Assessment of the suitability of rainfall–runoff models by coupling performance statistics and sensitivity analysis

Mun-Ju Shin and Chung-Soo Kim

ABSTRACT

Conceptual rainfall–runoff models are widely used to understand the hydrologic responses of catchments of interest. Modellers calculate the model performance statistics for the calibration and validation periods to investigate whether these models serve as satisfactory representations of the natural hydrologic phenomenon. Another useful method to investigate model suitability is sensitivity analysis (SA), which investigates structural uncertainty in the models. However, a comprehensive method is needed, which led us to develop a model suitability index (MSI) by combining the results of model performance statistics and SA. Here, we assessed and compared the suitability of three rainfall–runoff models (GR4J, IHACRES and Sacramento model) for seven Korean catchments using MSI. MSI showed that the GR4J and IHACRES models are suitable, having more than 0.5 MSI, whereas the Sacramento has less than 0.5 MSI, representing unsuitability for most of the Korean catchments. The MSI developed in this study is a quantitative measure that can be used for the comparison of rainfall–runoff models for different catchments. It uses the results of existing model performance statistics and sensitivity indices; hence, users can easily apply this index to their models and catchments to investigate suitability.

Key words | conceptual rainfall–runoff model, Korean catchments, model performance, model suitability index, sensitivity analysis

INTRODUCTION

Conceptual rainfall–runoff models are widely used to understand and predict the hydrologic responses of catchments of interest. In particular, parsimonious conceptual rainfall–runoff models have benefits of ease of use because of a smaller number of input data required, a faster computational time with reasonable results and a lower structural identifiability problem (Shin et al. 2013). The range of study using the conceptual models is broad: parameter calibration and validation (Vansteenkiste et al. 2014b; Willems 2014; Willems et al. 2014), comparison of objective functions (Jie et al. 2016), sensitivity analysis (SA) (Shin et al. 2013; Van Hoey et al. 2014; Massmann & Holzmann 2015), uncertainty analysis (Wagener et al. 2003; Blasone et al. 2008; Clark et al. 2008; Hailegeorgis & Alfredsen 2015; Shin et al. 2015), regionalisation (Li et al. 2014; Singh et al. 2014; Zhang et al. 2014) and climate change (Chiew et al. 2009; Vaze et al. 2010; Vansteenkiste et al. 2014a; Shin et al. 2016). These various studies demonstrate the usefulness of conceptual rainfall–runoff models.

To investigate whether the conceptual models sufficiently replicate the observed hydrographs, a routine method – the split-sampling test (Klemeš 1986) – is applied for the calibration and validation periods, and model performance statistics are calculated for these periods.
Modellers assess model performance using various objective functions and performance statistics to decide the suitability of the models to their study catchments. Hydrologic models with more parameters tend to have higher performance statistics compared to parsimonious models, but more complex models could have a serious interaction between parameters, which causes uncertainty in the results (Shin et al. 2013).

Another useful method to investigate and compare the suitability of models is SA. If all or many parameters are sensitive, then the models are structurally sound, which means that modellers can trust the results from these models. SA is an essential step to assess model parameters and, therefore, to investigate the model structure (Shin et al. 2013). The sensitivity of model parameters may vary for different study areas (Shin et al. 2013); hence, every study that makes use of hydrologic models with new catchment data needs SA. SA is categorised into global and local methods, and global SA methods are appropriate if there are interactions between model parameters, which results in a non-linear output (Saltelli & Annoni 2010).

Suitable rainfall–runoff models should have both good model performance and less uncertainty. There were several studies and discussions about the suitability or acceptability of models in terms of model performance and structural uncertainty (Beven & Freer 2001; Vrugt et al. 2003; Wagener 2003); however, these are not enough for users to determine which model should be selected as the suitable one. The aim of this study is to provide a guide for suitable model selection considering both model performance and uncertainty. To provide a clearer and easier guide, we developed the model suitability index (MSI) by coupling the results of model performance statistics and global SA. MSI is a quantitative index; hence, users can compare the different rainfall–runoff models using this measure. We compared three conceptual rainfall–runoff models using MSI for seven Korean catchments to investigate which models are suitable or require caution.

The subjects of each section are as follows: the second section describes the SA method, model performance criteria and MSI; the third section briefly describes rainfall–runoff models, catchments, data, the target functions for SA, the calibration method and the method used for this study; the fourth section shows the results of SA, performance statistics and MSI for the three conceptual rainfall–runoff models; and the final section provides a discussion and the conclusions of this study.

SA METHOD, MODEL PERFORMANCE CRITERIA AND MSI

Sobol method for SA

The Sobol method (Sobol 1993) is a variance-based quantitative global SA algorithm that has been applied to various studies (e.g., Nossent et al. 2011; Yang 2011; Herman et al. 2013; Shin et al. 2013). Through the decomposition of the model output variance, this method estimates the relative contributions of individual parameters and parameter interactions to the output variance. The sensitivity of a given parameter is, as a result, quantified by a ratio ranging from 0 to 1. This method calculates a first-order sensitivity index (FSI) and a total sensitivity index (TSI). FSI measures the sensitivity effect of each parameter on the output, whereas TSI measures the effect of each parameter and its interaction with the other parameters on the output (Saltelli et al. 2008).

As mentioned by Saltelli et al. (2008), TSI provides more reliable results than FSI for the investigation of the overall effect of each parameter on the output. Hence, we used TSI only for the investigation of parameter sensitivity, and it can be defined as (Saltelli & Annoni 2010):

\[
TSI = \frac{E_{X_i}(V_{X_i}(Y|X_{-i}))}{V(Y)} \tag{1}
\]

where \(X_i\) is the \(i\)th parameter and \(X_{-i}\) means the vector of all parameters not including \(X_i\). The inner variance in the numerator denotes that the variance of \(Y\), the scalar objective function values, is considered over all the possible values of \(X_i\) while keeping \(X_{-i}\) fixed. The outer expectation operator is considered over all the possible values of \(X_{-i}\), and the variance \(V(Y)\) in the denominator is the total (unconditioned) variance. The interpretation of the numerator of TSI is that it is the expected variance that would be left if all parameters are fixed except for \(X_i\) (Saltelli & Annoni 2010).
We applied Saltelli’s scheme (Saltelli 2002) in the R ‘sensitivity’ package (Pujol et al. 2012), which calculates TSI using the reduced number of samples from $n(2k + 2)$ to $n(k + 2)$. Here, $n$ represents the initial sample size used (10,000 in this study), and $k$ is the number of the parameters of the hydrological model. Hence, the total number of samples generated is 60,000 for four parameter models, and 70,000 and 150,000 for five and 13 parameter models, respectively. More details of the sampling method are described in the work by Shin et al. (2013).

