Article

Impact of Dataset Size on the Signature-Based Calibration of a Hydrological Model

Safa A. Mohammed ^{1,2}, Dimitri P. Solomatine ^{2,3,4}, Markus Hrachowitz ³ and Mohamed A. Hamouda ^{1,5,*}

¹ Department of Civil and Environmental Engineering, United Arab Emirates University, Al Ain P.O. Box 15551, United Arab Emirates; 201990188@uaeu.ac.ae

² IHE-Delft, Institute for Water Education, P.O. Box 3015, 2601 DA Delft, The Netherlands; d.solomatine@un-ihe.org

³ Water Resources Section, Faculty of Civil Engineering and Applied Geosciences, Delft University of Technology, P.O. Box 5048, 2628 CN Delft, The Netherlands; solomatine@un-ihe.org (D.P.S.); M.Hrachowitz@tudelft.nl (M.H.)

⁴ Water Problems Institute of RAS (Russian Academy of Sciences), 119991 Moscow, Russia

⁵ National Water and Energy Center, United Arab Emirates University, Al Ain P.O. Box 15551, United Arab Emirates

* Correspondence: m.hamouda@uaeu.ac.ae; Tel.: +971-3-713-5155

Abstract: Many calibrated hydrological models are inconsistent with the behavioral functions of catchments and do not fully represent the catchments' underlying processes despite their seemingly adequate performance, if measured by traditional statistical error metrics. Using such metrics for calibration is hindered if only short-term data are available. This study investigated the influence of varying lengths of streamflow observation records on model calibration and evaluated the usefulness of a signature-based calibration approach in conceptual rainfall-runoff model calibration. Scenarios of continuous short-period observations were used to emulate poorly gauged catchments. Two approaches were employed to calibrate the HBV model for the Brue catchment in the UK. The first approach used single-objective optimization to maximize Nash–Sutcliffe efficiency (NSE) as a goodness-of-fit measure. The second approach involved multiobjective optimization based on maximizing the scores of 11 signature indices, as well as maximizing NSE. In addition, a diagnostic model evaluation approach was used to evaluate both model performance and behavioral consistency. The results showed that the HBV model was successfully calibrated using short-term datasets with a lower limit of approximately four months of data (10% FRD model). One formulation of the multiobjective signature-based optimization approach yielded the highest performance and hydrological consistency among all parameterization algorithms. The diagnostic model evaluation enabled the selection of consistent models reflecting catchment behavior and allowed an accurate detection of deficiencies in other models. It can be argued that signature-based calibration can be employed for building adequate models even in data-poor situations.

Keywords: HBV model; hydrological signatures; multiobjective optimization; diagnostic evaluation approach; dataset size; lumped model calibration; poorly gauged catchments; Brue catchment

Citation: Mohammed, S.A.; Solomatine, D.P.; Hrachowitz, M.; Hamouda, M.A. Impact of Dataset Size on the Signature-Based Calibration of a Hydrological Model. *Water* **2021**, *13*, 970. <https://doi.org/10.3390/w13070970>

Academic Editor: Xing Fang

Received: 8 March 2021

Accepted: 29 March 2021

Published: 31 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Model calibration in a hydrological modeling context entails finding the most appropriate set of parameters to obtain the best model outputs resembling the observed system's behavior. Model calibration can be performed manually; however, it is an inefficient method because it is time-consuming and depends on the modeler's experience. Therefore, much effort has been made over the past decades to develop effective and efficient calibration methods such as automated (computer-based) calibration, especially in the view of advances in computer technology and algorithmic support for solving optimization problems [1,2]. Various metrics are used in model calibration. The most widely used

metrics are borrowed from classical statistical approaches, such as minimizing squared residuals (the difference between the observations and model simulation outputs), maximizing the correlation coefficient, or aggregating several metrics such as the Kling–Gupta efficiency [1,3–5].

The calibration of a hydrological model requires multiobjective optimization because no single metric can fully describe a simulation error distribution [6,7]. For the past two decades, evolutionary-based multiobjective optimization has been used for hydrological models [8]. Multiobjective optimization has a broad range of applications in engineering and water-resource management, particularly in hydrological simulations [8,9]. For exposure to this topic, readers may be directed, e.g., to Efstratiadis and Koutsoyiannis (2010) who reviewed several case studies that included multiobjective applications in hydrology [10]. In hydrological model calibration, multiobjective optimization can tradeoff between conflicting calibration objectives and can define a solution corresponding to Pareto front's knee point, which is nearest to the optimum point and can be considered the best individual tradeoff solution [11,12].

