Uncertainty in a monthly water balance model using the generalized likelihood uncertainty estimation methodology

Diego Rivera[1,*], Yessica Rivas[2] and Alex Godoy[3]

[1]*Laboratory of Comparative Policy in Water Resources Management, University of Concepcion, CONICYT/FONDAP 15130015, Concepcion, Chile.*
[2]*Department of Water Resources, Faculty of Agricultural Engineering, University of Concepcion, Concepcion, Chile.*
[3]*Facultad de Ingeniería, Universidad del Desarroll, Concepcion, Chile.*
[*]Corresponding author. e-mail: dirivera@udec.cl

Hydrological models are simplified representations of natural processes and subject to errors. Uncertainty bounds are a commonly used way to assess the impact of an input or model architecture uncertainty in model outputs. Different sets of parameters could have equally robust goodness-of-fit indicators, which is known as Equifinality. We assessed the outputs from a lumped conceptual hydrological model to an agricultural watershed in central Chile under strong interannual variability (coefficient of variability of 25%) by using the Equifinality concept and uncertainty bounds. The simulation period ran from January 1999 to December 2006. Equifinality and uncertainty bounds from GLUE methodology (Generalized Likelihood Uncertainty Estimation) were used to identify parameter sets as potential representations of the system. The aim of this paper is to exploit the use of uncertainty bounds to differentiate behavioural parameter sets in a simple hydrological model. Then, we analyze the presence of equifinality in order to improve the identification of relevant hydrological processes. The water balance model for Chillan River exhibits, at a first stage, equifinality. However, it was possible to narrow the range for the parameters and eventually identify a set of parameters representing the behaviour of the watershed (a behavioural model) in agreement with observational and soft data (calculation of areal precipitation over the watershed using an isohyetal map). The mean width of the uncertainty bound around the predicted runoff for the simulation period decreased from 50 to 20 m^3s^{-1} after fixing the parameter controlling the areal precipitation over the watershed. This decrement is equivalent to decreasing the ratio between simulated and observed discharge from 5.2 to 2.5. Despite the criticisms against the GLUE methodology, such as the lack of statistical formality, it is identified as a useful tool assisting the modeller with the identification of critical parameters.

1. Introduction

Hydrological models are simplified representations of natural processes, which are constituted by input variables; a processing box that mimics hydrological processes through a set of equations aimed at matching the observed and simulated values by a set of parameters and output variables. However, the incompleteness of knowledge about the state or process being modelled is defined as

Keywords. Monthly water balance; uncertainty; generalized likelihood uncertainty estimation methodology.

uncertainty (Caddy and Mahon 1995). As noted by Brazier *et al.* (2000), estimating uncertainty is not just a way to look for weaknesses in the model, it is a way to improve the model. Therefore, uncertainty should be estimated by modellers and communicated to the end-users.

Different sets of parameter combinations may generate acceptable outputs when comparing the measured data against simulated data; alternatively, some sets of parameters could have equally strong goodness-of-fit indicators. This non-uniqueness in representing a system is called equifinality (Beven and Freer 2001; Beven 2006). The Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Binley 1992) is a methodology to estimate uncertainty under real non-ideal applications (high complexity and/or high uncertainties) (Beven and Freer 2001; Beven *et al.* 2007; Vrugt *et al.* 2009). The main ideas behind GLUE are (Beven 2001): (1) good or poor model outputs are a function of the whole set of parameters, not of individual parameters, and (2) models having robust goodness-of-fit indicators have higher values of likelihood, which are indicative that the model, defined by a set of parameters, is correctly representing the system. Thus, goodness-of-fit indicators are less formal or informal likelihood measures because they do not include *a priori* the statistical distribution of the input parameters.

The equifinality concept could be considered as a starting hypothesis (to be proven), and the uncertainty bounds could be considered as a proxy for goodness-of-fit (because the bands are calculated from intensive Monte Carlo model realizations) in order to identify a single parameter set or a group of parameter sets with a given model structure as potential representations of the system (Wagener and Kollat 2007; Muñoz 2011). Thus, analysing changes in the uncertainty bounds allows the assessment of different sets of parameters.

The objective of this paper is to estimate the uncertainty in a simple hydrological model and exploit the use of uncertainty bounds. Then, we analyze the presence of equifinality in order to improve the identification of relevant hydrological processes.

2. Materials and methods

2.1 *The monthly water balance model*

The lack of stream flow data in many Chilean watersheds has made necessary the use of hydrological models for planning and design of water resources infrastructure. The first hydrological model used in Chile was the Brown, Ferrer y Ayala model (thereafter BFA model; Ferrer *et al.* 1973). This model has been widely used to estimate monthly stream flows in pluvial watersheds (e.g., Muñoz *et al.* 2011), employing monthly rainfall and evaporation as inputs. The BFA model has been used in simulating, for example, summer (dry season) water availability to assess the feasibility of irrigation projects. Later, water balance models have been developed from the BFA, like the MAGIC model used by the Chilean Water Authority (Zambrano *et al.* 2005).

The BFA model (figure 1) is a lumped model that considers a watershed as a double storage system: upper soil storage (SS) (water storage in the soil and runoff generation) and groundwater storage (GS). The GS accepts the overflows from the SS, and its overflow flows into the river generating the baseflow (ES) in the watershed. The input variables are the monthly rainfall (PM) and the monthly potential evapotranspiration (ETP), both acting over SS. Actual evapotranspiration (ER) is estimated using potential evapotranspiration values and available water stored in the soil. The only output is the total runoff (EST) at the watershed's outlet as the sum of both the direct runoff (EI) and ES.

The parameters of the model are described in table 1 (Muñoz *et al.* 2014). All parameters have physical meaning, but not all of them have a well-defined physical range. For example, $H_{\max}$, $C_{\max}$, D, $P_{\lim}$ and PORC are related to physical features of the soils in the watershed, integrating in one single value both the spatial and temporal variability, but are not necessarily a mean value for the watershed. In table 1, H1 corresponds to the initial value requiring iteration during each time step to achieve the closure of the water balance.

