

Calibración de Modelos Hidrológicos

Juan Cabrera, Civ. Eng.
Universidad Nacional de Ingeniería
Facultad de Ingeniería Civil

1. Introducción

El uso de modelos hidrológicos tiene por finalidad simular los fenómenos que ocurren en la realidad. Sin embargo, a pesar de los recientes avances en ciencia y tecnología, no existe un modelo hidrológico que simule a la perfección los complejos procesos que involucra la hidrología debido a la presencia de errores de diferente origen. Estos errores pueden ser minimizados mediante un adecuado proceso de calibración.

El presente artículo se enfoca en el proceso de calibración de modelos determinísticos, presentando conceptos tan importantes como incertidumbre, función objetivo y diferentes medidas de bondad de ajuste.

2. Modelamiento determinístico

Como se mencionó en el capítulo anterior, el modelamiento determinístico no considera términos aleatorios dentro del modelo; es decir, cada simulación que se realice producirá el mismo resultado. Bajo esta idea general, el proceso de calibración se centrará en minimizar las diferencias entre los datos observados y los datos simulados. Este “criterio de exactitud” se basa en el ajuste de parámetros como medio de calibración.

El uso de modelos determinísticos requiere el manejo de términos como: variables de ingreso, parámetros y variables de estado.

a. Variables de ingreso

Son series de datos a las cuales se les aplicará una serie de ecuaciones para intentar representar otra variable. Se asume que estas variables de ingreso son independientes, es decir, no dependen de ninguna otra. Algunos ejemplos son; precipitación, temperatura, etc.

b. Parámetros

Son parámetros del modelo que han sido asumidos constantes, como la conductividad hidráulica del suelo, la rugosidad, etc.

c. Variables de estado

Representan una “situación” que afectará los resultados del modelo; por ejemplo, el modelado de una cuenca bajo régimen húmedo tiene diferentes características que en régimen seco. Un ejemplo de variable de estado es la humedad de suelo, la cual varía estacionalmente afectando la concentración de escorrentía superficial.

3. Exactitud y precisión

Dos conceptos importantes a tener en cuenta antes de utilizar modelos hidrológicos son “exactitud” y “precisión”. La exactitud mide la performance de un modelo; es decir, la cercanía entre los valores simulados y los observados. Un modelo exacto no arroja los mismos resultados, pero estos se encuentran siempre dentro de un rango alrededor de la medida observada. La Figura 1 (Izquierda) ilustra este concepto.

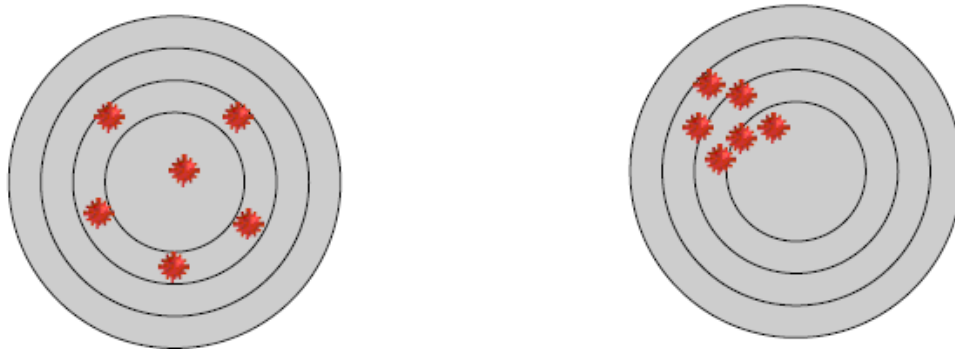


Figura 1 : Exactitud y precisión en mediciones. Izquierda: grupo de medidas consideradas exactas. Derecha: grupo de medidas consideradas precisas.

Fuente: Molnar, 2011.

La precisión mide la repetibilidad del experimento; es decir, la similitud entre valores simulados. Un modelo preciso arroja similares resultados, pero estos no son necesariamente cercanos a la medida observada. La Figura 1 (Derecha) ilustra este concepto.