Many researchers have argued that calibration of rainfall-runoff models should not be limited to ensuring the fitness of model simulations to observations; it should also be able to produce other hydrological variables to ensure robust model performance and consistency [13]. Martinez and Gupta (2011) approached the concept of hydrological consistency by recommending that the model structures and parameters produced in the classical maximum likelihood estimation should be constrained to replicate the hydrological features of the targeted process [14]. Euser et al. (2013) defined consistency as “the ability of a model structure to adequately reproduce several hydrological signatures simultaneously while using the same set of parameter values” [13]. To improve model calibration, hydrological signatures as objective functions have received more attention over the last decade [7,15–19]. Hydrological signatures reflect the functional behavior of a catchment [20], allowing the extraction of maximum information from the available data [21–26]. New model calibration metrics are continuously being developed to identify optimal solutions that are more representative of the employed hydrological signatures [7,16,17,19,27–29]. However, to the best of our knowledge, Shafii and Tolson (2015) were the first to consider numerous hydrological signatures with multiple levels of acceptability in the context of full multiobjective optimization to calibrate several models; they demonstrated the superiority of this approach over other approaches that are based on optimizing residual-based measures [29].

Streamflow records of several years are necessary for calibrating hydrological models [30], making the calibration of hydrological modeling of poorly gauged catchments or that in situations of considerable data gaps challenging. Several studies have investigated the possibility of using both limited continuous and discontinuous periods of streamflow observations to calibrate hydrological models [31–43]. Tada and Beven (2012) proposed an effective method to extract information from short observation periods in three Japanese basins. They examined calibration periods spanning 4–512 days; they randomly selecting their starting day and reported varying performances, concluding the challenge of pre-identifying the performing short periods in ungauged basins [37]. Sun et al. (2017) obtained performances similar to that of the full-length dataset model when calibrating a physically based distributed model with limited continuous daily streamflow data records (less than one year) in data-sparse basins [32].

Previous studies have explored the discontinuous streamflow data on two bases that can be classified into two categories suggested by Reynolds et al. (2020). One is on the basis that the available data are limited only to separated spots of discharge and the second that the continuous records of discharge are available but only for a few events [39]. In the context of the first category, Perrin et al. (2007) achieved robust parameter values for two rainfall-runoff models by random sampling of 350 discontinuous calibration days, including dry and wet conditions, in diverse climatic and hydrologic basins in the USA [38]. They concluded that in the driest catchments, stable parameter values are harder to achieve. Pool et al. (2017) investigated an optimal strategy for sampling runoff to constrain a rainfall-runoff model using only 12 daily runoff measurements in

one year in 12 basins in temperate and snow-covered regions throughout the east of the USA. They found that the sampling strategies comprising high-flow magnitude result in better hydrograph simulations, whereas strategies comprising low-flow magnitude result in better flow duration curve (FDC) simulations [43]. In the context of the second category, Seibert and McDonnell (2015) investigated the significance of limited streamflow observations and soft data in the Maimai basin (New Zealand) [40]. They found that 10 records of discharge data sampled from high flows used to inform the calibration of a simple rainfall-runoff model result in similar results when using three months of continuous discharge data for calibration. Reynolds et al. (2020) examined the hypothesis that limited flood-event hydrographs in a humid basin are adequate to calibrate a rainfall-runoff model. Their results indicate that two to four calibration events can substantially improve flood predictions for accuracy and uncertainty reduction; however, adding more events resulted in limited performance improvements [39].

In the context of signature satisfaction and absence of time-series, Gharari et al. (2014) proposed an alternative approach for parameter identification was based on prior experience (or professional knowledge) constraints. The algorithm of parameter searching was based on random search (stepwise Monte Carlo sampling) under these constraints. The parameter sets selected using this approach led to a consistent model, along with the potential to reproduce the functional behavior of the catchment [44]. In our study, we evaluate the hypothesis that using several hydrological signatures as objective functions will directly improve the calibration process in the situation of limited data.

To conclude, much of the previous research has focused on identifying and evaluating the minimum requirements of data for model calibration regarding quantity and methods to find the most informative sections of hydrographs and exploring approaches for optimal sampling strategies. This study predominantly focuses on the development and evaluation of a signature-based model calibration approach that incorporates several hydrological signatures to guide the parameter search toward regions of hydrological consistency in the search space under various data-availability scenarios. Different setups of the multiobjective signature-based (MO-SB) optimization approach are compared to single-objective (SO) calibration using traditional error metrics. The focus is primarily on cases where only continuous short-period observations were available. This study would provide a practical solution to successfully calibrate conceptual rainfall-runoff models in terms of both performance and consistency for poorly gauged catchments.

2. Study Area and Datasets

Selection of the case study was driven by the necessity of having enough observational data to conduct experiments with progressive reduction of data availability. The Brue catchment in the UK was chosen. It covers an area of 135 km² in the southwest of England, starting from Brewham and ending at Burnham-on-sea, and the outlet is at Lovington. This catchment is covered by three weather radars and a densely distributed rain gauge station. Researchers have comprehensively studied the area for modeling rainfall runoff and for precipitation forecasting, especially during the hydrological radar experiment [27,45–47]. The catchment is predominantly characterized by some low hills, discontinuous groups of rocks under clay soils, and major grasslands. Figure 1 shows the topography of the catchment and outlet's location.