Parameters A and B modify the available records of rainfall and evaporation (e.g., pan evaporation) to represent the watershed values (including, for

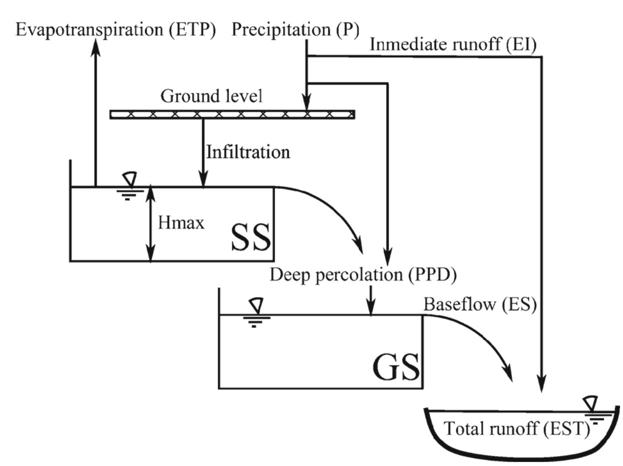

Figure 1. The BFA model: SS is the surface storage and GS is the underground storage. See text and table 1 for details.

Table 1. *Parameters of the BFA model: Values in parenthesis in the first column are the range of variation of the parameters used in the Rio Chillan model; values from Ferrer et al. (1973) and data collected by the authors.*

Parameter	Description	Variables
A (0.8–2.5)	Adjusts the station rainfall data to represent the watershed rainfall	$P_t = A \cdot PM_t$
B (0.6–1.0)	Adjusts the station evaporation data to represent the watershed evapotranspiration	$ETP_t = B \cdot EM_t$ $ER_t = ER_t(H_1, ETP_t, H_{\text{crit}})$
H_{max} (100–500 mm)	Maximum capacity of the soil to retain water	
C_{max} (0.2–0.6)	Maximum runoff generation (EI) under saturated conditions	$EI = C_{\text{max}} \cdot \frac{H_t^{(1)} + H_{t-1}}{2H_{\text{max}}} \cdot P_t$ $H_1 = H_{t-1} + P_t - EI_t - PPD_t$
D (0.1–0.6)	Percentage of rainfall transformed into deep percolation (PPD)	$PPD_t = D \cdot (P_t - P_{\text{lim}})$
P_{lim} (5–1000 mm)	Rainfall threshold over there is deep percolation	
PORC (20–60)	Fraction of H_{max} that defines the soil water content restricting the evaporation processes	$H_{\text{crit}} = PORC \cdot \frac{H_{\text{max}}}{100}$
Ck (0.2–0.9)	Constant value regulating the releases ES_t from the linear GS	$ES_t = Ck \cdot V_t$

instance, the orographic effect), ensuring the long-term water balance in the watershed. As in the Stanford Model (see Crawford and Burges 2004), A is a scaling factor adjusting the total rainfall over the watershed, RW, to the record from a single rain gauge, RS, as RW=A·RS (Gupta *et al.* 2005). If an isohyet (or evapotranspiration) map or more than a single rain gauge are available, it is possible to estimate *a priori* A (or B), following any procedure to calculate areal precipitation (Chow *et al.* 2005).

2.2 *A model for the Rio Chillan Watershed*

The Río Chillán (Chillan River) is a 757 km^2 watershed located in the Central Valley (CV, 33°–38°S). The CV is a highly productive agricultural area that depends heavily upon surface and groundwater supplies from upper Andean watersheds. In turn, water availability for multiple users depends on precipitation and temperature regimes, which are highly variable in both time and space due to ENSO events (Montecinos and Aceituno 2003) and the orographic effect of the Andes. Climate in the CV is Mediterranean – ca. 80% of precipitation occurs in winter (May–July) and, in contrast, during the summer months (September–March) precipitation is <10% of annual precipitation. Summer is also the irrigation season, so ES is essential for agriculture.

The headwaters of Chillan River are located at the piedmont in the Andes Mountains, flowing north–west through the Central Valley, reaching the Ñuble River, close to Chillán city. Part of the streamflow comes from snowmelt, but its magnitude is negligible (Toro 2009). Indeed, the hydrological behaviour can be considered as pluvial: high streamflow in the rainy season (June–July) and low flows from November to May (figure 2). The Chillan River feeds an extensive irrigation channel network from October (discharge at the abstraction 16 m^3 s^{-1}) to April (discharge at the abstraction 6 m^3 s^{-1}) (Toro 2009).

Stream flow data corresponds to official records for gauging stations Esperanza and Confluencia (figure 2, limnimetric stations managed by the Chilean Water Authority). The observed discharge time series used to calibrate the model was monthly mean values at the Confluencia gauging station plus the total discharge withdrawn by the irrigation channels (also supplied by the Chilean Water Authority). Monthly rainfall (automatic tipping bucket rain gauge) and monthly class-A pan evaporation were obtained from the Agrometeorological Station (Estación Agrometeorológica) at the Universidad de Concepción, Campus Chillán (36.57°S, 76.1°W; 144 masl). Both records span over the period January 1999–January 2006. As a result of the short length of the available time series (85 values), parameters and uncertainty bounds were obtained using the entire time series for calibration.

In the case of Río Chillán, we had access to a single meteorological station. Thus, for our first run we varied all parameters after which we contrasted GLUE-derived values (dotty plots, *a posteriori* distributions and uncertainty bands) with areal precipitation values using a hysoyetical map. It is important to note that the hysoyetical maps were derived 30 years ago and show a static picture of precipitation variability and so must be used carefully.