Estos dos conceptos están relacionados con otros dos: errores aleatorios y errores sistemáticos. Los errores aleatorios son errores propios de fluctuaciones aleatorias de la magnitud medida o del proceso de medición y no son controlables. Debido a que son aleatorios, la medición realizada se encontrará alrededor de la medida real, muy característico de las medidas exactas. Por otro lado, los errores sistemáticos son errores que se repiten permanentemente y aproximadamente en la misma magnitud; esta característica hace que todas las mediciones se encuentren “desviadas” en un mismo valor de la medida real (como las medidas precisas). De esta manera, si se determina el valor sistemático, es fácil corregir la medida y obtener medidas mucho más exactas.

4. Fuentes de incertidumbre

En general, existen cuatro fuentes principales de incertidumbre:

a. Errores (sistemáticos o aleatorios) en los datos de ingreso

Son errores propios de los datos de ingreso, los cuales no necesariamente son datos de campo. Por ejemplo, si se ingresa la precipitación media de una cuenca y esta serie se obtiene utilizando polígonos de Thiessen, la aplicación de este método produce errores.

b. Errores (sistemáticos o aleatorios) en los datos observados

Son errores propios de la medición en campo de las diferentes variables hidro-meteorológicas que se consideran en el modelo.

c. Errores debido a valores no óptimos de parámetros

Referido a los errores ocasionados por la asunción de valores de los parámetros del modelo que no son la consecuencia de un proceso de optimización.

d. Errores debido a la estructura del modelo

Errores propios del modelo y que no pueden ser eliminados sin modificar la estructura del mismo. El proceso de calibración no los afecta.

Las diferencias entre el modelo y la realidad se deben a la combinación de todas estas fuentes. El proceso de calibración puede minimizar las tres primeras; sin embargo, no altera la cuarta, por este motivo es importante discriminar los efectos de cada fuente de error.

5. Calibración del modelo

La calibración es el proceso por el cual se identifican los valores de los parámetros del modelo para los cuales la serie de datos simulado se ajusta de manera óptima a la serie de datos observados. Para evaluar la bondad de ajuste del modelo se utiliza una “función objetivo”.

Este proceso puede realizarse de dos formas:

a. Prueba y error

Es el método más utilizado y usualmente recomendado. Implica un ajuste manual de parámetros basado en el criterio del investigador.

Tiene valor de aprendizaje pero es lento y subjetivo.

b. Automática

Optimiza los valores de los parámetros utilizando técnicas numéricas. A diferencia de la “prueba y error”, este método es mucho más rápido y objetivo; sin embargo, es numéricamente complejo y presenta problemas en la compensación de errores y en la captura de valores mínimos locales. Algunos ejemplos son: técnicas de escalamiento (hill-climbing techniques), recocido simulado (simulated annealing), algoritmos genéticos, etc.

En la práctica, se sugiere aplicar el primer método para conseguir una primera aproximación y luego optimizarla utilizando una calibración automática.

6. Función objetivo

Para iniciar el proceso de calibración es necesario definir previamente la “función objetivo”, la cual permitirá estimar la incertidumbre de las series de caudales simulados respecto a los observados.

La función objetivo es usualmente una medida de bondad de ajuste (error de balance de masas, coeficiente de correlación, eficiencia de Nash-Sutcliffe, etc), pero también puede ser una combinación de estas (dependiendo de la importancia que el investigador asigne a las diferentes medidas de bondad de ajuste) recibiendo entonces el nombre de función “multiobjetivo”.

Usualmente, una función multiobjetivo (G) asume las siguientes formas:

$$G = \sum_{i=1}^n w_i F_i \quad G = \sum_{i=1}^n w_i |F_i| \quad G = \sum_{i=1}^n w_i F_i^2$$

donde F_i son las medidas de bondad de ajuste consideradas dentro de la función multiobjetivo, mientras que w_i son los pesos asignados a cada una de dichas medidas. Usualmente, no se hace distinción y se hace uso del término “función objetivo” para referirse a una función objetivo simple o a una multiobjetivo.

Dentro del proceso de calibración, es necesario “optimizar” el ajuste entre el modelo y los datos observados. La optimización se hace “minimizando” la función objetivo; es decir:

$$\frac{\partial G}{\partial X} = 0$$

donde X es el vector de parámetros θ_i (parámetros de los que depende la función objetivo).