Model performance criteria

The value of the objective function for the calibration of parameters can be used as the model performance statistics. We used three objective function values as model performance statistics to compare the performance of conceptual rainfall–runoff models. The first objective function, the Nash–Sutcliffe efficiency (NSE) (Nash & Sutcliffe 1970), is as follows:

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{n} (Q_{\text{obs}, i} - Q_{\text{sim}, i})^2}{\sum_{i=1}^{n} (Q_{\text{obs}, i} - Q_{\text{obs}})^2}$$  \hspace{1cm} (2)

where $n$ is the number of time steps, $Q_{\text{obs}, i}$ is the observed flow at time step $i$ (daily here), $Q_{\text{obs}}$ is the mean of the observed flow, and $Q_{\text{sim}, i}$ is the simulated flow. The range of NSE is $[-\infty, 1]$, where 1 represents the perfect match between the observed and simulated flow.

The other two objective functions are NSElog, which is NSE with log-transformed data and 0.5(NSE + NSElog). The NSE objective function relatively focuses on fitting high flow (Krause et al. 2005), whereas the NSElog objective function relatively focuses on fitting low flow caused by the log transformation of time series data. The objective function of 0.5(NSE + NSElog) focuses on fitting medium flow by giving equal weight on NSE and NSElog.

Model suitability index

The model performance statistics alone are not enough to decide the suitability of the model as they cannot assess uncertainty in the model. SA is one of the good measures for uncertainty investigation; hence, we developed MSI by coupling the model performance statistics and the SA results. The MSI is as follows:

$$\text{MSI} = 0.5 \times \left( \frac{1}{n} \sum_{i=1}^{n} \text{SR}_i \right) + 0.5 \times \left( \frac{1}{m} \sum_{j=1}^{m} \text{PS}_j \right)$$  \hspace{1cm} (3)

where SR is sensitivity ratio, which is the ratio of the number of sensitive parameters out of the total number of parameters having a range of [0, 1]. For example, if the number of sensitive parameters is three out of four for the GR4J model for the P1 period of the Andong catchment, then the SR is 0.75. If the SRs for the $n$ combination of periods and the target function for the Andong catchment are calculated and averaged, then the averaged value represents the SR of the GR4J model for the Andong catchment. For the sensitivity threshold, we used 0.2 TSI, the same value as that in Shin et al. (2013); hence, parameters having more than 0.2 TSI are sensitive parameters. The value of 0.2 TSI means that the parameter produces at least 20% of the variance of the target function value. The study of van Werkhoven et al. (2009) also suggested this value as the sensitivity threshold but this threshold is still somewhat arbitrary, therefore, care needs to be taken when the TSI values of parameters are near the threshold. PS is the performance statistics, and the averaged PS for the $m$ combination is calculated by applying the same method used for averaged SR calculation. The model performance statistics for the calibration and validation periods are both important; hence, we used both statistics for the calculation of the averaged PS. In addition, we used the NSE series objective functions; thus, the range of objective function values is $[-\infty, 1]$, but the optimised performance statistics are usually placed in $[0, 1]$. Therefore, we concluded that the range of PS is the same as that of SR and calculated the MSI.

We gave the same weight to the averaged SR and PS for the calculation of MSI because we believed that both measures are equally important. The SR is as important as the PS because over-parameterisation due to insensitive parameters is a well-known problem in rainfall–runoff models with large numbers of parameters (van Griensven et al. 2006), and this uncertainty problem could cause the
uncertain model simulation results, such as producing equi-
final multiple simulation hydrographs for the calibration
period (Beven 1993) but producing unequal multiple
simulation hydrographs for the validation periods by apply-
ing the multiple optimal parameter sets from the calibration,
which have significantly variable parameter values (Shin
et al. 2015). Therefore, the modellers could have problems
with decision-making due to this prediction uncertainty.
The SR can be coupled with the PS because we used the
quantitative SA method of Sobol for SR. Therefore, the
calculated MSI values can be used for the comparison of
any model with any catchment data. If the MSI value is
equal to 1, then it means that all the parameters are sensitive
and that the simulated hydrographs are perfectly matched to
the observed hydrographs.

RAINFALL–RUNOFF MODELS, CATCHMENTS,
DATA, TARGET FUNCTIONS AND THE CALIBRATION
METHOD

Rainfall–runoff models

We used three well-known conceptual rainfall–runoff
models with different complexities from four to 13 par-
parameters. The GR4J model (Perrin et al. 2003) with four
parameters uses two unit hydrographs and two stores for
the production and routing of water. The storage of rainfall,
evapotranspiration and percolation in the surface soil are
controlled by the production store, and the routing of effec-
tive rainfall is controlled by the routing store. In the routing
store, effective rainfall is separated by the ratio of 0.9 and
0.1, and then, two unit hydrographs route the portioned
effective rainfall. Quick flow is generated by 10% of effective
rainfall, and slow flow is generated using 90% of effective
rainfall (Le Moine et al. 2008). The GR4J model has been
applied in various studies (Moussu et al. 2011; Andréassian
et al. 2012; Pushpalatha et al. 2012; Shin et al. 2013, 2015).
The IHACRES model has many versions, which are
used in various studies (Viney et al. 2009; Vaze et al. 2010;
Post et al. 2012; Shin et al. 2013, 2015). We used the catchment
moisture deficit (CMD) version (Croke & Jakeman 2004) as
it has more physical meaning for the parameters. The CMD
version of the IHACRES model calculates the varying

catchment moisture at each time step. There are two
stores in the IHACRES model: the nonlinear store for the
generation of effective rainfall using nonlinear function to
deal with the raw rainfall and using accounting equation
to calculate the CMD output, and the linear store to convert
effective rainfall into quick and slow flow using unit hydro-
graphs. The summation of the quick and slow flow at each
time step becomes the total flow. Table 1 shows the par-
parameters of the IHACRES model, and we fixed the ‘e’
parameter as unity as it directly uses the potential evapotrana-
spiration data instead of the temperature data. Hence, five
parameters are used for the IHACRES model.

We used 13 instead of 16 parameters for the Sacramento
model (Burnash et al. 1973), which is the same number of
parameters used by Shin et al. (2015). The Sacramento
model has five runoff components: direct runoff from an
impervious area, surface runoff, interflow, supplementary
base flow and primary base flow. For a brief description of
the process, excess rainfall becomes runoff through a unit hygrograph, and the rest of the rainfall fills various depths
of soil moisture stores, which are interconnected. Loss
through evapotranspiration happens at the soil stores, and
the rest of the water becomes interflow and groundwater.
The summation of the surface and lateral flow become streamflow (Podger 2004). The Sacramento model has
been used in many studies including those of Shin et al.
(2013, 2015), Vaze et al. (2010), Van Werkhoven et al.
(2009), and Petheram et al. (2012).