Hourly precipitation data and discharge data from the Lovington gauging station are used in this study. Data were obtained from radar and rainfall gauges with a resolution of 15 min; in addition, data of potential evapotranspiration that was computed using a modified Penman method recommended by the Food and Agricultural Organization were obtained using automatic weather station data (temperature, solar radiation, humidity, wind speed) [48].

Figure 1. Brue catchment.

Three years and four months of hourly data from 1 September 1993 to 31 December 1996 were selected as the full dataset (FD) to calibrate the model, and one year and almost one month of data from 1 June 1997 to 3 June 1998 was selected as the validation dataset [45].

3. Methodology

The procedure for a signature-based model calibration followed sequential steps (Figure 2). The following subsections provide details on the implemented procedure.

Figure 2. Steps followed in the methodology.

3.1. Selection of Hydrological Signatures

The literature uses numerous hydrological signatures, either in hydrological model evaluation and calibration or catchment classification [20,29,49]. This study follows the guidelines or criteria for signature selection suggested by Mcmillan et al., (2017) [50]. The signatures were derived from the available time-series data as the basis of analysis. The selection process yielded 11 hydrological signatures listed in Table 1: three signatures extracted from three segments of the FDC, four signatures related to streamflow and precipitation, and four signatures characterizing the discharge statistics. The selected signatures have a distinct link to the hydrological process, leading to a better interpretation of the catchment's functional behavior. Moreover, their scales do not depend on the catchment size as they represent different parts of the flow hydrograph.

Table 1. Summary of hydrological and statistical signatures used in the study.

Symbol	Hydrological Signature	Equation	Comments	References
FHV (FDC)	High-flow segment volume of the flow duration curve	$\sum_{h=1}^H Q_h$	$h = 1, 2, 3, \dots, H$ are the indices of high flows; their probability of exceedance is < 0.02	[51]
FLV (FDC)	Low-flow segment volume of the flow duration curve	$-1 \times \sum_{l=1}^L [\log Q_l - \log Q_L]$	$l = 1, 2, \dots, L$ are the indices of low flows; their probability of exceedance is between 0.7 and 1.0 (L is the minimum flow index)	[51]
FMS (FDC)	Medium-flow segment of the flow duration curve	$\log Q_{m1} - \log Q_{m2}$	$m1$ and $m2$ are the lowest and highest flow exceedance probabilities within the mid-segment of FDC (0.2 and 0.7, respectively, in this study)	[51]
I_{BF}	Baseflow index	$Q_{Dt} = C Q_{Dt-1} + \frac{1+C}{2} (Q_t - Q_{t-1})$ $Q_{Bt} = Q_t - Q_{Dt}$ $I_{BF} = \sum_{t=1}^N \frac{Q_{Bt}}{Q_t}$	Q_{Dt} is the filtered surface runoff at t time-step, Q_t is the total flow (original streamflow) at t time-step, Q_{Bt} is the baseflow at t time-step, C is the filter parameter (0.925), I_{BF} is the baseflow index, and N is the total time steps of the study period	[20,52]
R_{QP}	Runoff ratio	$R_{QP} = \frac{Q}{P}$	R_{QP} is the runoff ratio, Q is the long-term average streamflow, and P is the long-term precipitation	[20,49,51]
R_{LD}	Rising limb density	$R_{LD} = \frac{N_{RL}}{T_R}$	N_{RL} is rising limbs (number of peaks of the hydrograph) and T_R is the total time that the hydrograph is rising	[13,20,49,53]
E_{QP}	Stream flow elasticity	$E_{QP} = \frac{dQ/Q}{dP/P} = \frac{dQ}{dP} \frac{P}{Q}$ $E_{QP} = \text{median} \left(\frac{Q_t - \bar{Q}}{P_t - \bar{P}} \frac{\bar{Q}}{\bar{P}} \right)$	dQ/Q is the proportional change in the streamflow, dP/P is the proportional change in precipitation, Q_t and P_t are streamflow and precipitation, respectively, at t time-step, and $\bar{Q}$ and $\bar{P}$ are the mean of streamflow and precipitation, respectively, in the long-term	[20,54]
Q_{mean}	Mean discharge	$\sum_{t=1}^N Q_t / N$	Q_t is the streamflow at t time-step and N is total time steps of the study period	[29,55,56]
Q_{median}	Median discharge	$M(Q)$		[29]
$DV(Q)$	Discharge variance	$\sqrt{\sum_{t=1}^N (Q_t - \bar{Q})^2 / (N - 1)}$	Q_t is the streamflow at t time-step, $\bar{Q}$ is the mean of the streamflow, and N is the total time steps of the study period	[29]
Q_{peak}	Peak discharge	$P(Q)$	$P(Q)$ is the peak of the streamflow data	[29]

3.2. Data Setup

To meet the study objective, several scenarios were created to obtain different dataset sizes representing various levels of information deficiency. The following steps were followed to set up the data.