Asimismo, la función objetivo permite realizar análisis de sensibilidad para evaluar la perturbación provocada por la variación de un parámetro determinado; es decir:

$$S_i = \frac{\partial G}{\partial \theta_i}$$

donde S_i es la sensibilidad y θ_i es el parámetro en evaluación.

Un ejemplo de uso de la función objetivo es mostrado en la Figura 2. La serie de datos simulada no ajusta perfectamente con la serie observada; para mejorar el ajuste se propone utilizar el criterio de Nash (escalar o logaritmico) y el error de volumen. El efecto que tiene el criterio de Nash sobre la simulación es de “desplazar verticalmente” hacia arriba o hacia abajo toda la serie, mientras que el error de volumen permite desplazar horizontalmente la serie. Para el ejemplo mostrado, la combinación de estos dos criterios dentro de la función objetivo permitirá controlar mejor el proceso de calibración del modelo.

El uso de las funciones objetivo brinda un criterio analítico y objetivo para la definir la calidad de ajuste del modelo; sin embargo, este análisis debe ser complementado con un análisis visual.

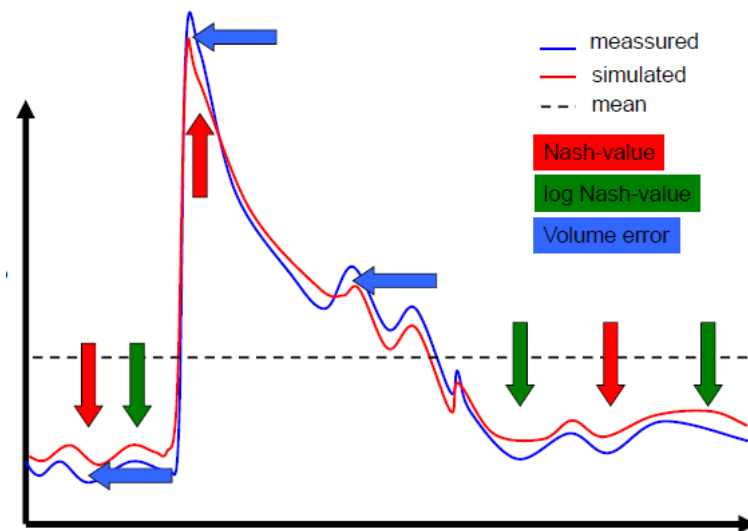


Figura 2 : Efectos de las medidas de ajuste sobre la calibración de un modelo.
Fuente: Finger, 2011.

7. Medidas de bondad de ajuste

La calibración de modelos usualmente se enfoca en un “criterio de exactitud”, el cual se apoya en la cuantificación de la bondad de ajuste del modelo. Para este fin, se hace uso de diferentes “medidas de bondad de ajuste”, entre las cuales podemos mencionar:

a. Coeficiente de calibración (r)

Expresa la dependencia lineal entre dos variables que, en nuestro caso, son los caudales observados y los caudales simulados. Se formula como:

$$r = \frac{S_{obs, sim}}{\sqrt{S_{obs} S_{sim}}}$$

donde $S_{obs, sim}$ es la covarianza sin sesgo entre los caudales observados y simulados:

$$s_{obs, sim} = \frac{1}{n-1} \sum_{i=1}^n (Q_i - \bar{Q})(Q_{sim,i} - \bar{Q}_{sim})$$

y, s_{obs} y s_{sim} son las varianzas sin sesgo de los caudales observados y simulados respectivamente:

$$s_{obs} = \frac{1}{n-1} \sum_{i=1}^n (Q_i - \bar{Q})^2 \quad y \quad s_{sim} = \frac{1}{n-1} \sum_{i=1}^n (Q_{sim,i} - \bar{Q}_{sim})^2$$

donde los valores con barras representan valores medios y n es número de datos simulados.

El coeficiente de correlación puede tomar valores entre $0 < r < 1$: a mayor valor de r , mejor ajuste.