Table 1 shows the descriptions of the parameters of the
three rainfall–runoff models. The parameter ranges for the
GR4J and Sacramento models are the same as those used
by Shin et al. (2013); hence, readers can refer to Shin et al.
(2013) for more description. We used the hydrologic
models, objective functions and calibration algorithm in
the Hydrological Model Assessment and Development
(Hydromad) (Andrews et al. 2011) modelling package,
which is an open-source software package in R. The Hydromad package is available at http://hydromad.catchment.org.

Catchments

The study area is composed of seven dam catchments in
South Korea (Figure 1). These study sites are unregulated
natural mountainous catchments, and are the major sources

Downloaded from http://iwaponline.com/hr/article-pdf/48/5/1192/365497/hr0481192.pdf by guest
of irrigation and city water. These catchments represent the major hydrologic characteristics of South Korea by having various sizes (Soyanggang, 2,703 km²; Chungju, 6,648 km²; Andong, 1,584 km²; Imha, 1,361 km²; Yongdam, 930 km²; Seomjingang, 763 km²; and Juam, 1,010 km²) and various spatial locations. The 5-year moving average of runoff ratio (%) for the seven catchments is shown in Figure 2. Similar patterns are shown for the runoff ratio between Soyanggang and Chungju, Andong and Imha, and Seomjingang and Juam. These mountainous catchments have relatively large magnitude and long duration of precipitation due to the monsoon climate, therefore, the dominant runoff is relatively large. Figure 2 shows the relatively large runoff ratios of at least 35% that represent the characteristic of wet catchments.

Data

The daily areal rainfall and potential evapotranspiration data for the seven catchments were used as input data for

Table 1 | Description of the parameters of the three rainfall–runoff models

| Parameter name | Parameter no. | Unit | Range | Description |
|----------------|---------------|------|-------|-------------|
| GR4J           |               |      |       |             |
| X1             | 1             | mm   | 50–5,000 | Maximum capacity of the production store |
| X2             | 2             | mm   | –15 to 4 | Groundwater exchange coefficient |
| X3             | 3             | mm   | 10–1,300 | 1-day-ahead maximum capacity of the routing store |
| X4             | 4             | day  | 0.5–5 | Time base of the unit hydrograph, UH1 |
| IHACRES-CMD    |               |      |       |             |
| f              | 1             | –    | 0.5–1.3 | CMD stress threshold as a proportion of d |
| e              | –             | –    | 1 (fixed) | Temperature to potential evapotranspiration (PET) conversion factor |
| d              | 2             | mm   | 50–550 | CMD threshold for producing flow |
| τs (tau_s)     | 3             | day  | 10–1,000 | Time constant for slow flow store |
| τq (tau_q)     | 4             | day  | 0–10 | Time constant for quick flow store |
| v_s            | 5             | –    | 0–1 | Fractional volume for slow flow |
| Sacramento     |               |      |       |             |
| UZTWM          | 1             | mm   | 1–150 | Upper zone tension water maximum capacity |
| UZFWM          | 2             | mm   | 1–150 | Upper zone free water maximum capacity |
| UZK            | 3             | 1/day | 0.1–0.5 | Upper zone free water lateral depletion rate |
| PCTIM          | 4             | –    | 0.000001–0.1 | Fraction of the impervious area |
| ADIMP          | 5             | –    | 0–0.4 | Fraction of the additional impervious area |
| ZPERC          | 6             | –    | 1–250 | Maximum percolation rate coefficient |
| REXP           | 7             | –    | 0–5 | Exponent of the percolation equation |
| LZTWM          | 8             | mm   | 1–500 | Lower zone tension water maximum capacity |
| LZFSM          | 9             | mm   | 1–1,000 | Lower zone supplementary free water maximum capacity |
| LZFPNM         | 10            | mm   | 1–1,000 | Lower zone primary free water maximum capacity |
| LZSK           | 11            | 1/day | 0.01–0.25 | Lower zone supplementary free water depletion rate |
| LZPK           | 12            | 1/day | 0.0001–0.25 | Lower zone primary free water depletion rate |
| PFREE          | 13            | –    | 0–0.6 | Direct percolation fraction from the upper to the lower zone free water storage |
| SIDE           | –             | –    | 0.0 (fixed) | Fraction of base flow that is draining to areas other than the observed channel |
| RSERV          | –             | –    | 0.3 (fixed) | Fraction of the lower zone free water that is unavailable for transpiration purposes |
| RIVA           | –             | –    | 0.0 (fixed) | Fraction of the riparian vegetation area |
the three conceptual rainfall–runoff models, and the observed flow data were used for the calibration and validation of the model parameters. The areal rainfall generated by Thiessen polygon method using data from multiple rainfall gauge stations, and the observed flow data were taken from the Water Resources Management Information System (http://www.wamis.go.kr). The potential evapotranspiration data were generated by the Penmann–Monteith method in the ‘SPEI’ package (Beguería & Vicente-Serrano 2015) using the daily point gauge climate data (maximum and minimum temperatures, wind speed and sunshine hours) from the Korea Meteorological Administration (http://www.kma.go.kr). We carefully selected the climate point gauge stations, which contain both long enough periods and good quality of data. The climate point gauge stations used for the Soyanggang, Chungju, Andong, Imha, Yongdam, Seomjingang and Juam catchments were the Chuncheon, Chungju, Yeongju, Andong, Jangsu, Imsil and Gwangju stations, respectively.

The total data periods are different for the seven dam catchments because of the different years of dam construction. To use the data efficiently, the data were split, having the same length of sub-periods for each catchment but different sub-periods for different catchments (Table 2). Hence, the calibration and validation periods were the same for each catchment but different over the seven catchments.
Table 2 Data period of the seven catchments for calibration and validation

| Catchment | Period name | Data period               |
|-----------|-------------|---------------------------|
| Soyanggang P1 | January 1, 1975–December 31, 1984 |
| P2        | January 1, 1985–December 31, 1994 |
| P3        | January 1, 1995–December 31, 2004 |
| P4        | January 1, 2005–December 31, 2014 |
| Chungju P1 | January 1, 1988–December 31, 1996 |
| P2        | January 1, 1997–December 31, 2005 |
| P3        | January 1, 2006–December 31, 2014 |
| Andong P1  | January 1, 1979–December 31, 1987 |
| P2        | January 1, 1988–December 31, 1996 |
| P3        | January 1, 1997–December 31, 2005 |
| P4        | January 1, 2006–December 31, 2014 |
| Imha P1    | January 1, 1995–December 31, 2004 |
| P2        | January 1, 2005–December 31, 2014 |
| Yongdam P1 | January 1, 2003–December 31, 2008 |
| P2        | January 1, 2009–December 31, 2014 |
| Seomjingang P1 | January 1, 1979–December 31, 1987 |
| P2        | January 1, 1988–December 31, 1996 |
| P3        | January 1, 1997–December 31, 2005 |
| P4        | January 1, 2006–December 31, 2014 |
| Juam P1   | January 1, 1993–December 31, 2003 |
| P2        | January 1, 2004–December 31, 2014 |

The duration of calibration and validation periods were selected for more than 6 years considering the studies of Anctil et al. (2004), Kim et al. (2011) and Yapo et al. (1996). For calibration and validation, 1 year warm-up period was used, that is, the warm-up period for the P1 period of the Soyanggang catchment in Table 2 is January 1, 1974, to December 31, 1974.