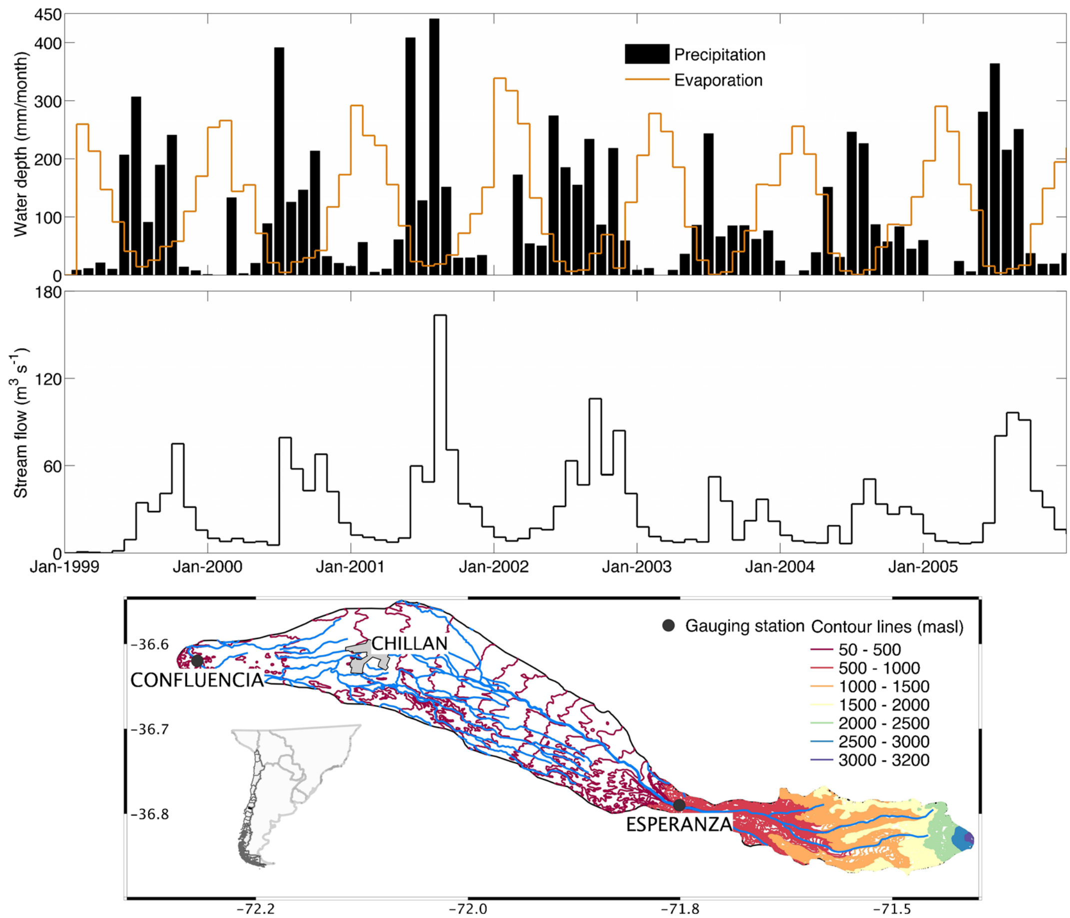

Figure 2. Upper panel: Monthly values for rainfall and pan evaporation in Chillan, and streamflow at Confluencia gauging station. Lower panel: watershed location and gauging stations, Esperanza (435 masl; drainage area 224 km^2) and Confluencia (70 masl.; drainage area 674 km^2).

2.3 *Estimating uncertainty bands*

The most common approach to estimate the output uncertainty is to analyse the error propagation (Caddy and Mahon 1995). For instance, for a fixed model structure and a given parameter set, the uncertainty in the model output, in probabilistic terms, can be computed by varying the input data, in a similar way to a sensitivity analysis. If the input dataset is considered reliable and if the model structure is fixed, it is possible to compute the uncertainty associated with a given set of parameters.

GLUE requires a number of subjective decisions, so the uncertainty bounds are, in essence, qualitative. Also, any effects of model nonlinearity, co-variation of parameter values, and errors in model structure, input data or observed variables are implicitly included (Beven and Freer 2001). The GLUE methodology has been subject to discussion regarding formal statistical issues (for details, see Mantovan and Todini 2006; Mantovan *et al.* 2007; Beven *et al.* 2007, 2008), as well as the influence on uncertainty bands of different likelihood measures and sampling procedures (Montanari 2007; Stedinger *et al.* 2008). However, formal and informal approaches to estimate uncertainty have shown similar results (Zhang *et al.* 2006; Vrugt *et al.* 2009).

The GLUE methodology uses Monte Carlo simulation to generate *a posteriori* distribution of the parameters, as well as confidence limits for the outputs (Khu and Werner 2003). In each Monte Carlo simulation, a parameter set is randomly sampled from the parameter space (the distribution of the parameter can be uniform, normal, or another model) to run the model. Each model realization can be defined as behavioural or non-behavioural. Behavioural models are those with values of one or more performance measures greater than a threshold defining realistic simulations (Beven 2001). Then, performance measures are converted to likelihood by a rescaling procedure. Finally, the prediction of each behavioural simulation (if a threshold is defined) is weighted by the likelihood for that simulation. Cumulative likelihood allows to construct *a posteriori* parameter's distributions, allowing an insight into the parameter sensitivities

(Mo *et al.* 2006) and the ability to define feasible parameter sets.

The likelihood function must be chosen *a priori*, and it is therefore subject to discussion because it affects the distribution of each model's realization (Wagener *et al.* 2001; Khu and Werner 2003; Beven 2006). A likelihood measure should increase monotonically, above zero, with increasing goodness of fit. Moreover, the sum of the likelihood values must be one (Beven 2001). We used as likelihood measures the Transformed Relative Error (TRD) and the Nash–Sutcliffe Efficiency (NSE) proposed by Nash and Sutcliffe (1970):

$$\mathrm{TRD} = 1 - \frac{\sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(Q_{\mathrm{obs}}^{i} - Q_{\mathrm{sim}}^{i}\right)^{2}}}{\overline{Q}_{\mathrm{obs}}} \tag{1}$$

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{m}(Q_{\mathrm{obs}}^{i} - Q_{\mathrm{sim}}^{i})^{2}}{\sum_{i=1}^{m}(Q_{\mathrm{obs}}^{i} - \overline{Q}_{\mathrm{obs}})^{2}} \tag{2}$$

where Q_{sim} is the simulated discharge, Q_{obs} is the observed discharge at the watershed's outlet, i is the ith month, m is the number of data points, and $\overline{Q}_{\mathrm{obs}}$ is the mean discharge for m points. The second term in the right hand side of equation (1) is the classical definition of the relative error as the mean square error and the mean value of the observations, so TRD=1 implies a good match and values ≤ 0 indicate a poor match. The NSE is an indicator that compares the mean square error generated by a particular model simulation to the variance of the target output, i.e., comparing the variability of the model's residuals against the variability of the observed series (for a review on the use of the NSE, see Schaefli and Gupta 2007).

The estimation of confidence levels and other results was conducted using the Monte Carlo Analysis Toolbox (MCAT Wagener and Kollat 2007) implemented in MATLAB®. This toolbox is a powerful tool for the quantitative and qualitative (visual) assessment of hydrological models (Wagener and Kollat 2007). In MCAT, for each point in time, a cumulative frequency distribution is generated using the likelihood values from equations (1) and (2), and the confidence intervals are calculated using linear interpolation for 5% and 95% (Wagener and Kollat 2007).

The range of the parameters in the BFA models are indicated in table 1, considering *a priori* uniform distributions for the Monte Carlo sampling. A detailed workflow for calculations of the BFA model can be found in Muñoz *et al.* (2014). In order to test the uncertainty associated with parameters we performed 10,000 model realizations considering a fixed model structure. The number of runs was defined after a trial-and-error process with a stop criterion of $R \geq 0.999$ as the correlation between the upper and lower limits of the bands of uncertainty for two different trials (e.g., comparing the uncertainty bound for 9000 realizations against 10,000 realizations) (Muñoz *et al.* 2014).

