FO evaluadas en dicha figura. Sin embargo, cuando se evaluaron el RMSE y 1-NSE (Fig. 2) se observó que un mismo conjunto de parámetros minimiza ambas funciones. Esto puede deberse a que ambas dependen en gran medida de la sumatoria del cuadrado de los errores. Es claro además que la función MAE tiene un comportamiento diferente de las otras dos, lo cual genera que no exista un único conjunto óptimo de parámetros.

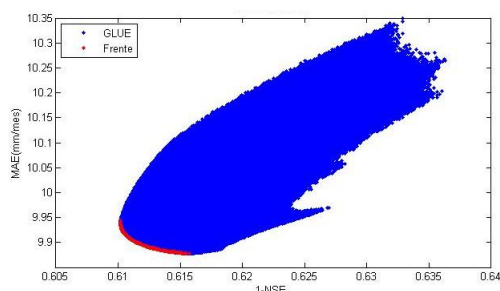


Figura 1. MAE vs 1-NSE de las simulaciones con GLUE y en rojo se resaltan las simulaciones que pertenecen al frente de Pareto.

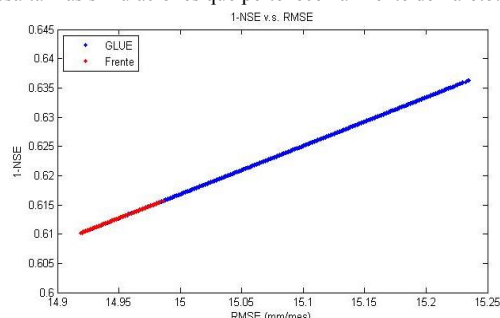


Figura 2. 1-NSE vs RMSE de las simulaciones con GLUE y en rojo se resaltan las simulaciones que pertenecen al frente de Pareto

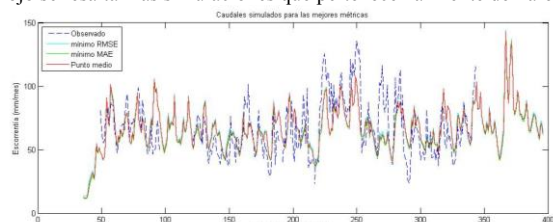


Figura 3. Hidrograma de escorrentía mensual con valores observados, con menor RMSE y MAE, y con un punto del frente.

La Tabla 1 muestra el valor mínimo y máximo para cada uno de los parámetros del modelo de Thomas, donde se puede ver que existe una gran rango para varios de los parámetros que pertenecen al conjunto de soluciones que son parte del frente de Pareto.

Tabla 1. Valores mínimos y máximos de los parámetros.

	mín.	máx.
a	0.9404	0.95
b	610	729.57
c	0.6114	0.6577
d	0.078	0.0999

La hidrógrafa de escorrentía mensual (figura 3) evidencia pocas diferencias en cuanto a picos y tiempo de ocurrencia de los mismos cuando se escogen diferentes conjuntos de parámetros dentro del frente.

6. Conclusiones y comentarios

El proceso de identificación de los individuos que pertenecen al frente está basado en la comparación de las FO, y el tiempo que esto requiere es proporcional al tamaño de la población muestreada. Por tanto, puede optimizarse el proceso mediante el descarte de aquellos conjuntos de parámetros que son dominados por otro conjunto en particular, que esté siendo comparado con todos los demás.

La poca influencia sobre la escorrentía que se presenta con la variación de los parámetros dentro del frente, puede deberse principalmente a la estructura del modelo. Posiblemente otros modelos sean realmente sensibles a los cambios de los parámetros, incluso dentro de los conjuntos que conforman el frente de Pareto.

Otra temática de gran importancia, es la comparación de diversas variables hidrológicas simuladas y observadas (caudal, humedad del suelo, etc.) que sirven para estimar los parámetros del modelo en la calibración MO.

Como trabajo futuro se plantea, el análisis de sensibilidad y de incertidumbre, así como realizar la identificación del frente a partir de algoritmos genéticos.

Agradecimientos

Al profesor Sergio Salazar por sus comentarios y apoyo, al profesor Nelson Obregón por su guía, y a los compañeros de la IAHR-YPN Bogotá y del GIREH.

Referencias

- Al-Lafta, H. S., Al-Tawash, B. S., & Al-Baldawi, B. A. (2013). Applying the “abcd” Monthly Water Balance Model for Some Regions in the United States. *Advances in Physics Theories and Applications*, 25(Figure 1), 36–48.
- Beven, K., & Binley, A. (1992). The future of distributed models: model calibration and uncertainty prediction. *Hydrological Processes*, 6(May 1991), 279–298.
- Dawson, C. W., Abrahart, R. J., & See, L. M. (2007). HydroTest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environmental Modelling and Software*, 22(7), 1034–1052.
- Gupta, H. V., Sorooshian, S., & Yapo, P. O. (1998). Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, 34(4), 751–763.
- Ngatchou, P., Zarei, a., & El-Sharkawi, a. (2005). Pareto Multi Objective Optimization. *Proceedings of the 13th International Conference On Intelligent Systems Application to Power Systems*, 84–91.
- Silberstein, R. P. (2006). Hydrological models are so good, do we still need data? *Environmental Modelling and Software*, 21(9), 1340–1352.