Target functions for SA

The target functions for SA are related to the high and low flow taken from Shin et al. (2013): NSE°, NSElog° and F20. The NSE° target function can be defined as follows:

\[
\text{NSE}^° = \frac{1 - \sum_{i=1}^{n} \left( Q_{\text{obs},i} - Q_{\text{sim},i} \right)^2 }{\sum_{i=1}^{n} \left( Q_{\text{obs},i} - Q_{\text{mean}} \right)^2}
\]

\[\text{(4)}\]

The modified NSE° was proposed by Mathevet et al. (2006), and it divides NSE by the NSE with a plus sign. It has the advantage of reducing the influence of large negative values without changing the interpretation of the objective function having a value range of \([-1, 1]\) rather than \([-\infty, 1]\) by NSE. The NSElog° target function is the same as the NSE° target function with log-transformed data. The F20 target function calculates the proportion of days when the modelled flow is below the historically observed 20th percentile flow level, that is, measures the frequency of low flow. The F20 target function is useful in identifying the sensitivity of the parameters related to the very low flow, that is, the ecological flow for a river system. For the SA, we assessed the parameters related to the high flow using the NSE°, and the parameters related to the low flow using the NSElog° and F20.

Calibration method

The shuffled complex evolution (SCE) (Duan et al. 1992, 1993) global optimisation method was used to calibrate the model parameters. The SCE algorithm is one of the popular methods for parameter calibration (Nicklow et al. 2010), and it has been used for various studies until recently (Serrat-Capdevila et al. 2011; Moreno et al. 2012; Shin et al. 2015, 2016). The SCE algorithm combines the strength of four methods: simplex procedure (Nelder & Mead 1965), controlled random search (Price 1987), competitive evolution (Holland 1975) and the concept of complex shuffling (Duan et al. 1992). A brief description of the procedure of the SCE method is as follows:

- **Step 1:** Select the initial ‘population’ randomly throughout the feasible parameter space.
- **Step 2:** Divide the ‘population’ into ‘complexes’, which represent communities that have parents to generate an offspring.
- **Step 3:** Evolve each complex independently by applying a competitive complex evolution strategy.
- **Step 4:** Shuffle the evolved complexes into a single, whole population to share the information.
- **Step 5:** Iterate Steps 2–4 until the results are converged to the preselected threshold.

Method used

First, we performed SA by applying the Sobol method for the parameters of the three models using the three target
functions and the daily data of the seven catchments. Then, we investigated whether each parameter has worked properly for its own purpose and the over-parameterisation problem by counting the number of sensitive parameters.

Second, we calibrated the parameters of the three models by applying the SCE algorithm using the three objective functions and the daily data of the seven catchments. Then, we calculated the model performance statistics for the calibration and validation periods, and compared the three models. To investigate the overall model ability of replication of natural flow and prediction, we generated boxplots using the performance statistics for each model of the seven catchments, and the objective function grouped by calibration and validation periods. A larger dispersion of the boxplot for the validation period represents a lower prediction ability.

Lastly, we assessed the suitability of the rainfall–runoff models by applying MSI. We compared the three conceptual models for the seven catchments using the bar plot of MSI. The bar plot is an easy and quick method to investigate which models are suitable and which models require caution.

**RESULTS**

**SA for the parameters of the three conceptual models**

The maximum and minimum TSIs of the parameters of the three rainfall–runoff models for the seven catchments are shown in Figures 3–5. Each plot shows the results for the NSE*, NSElog* and F20 objective functions. The maximum and minimum TSIs represent the variability of the parameter sensitivity over the different data periods for the catchment of interest with respect to the target function of interest. If the maximum and minimum TSIs are on the diagonal line (one-to-one line) as they are equal for a parameter, then that parameter has the same TSI for all periods so these parameters are insensitive regardless of the data period. The threshold for the sensitivity of TSI in this study is 0.2, as mentioned in sub-section ‘MSI’.

For the GR4J model, parameter no. 4 (X4) related to the unit hydrograph is sensitive to the high flow target function (Figure 3(a)), and parameter no. 2 (X2) related to the groundwater is sensitive to the low flow target function (Figure 3(b)). Parameter no. 3 (X3) related to the routing store of 1-day-ahead is more sensitive to the low flow target function (Figure 3(c)) as it is related to the antecedent soil moisture condition. Parameter no. 1 (X1) is always sensitive regardless of the target functions except for the Imha and Seomjingang catchments for the NSElog* target function (Figure 3(b)) because it is for producing the streamflow. It is worth noting that the NSElog* target function is a statistical measure, which calculates the overall performance by comparison of modelled and observed streamflow, therefore, it sometimes produces ‘type I error’ (incorrectly classifying a sensitive parameter as being insensitive). As the F20 target function is a scalar value calculated from the modelled streamflow time series, it is sometimes more reliable to judge the sensitivity of the low flow parameter compared to the NSElog* target function. Figure 3(c) shows that parameter no. 1 (X1) is sensitive to the F20 target function; hence, this parameter is sensitive to the low flow target function. Also note that this parameter could be less sensitive to the F20 target function if the characteristics of catchment and data are very dry; therefore, there is not enough information in the data for SA as demonstrated by Shin et al. (2013). The seven study catchments and data have wet characteristics as mentioned in the sub-section ‘Catchments’, therefore, these catchments are not cases for less sensitivity for parameter no. 1 (X1). Nos. 3 (for Juam), 1 (for Chungju) and 3 (for Yongdam) parameters for the NSE* target function (Figure 5(a)) are near the diagonal line that represents similar maximum and minimum TSIs for all periods. This model would not have insensitive nos. 1 and 3 parameters for this target function if a different data period for these catchments was used.

For the IHACRES model, parameter no. 5 (νs) related to slow flow is sensitive to all target functions (Figure 4) as this parameter is also directly related to the quick flow calculated as 1 – the volume of quick flow. Parameter no. 4 (τq) related to the quick flow is sensitive to the NSE* target function, and parameter no. 1 (f) related to the evapotranspiration is relatively sensitive to the NSElog* and F20 target functions with respect to the maximum TSI as expected. Parameter no. 2 (d), which is related to the soil depth for flow production, is more sensitive to the NSE* target function with respect to the maximum TSI than the
other target functions related to the low flow as the change of high flow magnitude by the change of this parameter is relatively larger than the change of low flow magnitude. Parameter no. 3 ($\tau_s$) related to the slow flow is more sensitive to the $F_{20}$ target function (Figure 4(f)) compared to the sensitivity of this parameter to the NSElog* target function (Figure 4(d)) with respect to the maximum TSI. However, the minimum TSI values in Figure 4(d) and 4(f) have much lower TSI values, which means that the IHACRES model could be relatively weak for the low flow simulation.