3. Results and discussion

In terms of likelihood (the closer the value to 1, the better the model), figure 3 displays the value of TRD associated with each parameter for each model realization. Most of the parameters have similar mapping: higher density of points, uniformly distributed, in a horizontal band close to the upper limit for likelihood (for the 10,000 runs). This concentration of points shows that it is possible to find several parameter sets with high likelihood (or in other words, with low relative error) within the predefined range, which is an expression of equifinality. The exception is the parameter A where the likelihood values are localized in a narrow non-horizontal band, showing a maximum value close to 1.5. Values of the parameter A from the 1.4–1.8 range strongly decrease the goodness-of-fit (lower likelihood) showing that the model is highly sensitive to this parameter, i.e., it is identifiable.

In order to identify in a better way, the range of parameter values with higher likelihood, the GLUE methodology allows us to estimate *a posteriori* the frequency distribution of the parameters (figure 4). The initial assumption about the uniform distribution of the parameters is (weakly) true for most of the parameters. Again, the exception is the A parameter, which shows a normal-like distribution, indicating that the most likely value for this parameters is close to 1.5.

Table 2 shows the best parameter sets considering TRD, as well as NSE. The highest values of likelihood (rectangles in figure 3) for $C_{\max}$ and $H_{\max}$ are 0.6 and 500, respectively, both being close to the upper limits previously defined (see table 1, indicating a high capacity to store water in the soil). Indeed, the low value of $P_{\mathrm{lim}} = 100$ mm and $D = 0.4$ suggest that in the watershed the rainfall threshold for deep percolation is very low but half of the water available for deep percolation stays in the SS. On the other hand, a low value of $\mathrm{Ck} = 0.3$ indicates that the water stored in the soil is slowly released to the river because Ck linearly regulates the release of water from the groundwater storage.

A closer inspection of table 2 leads us to say that there is equifinality in the model outputs, that is,

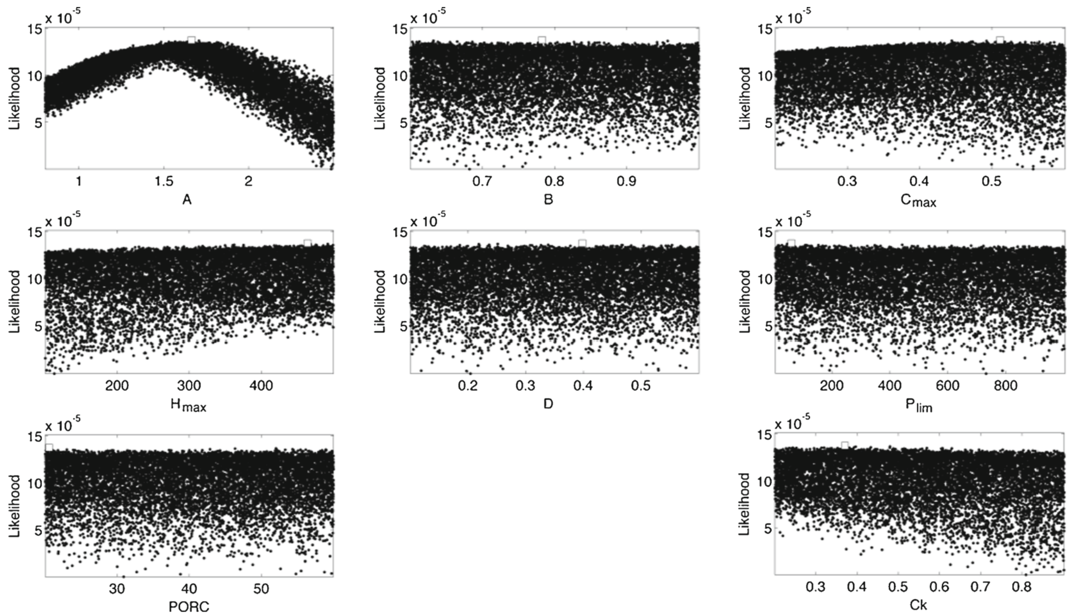

Figure 3. Dotty (scatter) plots for each parameter of the BFA model considering the TRD as likelihood measure. The white square indicates the best parameter set.

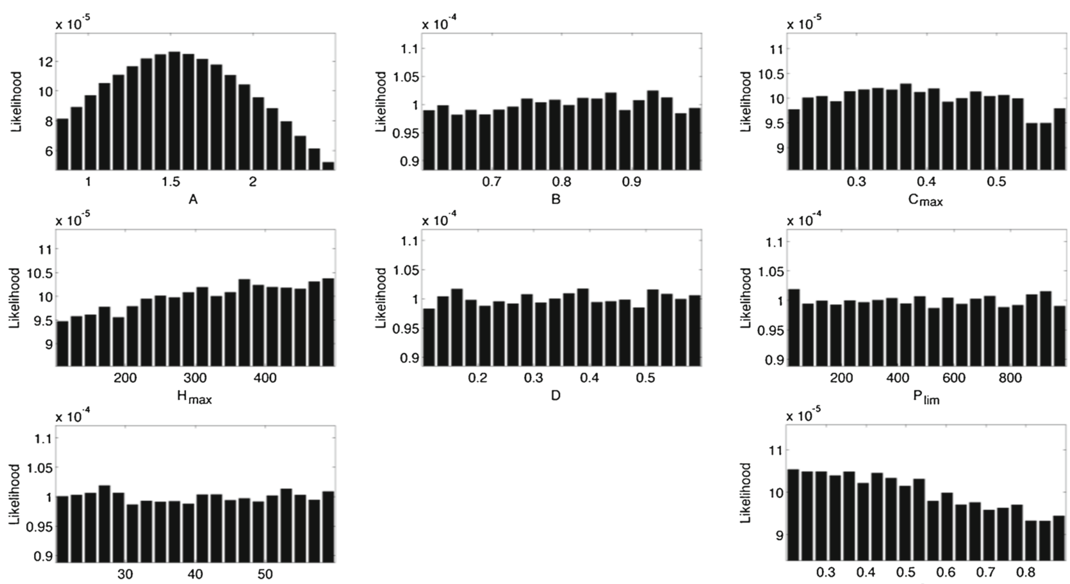

Figure 4. *A posteriori* probability distribution for each parameter of the BFA model (TRD as likelihood measure).