El uso de esta medida de bondad de ajuste se recomienda cuando se espera una relación lineal entre las variables observadas y simuladas.

b. Criterio de Schultz (D)

El criterio de Schultz representa la desviación de los caudales simulados respecto de los observados. Se calcula como:

$$D = 200 \frac{\sum_{i=1}^n |Q_{sim,i} - Q_i| Q_i}{n(Q_{max})^2}$$

donde Q_{max} es la descarga máxima observada en el periodo en estudio.

El uso de este criterio es recomendado para análisis de eventos de alta resolución temporal.

Tabla 1. Valores referenciales del Criterio de Schultz.

D	Ajuste
0 - 3	Muy bueno
3 - 10	Bueno
10 - 18	Suficiente
> 18	Insuficiente

Fuente: Molnar, 2011.

c. Desviación media acumulada (MAD)

Se define como:

$$MAD = \frac{\sum_{i=1}^n |Q_{sim,i} - Q_i|}{n}$$

La desviación media acumulada debe ser minimizada en el proceso de calibración para obtener un mejor ajuste de la variable simulada respecto a la observada.

d. Eficiencia de Nash-Sutcliffe (E)

El criterio de Nash-Sutcliffe es uno de los más usados en Hidrología. Se define como:

$$E = 1 - \frac{\sum_{i=1}^n (Q_{sim,i} - Q_i)^2}{\sum_{i=1}^n (Q_i - \bar{Q})^2}$$

y mide cuánto de la variabilidad de las observaciones es explicada por la simulación. Si la simulación es perfecta, $E=1$; si se intentase ajustar las observaciones con el valor promedio, entonces $E=0$. Algunos valores sugeridos para la toma de decisiones son resumidos en la siguiente tabla.

Tabla 2. Valores referenciales del Criterio de Nash-Sutcliffe

E	Ajuste
<0.2	Insuficiente
0.2 – 0.4	Satisfactorio
0.4 – 0.6	Bueno
0.6 -0.8	Muy bueno
> 0.8	Excelente

Fuente: Molnar, 2011.

Este indicador puede ser utilizado para cualquier escala de tiempo.

Cuando los valores de la variable simulada son demasiado grandes, se suele utilizar la “eficiencia logarítmica de Nash-Sutcliffe”:

$$E_{\log} = 1 - \frac{\sum_{i=1}^n (\log(Q_{sim,i}) - \log(Q_i))^2}{\sum_{i=1}^n (\log(Q_i) - \overline{\log(Q_i)})^2}$$

e. Error de balance de masas (m).

Expresa cuantitativamente la relación entre el volumen del hidrograma observado y el simulado. Se expresa como:

$$m = 100 \frac{\sum_{i=1}^n (Q_{sm} - Q_i)}{\sum_{i=1}^n Q_i}$$

donde m es un porcentaje.

Si el ajuste es perfecto, $m=0$. En el proceso de calibración, se debe buscar alcanzar el valor más bajo de m .

Esta media de bondad de ajuste es adecuada para análisis mensuales y anuales, y para casos donde se requiere reproducir el balance se aguas.

f. Raíz del error cuadrático medio (RMSE)

La raíz del error cuadrático medio permite cuantificar la magnitud de la desviación de los valores simulados respecto a los observados. Se formula como:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Q_{sm} - Q_i)^2}{n}}$$

El rango de valores que puede tomar va desde 0 hasta el infinito (positivo). El 0 corresponde a un ajuste perfecto mientras que valores más grandes indican un menor ajuste.

8. Referencias

Beven, K. (2006). "Rainfall-Runoff Modelling. The Primer". John Wiley & Sons Ltd. West Sussex, England.

Finger, D. (2011). "Calibration of Hydrological Models". Hydrological Processes and Modelling, SS 2011. Swiss Federal Research Institute WSL. Switzerland.

Molnar, P. (2011). "Calibration". Watershed Modelling, SS 2011. Institute of Environmental Engineering, Chair of Hydrology and Water Resources Management, ETH Zürich. Switzerland.

Zappa, M. (2011). "PREVAH – Calibration and Verification". Hydrological Processes and Modelling, SS 2011. ETH Zürich. Switzerland.