The parameter nos. 1 and 3 for the Juam and Seomjingang catchments are sensitive to the $F_{20}$ target function in terms of the maximum TSI (Figure 4(e) and 4(f)) but the sensitivities of these parameters are below the threshold for the Chungju catchment (Figure 4(f)). This represents that the parameter sensitivity is dependent on the catchment’s characteristics.

For the Sacramento model, parameter nos. 2, 5, 8, 9, 11 and 12 are relatively sensitive (at least one parameter is sensitive) to the NSE* target function with respect to the

---

**Figure 3** | Plot of the Sobol total sensitivity indices (TSI) for the different target functions and the GR4j model parameters for the seven catchments. Minimum and maximum TSI were calculated from the period values. (a) NSE*, (b) NSElog* and (c) $F_{20}$. Diagonal line is one-to-one line.
maximum TSI; parameter nos. 1, 12 and 13, to the NSElog* target function; and parameter nos. 1, 11 and 12, to the F20 target function. Hence, eight parameters (nos. 1, 2, 5, 8, 9, 11, 12 and 13) are relatively sensitive, and five parameters (nos. 3, 4, 6, 7 and 10) are insensitive to both high and low flow target functions. These insensitive parameters...
Figure 5 | Plot of the Sobol total sensitivity indices (TSI) for the different target functions and the Sacramento model parameters for the seven catchments. Minimum and maximum TSI were calculated from the period values. (a) NSE*, (b) zoomed image of (a), (c) NSElog*, (d) zoomed image of (c), (e) F20 and (f) zoomed image of (e). Diagonal line is one-to-one line.
could cause an identifiability problem through a serious interaction with the other parameters (Shin et al. 2015). Therefore, the Sacramento model has parametric and structural identifiability problems.

**Boxplots to assess the overall sensitivity of the parameters**

To investigate the overall sensitivity of the parameters for the Korean catchments, we plotted the boxplots using the TSI values over the seven catchments for the three conceptual models for each target function. Although the whisker of the X1 parameter for the NSElog* target function is slightly less sensitive, all the parameters of the GR4J model are sensitive to the related target function (Figure 6), which complements the results in the sub-section ‘SA for the parameters of the three conceptual models’. The parameters related to the high flow (tau_q and v_s) of the IHACRES model are sensitive to the NSE* target function, and the parameter d has a relatively lower sensitivity over the three target functions (Figure 7). The sensitivity of parameter f, which is related to the evapotranspiration, to the NSElog* and F20 target functions depends on the characteristics of the data that are shown by the large variation of the boxplot over the different catchments and periods. The boxplot of parameter tau_s is below or across the threshold of 0.2 TSI to the low flow target functions (NSElog* and F20), which implies that the IHACRES model could have a lower model performance for the low flow simulation for some periods and/or catchments.

The Sacramento model has five relatively sensitive parameters, which are over the threshold, at least the whisker,
to the NSE* target function (UZFWM, ADIMP, LZTWM, LZSK and LZPK), four relatively sensitive parameters to the NSElog* target function (UZTWM, LZSK, LZPK and PFREE) and three relatively sensitive parameters to the F20 target functions (UZTWM, LZSK and LZPK) (Figure 8), which means that around seven parameters are relatively sensitive (UZTWM, UZFWM, ADIMP, LZTWM, LZSK, LZPK and PFREE) and six parameters are not sensitive enough over the three target functions. These boxplots also support the results of Shin et al. (2013, 2015) in terms of the uncertainty of the Sacramento model due to over-parameterisation. Interestingly, parameter no. 13 (PFREE), which is insensitive as demonstrated by Shin et al. (2013, 2015), is sensitive to the NSElog* target function; hence, the sensitivity of the parameters can change with the data having very different hydrological characteristics.

**Comparison of the three model performances for the three objective functions**

Figures 9–11 show the three model performances over the seven catchments for the three objective functions related to fitting high, intermediate and low flow. The model performances of the three models are shown for each catchment (refer to the titles of each figure). The subtitles represent the calibration periods (e.g., P1 and P2). The x scale for each subfigure represents the simulation periods, and the y scale represents the values of the objective function. As shown in the legend, the cross and the circle in the figures represent the objective function values for the calibration and validation periods, respectively.

In Figure 9, for the NSE objective function, the three models have good model performances over the seven catchments except for the P2 periods of the Seomjingang
catchment. For the P2 periods, the three models have relatively low model performances for the calibration and validation periods, which also occurs in Figures 10 and 11. It could be due to the significantly lowered runoff ratio from 1990 to 1994 in Figure 2 (5-year moving average) or a significant error in the observed data.

The GR4J model, with only four parameters, dominates the other two models in terms of model performances with the NSE objective function for four catchments (Soyanggang, Andong, Imha and Juam) and has similar model performances for two catchments (Yongdam and Seomjingang) (Figure 9). For the Chungju catchment, the GR4J model has a slightly lower model performance than the other models, but still has a good performance having more than 0.7 NSE. Overall, the Sacramento and IHACRES models follow the GR4J model, with a reasonable model performance for high flow. The GR4J model has a higher or similar model performance for the \(0.5(\text{NSE} + \text{NSElog})\) and NSElog objective functions over the seven catchments compared to the other two models.

For a more detailed analysis of the Chungju catchment, the GR4J model has a similar or slightly lower model performance than the other models for the NSE objective function, but this model has a similar model performance for the \(0.5(\text{NSE} + \text{NSElog})\) objective function, and a slightly higher or similar model performance for the NSElog objective function. The model performances of the IHACRES model with the NSE objective function is higher than with the NSElog objective function, whereas the model performances of the GR4J are opposite, which implies that the GR4J model is more suitable for the low flow simulation and the IHACRES model is more suitable for the high flow simulation for some catchments. These results complement the study of Shin et al. (2013). For the P3 calibration period and other periods for validation, the Sacramento model has similar or better model performance than the other models.
performances for the NSE and $0.5(NSE + NSE_{log})$ objective functions but a lower model performance for the $NSE_{log}$ objective function than the other two models. It means that more parameters for groundwater in the Sacramento model have not functioned enough for low flow fitting due to the insensitivity discussed in the subsection ‘Boxplots to assess the overall sensitivity of the parameters’.

Figure 9 | IHACRES, GR4J and Sacramento model performances for the NSE objective function.
Boxplots to assess the overall replicability and predictability of the models

Figures 12 and 13 show the boxplots of the model performances for each objective function using all the results of the seven catchments. Figure 12, which is for calibration, is to investigate the replicability of the natural flow of the models, and Figure 13, which is for validation, is to investigate the predictability of the models for each objective function.
In Figure 12, the GR4J model has a higher or similar model performance overall. The median value of NSE is higher than the other two models. The range of the boxplot for the 0.5(NSE + NSElog) performance statistic is narrower, which represents a more stable medium flow replicability than the other models. The Sacramento model follows the GR4J model (see the median value of the boxplot), and the IHARES model has a relatively lower model performance but has a narrow range over the objective functions, representing a more stable replicability of the catchment response.