the listed parameter sets have very similar performance indicators but there are also clear differences between the sets. The main difference is the subset $\{P_{\lim}$, PORC, Ck$\}$. These differences lead to very similar outputs, therefore the modeller must carefully choose the parameter set that is most representative of the watershed behaviour, beyond the match between observed and simulated data. For instance, in the case study, the subsets $\{D{=}0.26$, $P_{\lim}{=}36$, PORC$=26\}$ and $\{D{=}0.1$, $P_{\lim}{=}236$, PORC$=56\}$ suggest different internal mechanisms within the watershed (e.g., groundwater storage and release of ES, evapotranspiration, ER, under restrictive conditions, and also less studied processes in Andean watershed (like fracture storage). For example, rainfall is transformed into Deep Percolation (PPD) as $PPD_t = D \cdot (P_t - P_{\lim})$, so for a monthly precipitation of 250 mm, the combination $\{D{=}0.26,\ P_{\lim}{=}36\}$ yields a deep percolation amount of 55.64, favouring groundwater storage, while the combination $\{D{=}0.1,\ P_{\lim}{=}236\}$ attains 1.4 mm of deep percolation favouring direct runoff. Low values of PORC imply that evapotranspiration has a lower limit for vegetation stress.

Table 2. *Best parameter sets of the BFA model considering the transformed relative error (TRD) and the Nash–Sutcliffe Efficiency (NSE) as objective functions.*

A	B	$C_{\max}$	$H_{\max}$	D	$P_{\lim}$	PORC	Ck	RD	NSE
TRD									
1.75	0.8	0.59	494	0.4	80	52	0.23	0.7	0.91
1.67	0.68	0.47	450	0.51	237	56	0.27	0.7	0.91
1.63	0.75	0.47	410	0.26	36	42	0.38	0.7	0.9
1.72	0.61	0.49	497	0.15	584	45	0.35	0.7	0.9
1.74	0.93	0.49	421	0.47	130	48	0.26	0.69	0.9
NSE									
1.75	0.8	0.59	494	0.4	80	52	0.23	0.7	0.91
1.89	0.61	0.48	485	0.27	44	47	0.21	0.66	0.91
1.82	0.67	0.48	477	0.29	154	42	0.22	0.69	0.91
1.74	0.67	0.49	404	0.36	55	30	0.23	0.69	0.91
1.55	0.64	0.52	479	0.59	469	54	0.39	0.68	0.91

The output's uncertainty is shown in figure 5(a) (without setting a threshold for behavioural models, as we kept all model runs). The output uncertainty band brackets the observations, being wider in the wet season, indicating that the models fail in reproducing high flows, but become more reliable in the low flow season. Thus, the model is less uncertain when there is no rainfall entering the system, suggesting that the current structure of the model is not the most optimal for accounting stream flows in the wet season, or for 'fast' response. It is also worth noting that the parameters of the model are constant for the whole watershed and the complete simulation period, making it difficult to capture the spatial and temporal changes in the watershed. One approach to overcome this shortcoming is to change the model structure to allow changes in the parameter values depending on the rainfall.

The parameter A strongly affects the model's performance because the only water input into the watershed considered by the BFA model is the rainfall. Indeed in the strictest sense, A is not a parameter but an input value dependent upon the characteristics of the watershed. Based on the available isohyetal map for the area (Toro 2009), we calculated the areal precipitation as described in Chow *et al.* (2005) and then calculated the ratio between areal precipitation and the precipitation for the weather station. This calculation yields a value of 1.6, i.e., the mean precipitation in the watershed is 1.6 times the monthly precipitation at Chillán. The likelihood was greatest around the same value (figure 3), so the parameter A was fixed to 1.6. In order to show the iterative processes of model selection, we carried out 10,000 Monte Carlo simulations considering only seven model parameters.

As expected, after fixing the parameter A, there is a significant reduction in the uncertainty (figure 5b). It is worth noting that after fixing A, the dotty plots, as well as the probability distribution function for each parameter changed slightly. As a result we show the uncertainty bounds as an integration of those changes. Figure 7 compares the *a posteriori* distribution of each parameter. The first run – all parameters varying – made no assumption regarding parameters A and B, resembling first stages of hydrological modelling. The parameter A was then fixed, and in a third run, we also fixed parameter B. As we used pan evaporation, we fixed the value to $B = 0.85$. This weak identifiability evidences that the model will adjust the rest of parameters to attain a long-term water balance. The rest of the parameters do not show identifiable ranges (no changes in the *a posteriori* distribution) due to overparametrization of the model and parameters' interdependency (Li *et al.* 2009; Muñoz *et al.* 2014). After fixing parameters A and B there is no significant reduction in the width of the uncertainty band (results not shown).

Thus, by sequentially fixing parameters that show identifiable ranges (non-uniform *a posteriori* probability distributions), dotty plots and uncertainty bands are valuable tools for this task. Indeed, Muñoz *et al.* (2014) developed within this concept a simple method aimed at constraining the equifinal parameters and reducing the uncertainty bands of model outputs. The model discards equifinal solutions by inspecting the identifiability plots (gradient of the cumulative distribution derived from the *a posteriori* distribution), and narrowing the range of each parameter used in Monte Carlo realizations by inspecting dotty plots.

In order to compare how much uncertainty decreases after fixing parameter A, we calculated the width of the uncertainty bounds, ΔCF, as the difference between upper (95%) and lower (5%) confidence limits for each simulation step. As seen in figure 6(a), in both cases, ΔCF increases during the precipitation season and decreases during the

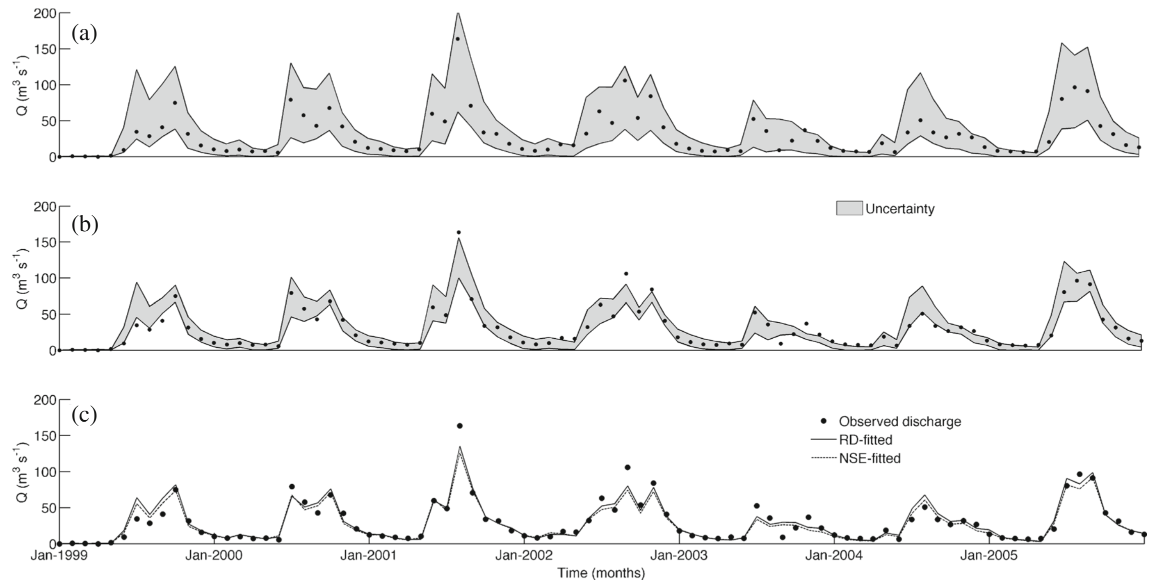

Figure 5. Uncertainty band for monthly streamflow simulated using the BFA model for (**a**). All parameters subject to Monte Carlo sampling (**b**) only parameter A=1.60. (**c**) Comparison between the best realization considering the TRD and NSE (A fixed). The confidence limits are 95% and 5% for (a) and (b).

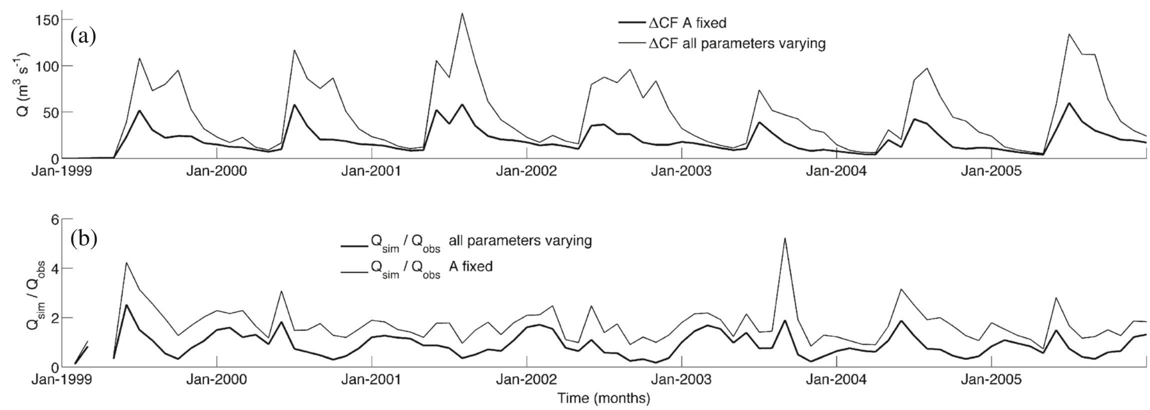

Figure 6. Width of the uncertainty band, ΔCF, for monthly streamflow simulated using the BFA model considering all parameters varying and parameter A fixed. (b) Ratio between simulated and observed values considering all parameters varying and parameter A fixed.

baseflow periods but the magnitude of ΔCF is lower after fixing parameter A. The mean of ΔCF for all varying parameters is 50 m^3 s^{-1}, while for parameter A fixed to 1.6, the value is 20 m^{3}s^{-1}. Figure 7(b) displays the ratio between simulated (averaged over 10,000 model realizations) and observed discharges for both approaches: after fixing the parameter A, the maximum value of the ratio decreases from 5.2 to 2.5. Even though there is a reduction of the width of the uncertainty band, a side effect appears. As the band has been narrowed, the reliability – the percentage of observations bracketed by the uncertainty band – decreases. This result implies that there must be a trade-off between reliability and uncertainty bands. For the BFA model, this decrease in uncertainty means that peak flows are not bracketed, suggesting a structural failure of the model to capture this process.

Comparing the best outputs with consideration to TRD and NSE as the objective functions (figure 5c), the results are similar. However that may not always be true, depending upon the watershed to be modelled. A closer inspection of figure 8 shows that the best 10 parameter sets (A fixed) perform in a similar way, however between January 2003 and June 2005, there are two clear groups. The first group, Group F (B≈0.8; P_{lim}≈40 mm; Ck≈0.3), is a better fit of the observed stream flow under predominant baseflow conditions (streamflow after two consecutive years with rainfall below the average). Low values for P_{lim} imply an increase

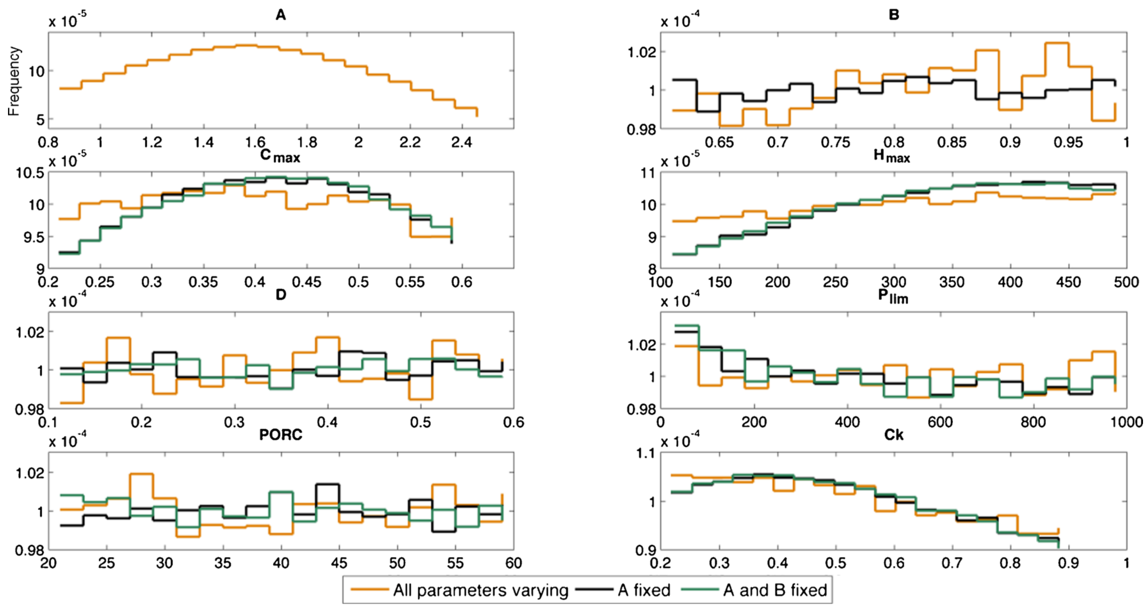

Figure 7. Changes on the *a posteriori* distribution of parameters for three sequential cases: all parameters varying, A-fixed, and both A and B fixed.