Figure 11 | IHACRES, GR4J and Sacramento model performances for the NSElog objective function.
In Figure 13, the GR4J model has the highest median value of the boxplot for NSE values, but the range of NSE values is relatively higher than the other models; hence, it has a relatively unstable prediction ability for high flow. However, the difference of the NSE value range between the GR4J model and Sacramento model is less than 0.1, representing little difference. This phenomenon was also found in the results for the $0.5(\text{NSE} + \text{NSElog})$ and NSElog objective functions. The model performances of the Sacramento model follow those of the GR4J model, and the IHACRES model also shows reasonable model performances with more than 0.5 of performance statistics.

Comparison of rainfall–runoff models using MSI

It is hard to judge which model is more suitable when one considers the sensitivity indices and performance statistics from the result of the sub-sections above. For example, the GR4J model had good and clear parameter sensitivities capturing their purposes, and the median values of the model performance statistics were higher than the other models. However, the range of the performance statistics were sometimes wider than the other models; hence, the simulation uncertainty was higher. The IHACRES model had better parameter sensitivities than the Sacramento model but the median values of the model performance statistics were
similar to or lower than the Sacramento model. The higher narrowness of the performance range compared to the Sacramento model depended on the objective functions. The Sacramento model had more uncertainty due to possessing more insensitive parameters than the other models, but the median values of the model performance statistics were better than the IHACRES model. The MSI, which comprehensively considers both sensitivity indices and performance statistics, has the strength of ease and clearness to judge the superiority and inferiority of the models compared to the other two methods.

We decided that 0.5 of MSI is the threshold of good MSI as NSE values greater than 0.5 are good (Moriasi et al. 2007). Moreover, the study of van Werkhoven et al. (2009) undertook the Sobol SA of the Sacramento model in order to fix insensitive parameters that have TSIs less than 0.2 threshold while maintaining the same predictive performance to the non-fixed parameter case. As a result, this sensitivity threshold reduced around half of the total number of Sacramento parameters. As more fixed parameters result in the poorer model performances, we believed that the model with more than half of the sensitive parameters for the arbitrary target function is suitable. As described in the sub-section ‘MSI’, we give the same weight to model performance and sensitivity; therefore, the threshold value of good MSI is 0.5.

The MSIs using all the model performances and sensitivity indices in the sub-sections above are shown in Figure 14. In Figure 14, the GR4J model has the highest MSI over the seven catchments, followed by the IHACRES model. Hence, the GR4J and IHACRES models are suitable models, at least for these seven Korean catchments. Interestingly, the Sacramento model has slightly higher performance statistics than the IHACRES model in the sub-sections ‘Comparison of the three model performances for the three objective functions’ and ‘Boxplots to assess the overall replicability and predictability of the models’; however, the MSI shows opposite results due to the many insensitive parameters in the Sacramento model shown in sub-sections ‘SA for the parameters of the three conceptual models’ and ‘Boxplots to assess the overall sensitivity of the parameters’. The Sacramento model has many MSI values of less than 0.5 except for the Juam catchment. The threshold of MSI is reasonable because it is based on the two aforementioned studies; therefore, the Sacramento model is not suitable to be used for the six Korean catchments.

**DISCUSSION AND CONCLUSIONS**

MSI considers both model performance and uncertainty quantitatively; hence, it can be applied to any model and any catchment for the purpose of comparison. Furthermore, this index is very flexible as it uses various target functions and objective functions for sensitive analysis and performance statistics, respectively. This reliability of MSI can be improved by adding more sensitivity indices and performance statistics. We used the quantitative SA method of Sobol, but the qualitative SA method, such as the Morris (1991) method, cannot be used for MSI calculation as MSI comparisons between models or catchments are not...
possible. We gave the same weight to the model performance and sensitivity results, but this weight can be adjusted by users with valid reasons. The analysis of sensitivity to the weights given to these two parts is out of bounds for this study so we have not done this analysis. However, this analysis can be an interesting study because the MSI can be changed with different weights; therefore, this work will be researched in the future.

The parameters of the models used in this study were related to the input rainfall and its response of runoff, therefore, they can be used for the SA, model performance test and, finally, the MSI analysis. However, if some of the parameters from a conceptual model are related to snow accumulation and melt, and are applied to some tropical catchments, then the MSI of the model will be low because the parameters will be insensitive, even if the other parameters of the model capture the runoff reasonably. Therefore, the selection of relevant model parameters is an essential prerequisite step before the MSI analysis.

While this paper focused on three conceptual rainfall–runoff models, the method can be applied to other physically based distributed rainfall–runoff models. The detailed analysis for the sensitivity indices in the sub-section ‘SA for the parameters of the three conceptual models’ can be used to investigate the parametric and structural uncertainty of the model, and the detailed analysis for the model performances in the sub-section ‘Comparison of the three model performances for the three objective functions’ can be used to investigate the strength of the model, i.e., the extent of capturing the dynamics of rainfall–runoff processes. The overall analysis using all the sensitivity indices or model performances in the sub-sections ‘Boxplots to assess the overall sensitivity of the parameters’ and ‘Boxplots to assess the overall replicability and predictability of the models’ can be a useful tool for gaining intuition about model suitability. The MSI analysis in the sub-section ‘Comparison of rainfall–runoff models using MSI’ can be a pragmatic measure to screen out the suitable model.

To summarise, we have investigated the suitability of three conceptual rainfall–runoff models for the Korean catchments through the comparison of the sensitivity of model parameters, model performances and MSI by coupling the results of SA and the model performance statistics. We used three well-known and widely used hydrological models – GR4J, IHACRES and Sacramento – for the seven unregulated dam catchments, which have different periods of data sets from 12 to 40 years. Three target functions for high, low and very low flow were used for SA, and three objective functions for high, medium and low flow were used for calibration. The SCE algorithm was used for parameter estimation, and the split-sample independent data sets were used for the calibration and validation periods.

As a result, the parameters of the GR4J model had good sensitivity, representing well-identified parameters, and this model with only four parameters had the best model performance overall; hence, it had the best MSI values. The IHACRES model had well-identified parameters overall, but had one parameter with a relatively low sensitivity and this parameter could cause a lower performance for the low flow simulation. The IHACRES model ranked as second best in terms of MSI values over the seven catchments. Both GR4J and IHACRES were suitable models, with more than 0.5 of MSI. However, the Sacramento model had MSI values of less than 0.5 for most catchments as it has many insensitive parameters. The Sacramento model had an over-parameterisation problem, and this uncertainty problem could be propagated to the model results as demonstrated by Shin et al. (2015); hence, care needs to be taken. The Sacramento model has relatively more parameters for groundwater simulation compared to the GR4J and IHACRES models, but its model performance for the low flow simulation was not sufficiently better. This implies that the use of more parameters cannot guarantee better model performance, as warned by Jakeman & Hornberger (1993).