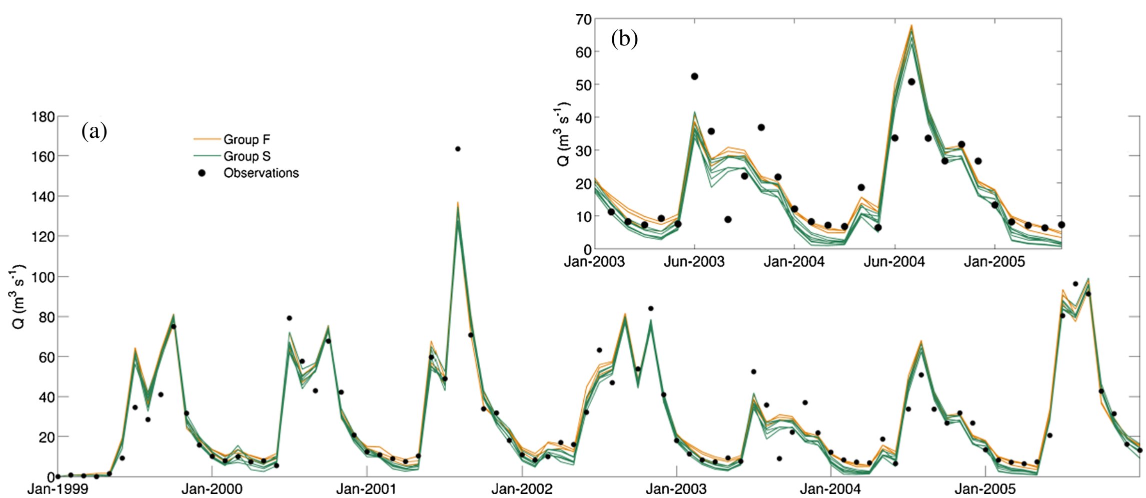

Figure 8. (**a**) Comparison of observed and simulated values using the best 10 parameter sets; (**b**) same as (a), but zoomed from January 2003 to June 2005.

of groundwater recharge as precipitation volumes greater than P_{lim} mainly travel to the groundwater storage. On the other hand, low values for Ck imply a more stable release from the groundwater storage. Thus, the combination of low values for both P_{lim} and Ck lead to a higher and more stable base flow during the dry season. On the contrary, Group S ($B\approx0.7$; $P_{\mathrm{lim}}\approx90$ mm; Ck$\approx$0.4) is related to less percolation, less evapotranspiration and quicker release of water from the groundwater storage, leading to a less persistent baseflow, as well as less discharge. It is worth noting that Group S produces similarly good results, especially for some peaks flows. This is indicative of the model's inability to reproduce all variations showing the streamflow data with a single set of parameters. Values for Group S are in line with the high storage capacity observed in the basins of volcanic origin located in central Chile that drain from the Andes (Muñoz *et al.* 2014), so we define this parameter group as the behavioural set.

The BFA model has been used in a wide set of watersheds, with its structure unchanged. As 'it is unlikely that a single model structure provides the best stream flow simulation from multiple basins in different climate regions' (Clark *et al.* 2008), it must be stressed that the results presented here are valid for a given structure and parameterization of the model.

In this case study, both performance measures, TRD and NSE, gave similar results but it is worth noting that using different indices could lead to different results (linked to parameter sets). However, increasing the number of performance measures does not ensure a better model evaluation (Viola *et al.* 2009; Xiong *et al.* 2009). As discussed elsewhere, goodness-of-fit indices are intended to assess different parts of the hydrographs (base flow, peaks, recession). Even though a multi-objective approach is appealing, we use a 2-objective approach to keep the case study simple using widely used and easy-to-implement measures, namely Nash–Sutcliffe Efficiency and Relative Differences.

Model-related uncertainties are not the only source of uncertainties (Shirmohammadi *et al.* 2006). Input parameters and measurement (or the lack of field measurement) are also sources of uncertainties, as these make the process of model evaluation difficult. GLUE aggregates different sources of errors (input data, structure, relationships among parameters), it is therefore difficult to draw conclusions about unacknowledged errors or uncertainties. As a consequence of the above-mentioned phenomena, "there is a large difference between a better fit (to observed data) for a given model and a better model" (Ebel and Loague 2006, p. 2889).

4. Conclusions

The application of the GLUE methodology demonstrates that the Monthly Water Balance model used to represent the monthly discharge at the outlet of the Rio Chillan watershed exhibits, at a first stage, equifinality, i.e., different sets of parameters generate acceptable outputs. The outputs are highly sensitive to the parameters A, $H_{\max}$, $P_{\lim}$, and Ck, representing the water input, the soil and the groundwater systems. However, it was possible to narrow the range for the parameters and eventually identify a set of parameters representing the behaviour, i.e., a behavioural model, of the watershed in agreement with observational and soft data.

In the case study, the BFA model presents a wider band of uncertainty for the rainfall season, showing some problems in representing the hydrological processes leading to peak stream flows. The uncertainty shows a significant reduction after the parameter A is fixed or with a narrower range of distribution, considering prior knowledge. We propose that A values should not be considered as a parameter (and is therefore not subject to equifinality), because it could be estimated *a priori* or contrasted against rainfall data.

To summarize, we applied GLUE knowing that it is a good tool for assessing a model in the early stages of model development instead of a conclusive tool. The GLUE methodology is a useful tool, assisting the modeller with the identification of critical parameters, as well as with the proper communication of model uncertainties. The last decision, however, always lies with the modellers, who are best placed to identify the best models and parameters based on their judgment and knowledge.

Our approach is simple, easy to implement, and has the potential to be used in the practice of hydrology.

Acknowledgements

The BFA model was coded in MATLAB by Dr© FulvioRivano and Leonardo Vega. Codes and datasets are available on request. This research was benefited from FONDECYT Grant Nos. 11090032 and 1110298. We thank Dr Claire Baffaut and Dr José Luis Arumí for their comments and suggestions. Thanks to Roto Quezada.

References

Beven K 2001 Rainfall-runoff modelling: The primer; John Wiley & Sons Inc, Sussex, England.

Beven K 2006 A manifesto for the equifinality thesis; *J. Hydrol.* **320(1–2)** 18–36.

Beven K and Binley A 1992 The future of distributed models: Model calibration and uncertainty prediction; *Hydrol. Process.* **6(3)** 279–298.