REFERENCES

Anctil, F., Perrin, C. & Andréassian, V. 2004 Impact of the length of observed records on the performance of ANN and of conceptual parsimonious rainfall-runoff forecasting models. Environ. Modell. Softw. 19 (4), 557–568.
Andréassian, V., Lerat, J., Le Moine, N. & Perrin, C. 2012 Neighboors: nature’s own hydrological models. J. Hydrol. 414–415 (11), 49–58.
Andrews, F. T., Croke, B. F. W. & Jakeman, A. J. 2011 An open software environment for hydrological model assessment and development. Environ. Modell. Softw. 26 (10), 1171–1185.
Beguería, S. & Vicente-Serrano, S. M. 2005 Package ‘SPEI’. R Package Version 1.6. http://CRAN.R-project.org/package=SPEI.
Beven, K. 1993 Prophecy, reality and uncertainty in distributed hydrological modelling. *Adv. Water Resour.* 16 (1), 41–51.

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

Blasone, R. S., Vrugt, J. A., Madsen, H., Rosbjerg, D., Robinson, B. A. & Zyvoloski, G. A. 2008 Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov Chain Monte Carlo sampling. *Adv. Water Resour.* 31 (4), 630–648.

Burnash, R. J. C., Ferral, R. L. & McGuire, R. A. 1973 A Generalized Streamflow Simulation System: Conceptual Modeling for Digital Computers. Technical Report, US National Weather Service, Sacramento, CA, USA.

Chiew, F. H. S., Teng, J., Vaze, J., Post, D. A., Perraud, J. M., Kirono, D. G. C. & Viney, N. R. 2009 Estimating climate change impact on runoff across southeast Australia: method, results, and implications of the modeling method. *Water Resour. Res.* 45 (10).

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T. & Hay, L. E. 2008 Framework for understanding structural errors (FUSE): a modular framework to diagnose differences between hydrological models. *Water Resour. Res.* 44 (12).

Croke, B. F. W. & Jakeman, A. J. 2004 A catchment moisture deficit module for the IHACRES rainfall-runoff model. *Environ. Modell. Softw.* 19 (1), 1–5.

Duan, Q., Sorooshian, S. & Gupta, V. 1992 Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resour. Res.* 28 (4), 1015–1031.

Duan, Q. Y., Gupta, V. K. & Sorooshian, S. 1993 Shuffled complex evolution approach for efficient and effective global minimization. *J. Optimiz. Theory App.* 76 (3), 501–521.

Hailegeorgis, T. T. & Alfredsen, K. 2015 Comparative evaluation of performances of different conceptualisations of distributed HBV runoff response routines for prediction of hourly streamflow in boreal mountainous catchments. *Hydrol. Res.* 46 (4), 607–628.

Herman, J. D., Kollat, J. B., Reed, P. M. & Wagener, T. 2013 Technical note: method of Morris effectively reduces the computational demands of global sensitivity analysis for distributed watershed models. *Hydrol. Earth Syst. Sci.* 17 (7), 2893–2903.

Holland, J. H. 1975 Adaptation in Natural and Artificial System: An Introduction with Application to Biology, Control and Artificial Intelligence. Michigan University Press, Ann Arbor, MI, USA.

Jakeman, A. J. & Hornberger, G. M. 1993 How much complexity is warranted in a rainfall-runoff model? *Water Resour. Res.* 29 (8), 2637–2649.

Jie, M. X., Chen, H., Xu, C.-Y., Zeng, Q. & Tao, X. E. 2016 A comparative study of different objective functions to improve the flood forecasting accuracy. *Hydrol. Res.* 47 (4), 718–735. doi:10.2166/nh.2015.078.

Kim, H. S., Croke, B. F. W., Jakeman, A. J. & Chiew, F. H. S. 2011 An assessment of modelling capacity to identify the impacts of climate variability on catchment hydrology. *Math. Comput. Simulat.* 81 (7), 1419–1429.

Klemeš, V. 1986 Operational testing of hydrological simulation models. *Hydrol. Sci. J.* 31 (1), 13–24.

Krause, P., Boyle, D. P. & Bäse, F. 2005 Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.* 5, 89–97.

Le Moine, N., Andréassian, V. & Mathevet, T. 2008 Confronting surface- and groundwater balances on the La Rochefoucauld-Touvre karstic system (Charente, France). *Water Resour. Res.* 44 (5), W05403.

Li, F., Zhang, Y., Xu, Z., Liu, C., Zhou, Y. & Liu, W. 2014 Runoff predictions in ungaged catchments in southeast Tibetan Plateau. *J. Hydrol.* 511, 28–38.

Massmann, C. & Holzmann, H. 2015 Analysing the sub-processes of a conceptual rainfall-runoff model using information about the parameter sensitivity and variance. *Environ. Model. Assess.* 20 (1), 41–53.

Mathevet, T., Michel, C., Andréassian, V. & Perrin, C. 2006 A bounded version of the Nash–Sutcliffe criterion for better model assessment on large sets of basins. In: *Large Sample Basin Experiments for Hydrological Model Parameterization: Results of the Model Parameter Experiment–MOPEX*. V. Andréassian, A. Hall, N. Chahinian & J. Schaake (eds), IAHS Publ., 307, Wallingford, UK, pp. 211–219.

Moreno, H. A., Vivoni, E. R. & Ghochis, D. J. 2012 Utility of quantitative precipitation estimates for high resolution hydrologic forecasts in mountain watersheds of the Colorado front range. *J. Hydrol.* 438, 66–83.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D. & Veith, T. L. 2007 Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* 50 (3), 885–900.

Morris, M. D. 1991 Factorial sampling plans for preliminary computational experiments. *Technometrics* 33 (2), 161–174.

Moussu, P., Oudin, L., Plagnes, V., Mangin, A. & Bendjoudi, H. 2011 A multi-objective calibration framework for rainfall-discharge models applied to karst systems. *J. Hydrol.* 400 (3–4), 364–376.

Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models part I–A discussion of principles. *J. Hydrol.* 10 (3), 282–290.

Nelder, J. A. & Mead, R. 1965 A simplex method for function minimization. *Comput. J.* 7 (4), 308–315.

Nicklow, J., Reed, P., Savic, D., Dessalegne, T., Harrell, L., Chan-Hilton, A., Karamouz, M., Minser, B., Ostfeld, A., Singh, A. & Zechman, E. 2010 State of the art for genetic algorithms and beyond in water resources planning and management. *J. Water Res. Pl.–ASCE* 136 (4), 412–432.