Beven K and Freer J 2001 Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology; *J. Hydrol.* **249(1–4)** 11–29.

Beven K, Smith P and Freer J 2007 Comment on "Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology" by Pietro Mantovan and Ezio Todini; *J. Hydrol.* **338(3–4)** 315–318.

Beven K, Smith P and Freer J 2008 So just why would a modeller choose to be incoherent; *J. Hydrol.* **354(1–4)** 15–32.

Brazier R, Beven K, Freer J and Rowan J 2000 Equifinality and uncertainty in physically based soil erosion models: Application of the GLUE methodology to WEPP – the Water Erosion Prediction Project – for sites in UK and USA; *Earth Surf. Process. Landforms* **25(8)** 825–845.

Caddy J and Mahon R 1995 Reference points for fisheries management; In: FAO Fisheries Technical Paper, No. 347, Food and Agriculture Organization of the United Nations, Rome.

Chow V, Maidment Dm and Mays L 2005 *Applied hydrology*; McGraw-Hill, New York, USA.

Clark M, Slater A, Rupp D, Woods R, Vrugt J, Gupta H, Wagener T and Hay L 2008 Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models; *Water Resour. Res.* **44** W00B02.

Crawford N and Burges S 2004 History of the Stanford watershed model; *Water Resources Impact* **6(2)** 3–5.

Ebel B and Loague K 2006 Physics-based hydrologic-response simulation: Seeing through the fog of equifinality; *Hydrol. Process.* **20(13)** 2887–2900.

Ferrer P, Brown E and Ayala L 1973 Simulación de gastos medios mensuales en una cuenca pluvial; In: II Coloquio Nacional de Ingeniería Hidráulica.

Gupta H, Beven K and Wagener T 2005 Model calibration and uncertainty estimation; In: *Encyclopedia of Hydrologic Sciences* (ed.) Anderson M E A, John Wiley & Sons, pp. 1–17.

Khu S and Werner M 2003 Reduction of Monte-Carlo simulation runs for uncertainty estimation in hydrological modelling; *Hydrol. Earth Syst. Sci.* **7(5)** 680–692.

Li L, Xia J, Xu C Y, Chu J and Wang R 2009 Analyse the sources of equifinality in hydrological model using GLUE methodology, In: *Hydroinformatics in Hydrology, Hydrogeology and Water Resources*, Proceedings of Symposium JS, **4** 130–138.

Mantovan P and Todini E 2006 Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology; *J. Hydrol.* **330(1–2)** 368–381.

Mantovan P, Todini E and Martina M L 2007 Reply to comment by Keith Beven, Paul Smith and Jim Freer on hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology; *J. Hydrol.* **338(3–4)** 319–324.

Mo X, Pappenberger F, Beven K, Liu S, De Roo A and Lin Z 2006 Parameter conditioning and prediction uncertainties of the LISFLOOD-WB distributed hydrological model; *Hydrol. Sci. J.* **51(1)** 45–65.

Montanari A 2007 What do we mean by "uncertainty"? The need for a consistent wording about uncertainty assessment in hydrology; *Hydrol. Process.* **21(6)** 841–845.

Montecinos A and Aceituno P 2003 Seasonality of the ENSO-related rainfall variability in central Chile and associated circulation anomalies; *J. Climate* **16** 281–296.

Muñoz E 2011 Perfeccionamiento de un modelo hidrológico: aplicación de análisis de identificabilidad dinámico y uso de datos grillados; PhD Thesis, University of Concepcion, Chile.

Muñoz E, Álvarez C, Billib M, Arumí J L and Rivera D 2011 Comparison of gridded and measured rainfall data for basin-scale hydrological studies; *Chilean J. Agricultural Res.* **71(3)** 459–468.

Muñoz E, Rivera D, Vergara F, Tume P and Arumí J L 2014 Identifiability analysis: Towards constrained equifinality and reduced uncertainty in a conceptual model; *Hydrol. Sci. J.*, doi: 10.1080/02626667.2014.892205.

Nash J and Sutcliffe J 1970 River flow forecasting through conceptual models. Part I: A discussion of principles; *J. Hydrol.* **10(3)** 282–290.

Schaefli B and Gupta H 2007 Do Nash values have value? *Hydrol. Process.* **21(15)**.

Shirmohammadi A, Chaubey I, Harmel R, Bosch D, Munoz-Carpena R, Dharmasri C, Sexton A, Arabi M, Wolfe M and Frankenberger J *et al.* 2006 Uncertainty in TMDL Models; *Trans. ASABE* **49** 1033–1049.

Stedinger J, Vogel R, Lee S and Batchelder R 2008 Appraisal of the generalized likelihood uncertainty estimation (GLUE) method; *Water Resour. Res.* **44** W00B06.

Toro P 2009 Análisis del manejo de los recursos hídricos del río chillán; Master's thesis, Department of Civil Engineering, University of Concepcion.

Viola F, Noto L, Cannarozzo M and La Loggia G 2009 Daily streamflow prediction with uncertainty in ephemeral catchments using the GLUE methodology; *Phys. Chem. Earth* **34(10–12)** 701–706.

Vrugt J, terBraak C, Gupta H and Robinson B 2009 Equifinality of formal (DREAM) and informal (GLUE) bayesian approaches in hydrologic modeling; *Stoch. Env. Res. Risk A* **23(7)** 1011–1026.

Wagener T and Kollat J 2007 Numerical and visual evaluation of hydrological and environmental models using the Monte Carlo analysis toolbox; *Environ. Modell. Softw.* **22(7)** 1021–1033.

Wagener T, Lees M and Wheater H 2001 A toolkit for the development and application of parsimonious hydrological models; In: *Mathematical models of small watershed hydrology* (eds) Singh Frevert and Meyer, **2** Water Resources Publications, LLC, USA.

Xiong L, Wan M, Wei X and O'Connor K 2009 Indices for assessing the prediction bounds of hydrological models and application by generalised likelihood uncertainty estimation; *Hydrol. Sci. J.* **54(5)** 852–871.

Zambrano M, Pagliero L, Rojas R and Gangas A 2005 Magic: Modelación integrada de cuencas y acuíferos; Serie Informes Técnicos 107, Dirección General de Aguas, Departamento de Estudios y Planificación.

Zhang D, Beven K and Mermoud A 2006 A comparison of non-linear least square and GLUE for model calibration and uncertainty estimation for pesticide transport in soils; *Adv. Water Resour.* **29(12)** 1924–1933.

MS received 27 February 2014; revised 9 September 2014; accepted 11 September 2014