Nossent, J., Elsen, P. & Bauwens, W. 2011 Sobol sensitivity analysis of a complex environmental model. *Environ. Modell. Softw.* 26 (12), 1515–1525.

Perrin, C., Michel, C. & Andrèassian, V. 2005 Improvement of a parsimonious model for streamflow simulation. *J. Hydrol.* 279 (1), 275–289.

Petheram, C., Rustomji, P., Chiew, F. H. S. & Vleeshouwer, J. 2012 Rainfall-runoff modelling in northern Australia: a guide to modelling strategies in the tropics. *J. Hydrol.* 462–463, 28–41.
Podger, G. 2004 RRL Rainfall Runoff Library User Guide. Cooperative Research Centre for Catchment Hydrology. http://www.toolkit.net.au.

Post, D. A., Chiew, F. H. S., Teng, J., Viney, N. R., Ling, F. L. N., Harrington, G., Crosbie, R. S., Graham, B., Marvanek, S. & McLoughlin, R. 2012 A robust methodology for conducting large-scale assessments of current and future water availability and use: a case study in Tasmania, Australia. J. Hydrol. 412–413, 233–245.

Price, W. L. 1987 Global optimization algorithms for a CAD workstation. J. Optimiz. Theory App. 55 (1), 133–146.

Pujol, G., Iooss, B. & Janon, A. 2012 Package ‘sensitivity’: Sensitivity Analysis. R package version 1.5. Available from: http://CRAN.R-project.org/package=sensitivity.

Pushpalatha, R., Perrin, C., Le Moine, N. & Andréassian, V. 2012 A review of efficiency criteria suitable for evaluating low-flow simulations. J. Hydrol. 420–421, 171–182.

Saltelli, A. 2002 Making best use of model evaluations to compute sensitivity indices. Comput. Phys. Commun. 145 (2), 280–297.

Saltelli, A. & Annoni, P. 2010 How to avoid a perfidious sensitivity analysis. Environ. Modell. Softwa. 25 (12), 1508–1517.

Saltelli, A., Chan, K. & Scott, E. M. 2008 Sensitivity Analysis. Wiley, New York, USA.

Serrat-Capdevila, A., Scott, R. L., James Shuttleworth, W. & Valdés, J. B. 2011 Estimating evapotranspiration under warmer climates: insights from a semi-arid riparian system. J. Hydrol. 399 (1–2), 1–11.

Shin, M.-J., Guillaume, J. H. A., Croke, B. F. W. & Jakeman, A. J. 2013 Addressing ten questions about conceptual rainfall-runoff models with global sensitivity analyses in R. J. Hydrol. 503, 135–152.

Shin, M.-J., Guillaume, J. H. A., Croke, B. F. W. & Jakeman, A. J. 2015 A review of foundational methods for checking the structural identifiability of models: results for rainfall-runoff. J. Hydrol. 520, 1–16.

Shin, M.-J., Eum, H.-I., Kim, C.-S. & Jung, I.-W. 2016 Alteration of hydrologic indicators for Korean catchments under CMIP5 climate projections. Hydro. Process. 30 (24), 4517–4542. doi:10.1002/hyp.10948.

Singh, R., Archfield, S. A. & Wagener, T. 2014 Identifying dominant controls on hydrologic parameter transfer from gauged to ungauged catchments—A comparative hydrology approach. J. Hydrol. 517, 985–996.

Sobol, I. M. 1995 Sensitivity analysis for nonlinear mathematical models. Math. Model. Comput. Exp. 1 (4), 407–414.

van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M. & Srinivasan, R. 2006 A global sensitivity analysis tool for the parameters of multi-variable catchment models. J. Hydrol. 324 (1–4), 10–23.

Van Hoey, S., Seuntjens, P., van der Kwaast, J. & Nopens, I. 2014 A qualitative model structure sensitivity analysis method to support model selection. J. Hydrol. 519, 3426–3435.

van Werkhoven, K., Wagener, T., Reed, P. & Tang, Y. 2009 Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models. Adv. Water Resour. 32 (6), 1154–1169.

Vansteenkiste, T., Tavakoli, M., Ntegeka, V., De Smedt, F., Bataeaan, O., Pereira, F. & Willems, P. 2014 Intercomparison of hydrological model structures and calibration approaches in climate scenario impact projections. J. Hydrol. 519, 743–755.

Vansteenkiste, T., Tavakoli, M., Van Steenbergen, N., De Smedt, F., Bataeaan, O., Pereira, F. & Willems, P. 2014b Intercomparison of five lumped and distributed models for catchment runoff and extreme flow simulation. J. Hydrol. 511, 335–349.

Vaze, J., Post, D. A., Chiew, F. H. S., Perraud, J. M., Viney, N. R. & Teng, J. 2010 Climate non-stationarity—validity of calibrated rainfall-runoff models for use in climate change studies. J. Hydrol. 394 (3–4), 447–457.

Viney, N. R., Bormann, H., Breuer, L., Bronstert, A., Croke, B. F. W., Frede, H., Gräff, T., Hubrechts, L., Huismans, J. A., Jakeman, A. J., Kite, G. W., Lanini, J., Leavesley, G., Lettenmaier, D. P., Lindström, G., Seibert, J., Sivapalan, M. & Willems, P. 2009 Assessing the impact of land use change on hydrology by ensemble modelling (LUCHEM) II: ensemble combinations and predictions. Adv. Water Resour. 32 (2), 147–158.

Vrugt, J. A., Gupta, H. V., Bouten, W. & Sorrooshian, S. 2003 A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. Water Resour. Res. 39 (8), 1201.

Wagener, T. 2003 Evaluation of catchment models. Hydro. Process. 17 (16), 3375–3378.

Wagener, T., McIntyre, N., Lees, M. J., Wheater, H. S. & Gupta, H. V. 2005 Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. Hydro. Process. 17 (2), 455–476.

Willems, P. 2014 Parsimonious rainfall-runoff model construction supported by time series processing and validation of hydrological extremes – Part 1: Step-wise model-structure identification and calibration approach. J. Hydrol. 510, 578–590.

Willems, P., Mora, D., Vansteenkiste, T., Taye, M. T. & Van Steenbergen, N. 2014 Parsimonious rainfall-runoff model construction supported by time series processing and validation of hydrological extremes – Part 2: Intercomparison of models and calibration approaches. J. Hydrol. 510, 591–609.

Yang, J. 2011 Convergence and uncertainty analyses in Monte-Carlo based sensitivity analysis. Environ. Modell. Softwa. 26 (4), 444–457.

Yapo, P. O., Gupta, H. V. & Sorooshian, S. 1996 Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. J. Hydrol. 181 (1–4), 23–48.

Zhang, Y., Vaze, J., Chiew, F. H., Teng, J. & Li, M. 2014 Predicting hydrological signatures in ungauged catchments using spatial interpolation, index model, and rainfall-runoff modelling. J. Hydrol. 517, 936–948.