1. Select a long-period dataset as an FD for model calibration (benchmark dataset) (Table 2);
2. Select an additional dataset for model validation (Table 2);
3. Divide the FD into partial datasets progressively decreasing in size, from long-term- to short-term data using four scenarios (Table 2):
  1. Scenario 1: Each new data subset is composed by removing a certain amount of data (a certain percentage, e.g., remove 25% of the total data) from the end of the FD (Figure 3);
  2. Scenario 2: The new data subset is created by removing an equal amount of data from both the start and end of the FD (Figure 3);
  3. Scenario 3: A section of the FD represents a short continuous dry period (no precipitation);
  4. Scenario 4: A section of the FD represents a short continuous wet period (frequent and intensive precipitation).

Table 2. Datasets used in the study.

Dataset	Date (from–to)	Number of Data Records
FD	1 September 1993 00:00–31 December 1996 23:00	29,232
Validation dataset	1 June 1997 01:00–30 June 1998 23:00	9478
Scenario 1		
75% FRD	1 September 1993 00:00–2 March 1996 11:00	21,924
50% FRD	1 September 1993 00:00–2 May 1995 22:00	14,615
25% FRD	1 September 1993 00:00–2 July 1994 11:00	7308
10% FRD	1 September 1993 00:00–31 December 1993 23:00	2928
5% FRD	1 September 1993 00:00–31 October 1993 21:00	1462
Scenario 2		
75% FRD	31 January 1994 05:00–1 August 1996 17:00	21,924
50% FRD	2 July 1994 10:00–2 March 1996 11:00	14,615
25% FRD	1 December 1994 15:00–2 October 1995 05:00	7308
10% FRD	2 March 1995 22:00–2 July 1995 21:00	2928
5% FRD	2 April 1995 08:00–2 June 1995 10:00	1462
Scenario 3		
Dry-period dataset	1 June 1994 00:00–1 August 1994 23:00	1488
Scenario 4		
Wet-period dataset	27 December 1994 10:00–1 February 1995 13:00	868

Figure 3. Example of two scenarios to obtain 75% Fraction Retained Dataset (FRD) of the full dataset (FD): scenario 1 (left) and scenario 2 (right).

3.3. HBV Model Setup

HBV is a conceptual model that can simulate runoff in different climate zones using precipitation, temperature, and potential evapotranspiration as inputs; it was developed by the Swedish Meteorological and Hydrological Institute and has been applied in more than 30 countries [57,58]. Various variants of the model have been suggested, e.g., HBV-Light by Seibert (1997) [59] and HBV-96 by Lindström et al. (1997) [60]. The model comprises various routines, namely, precipitation, snow, soil, response, and routing routines. Table 3 presents the HBV parameters that are calibrated herein using the following methodology.

Table 3. Parameters of the HBV model targeted for calibration in this study.

Parameter	Explanation	Unit
Precipitation Routine		
LTT	Lower temperature threshold	°C
UTT	Upper temperature threshold	°C
RFCF	Rainfall corrector factor	—
SFCF	Snowfall corrector factor	—
Snow Routine		
CFMAX	Day degree factor	mm °C ⁻¹ h ⁻¹
TTM	The temperature threshold for melting	°C
CFR	Refreezing factor	—
CWH	Water holding capacity	—
Soil and Evaporation Routine		
FC	Maximum soil moisture	Mm
ETF	Total potential evapotranspiration	mm h ⁻¹
LP	Soil moisture threshold for evaporation reduction (wilting point)	—
E_CORR	Evapotranspiration corrector factor	—
BETA	Shape coefficient	—
C_FLUX	Capillary flux in the root zone	mm h ⁻¹
Response Routine		
K	Upper zone recession coefficient	h ⁻¹
K1	Lower zone recession coefficient	h ⁻¹
PERC	Maximum percolation rate from the upper to the lower tank	mm h ⁻¹
ALPHA	Response box parameter	—
Routing Routine		
MAXBAS	Routing, length of weighting function	H

3.4. Model Calibration Approaches

The calibration comprises the single objective (SO) optimization and Multi-objective signature-based (MO-SB) optimization approaches.

3.4.1. Formulation of SO Optimization Approach

In the SO approach, the constrained SO optimization algorithm is used to maximize Nash–Sutcliffe efficiency (NSE), equivalent to minimizing the mean-squared error divided by observation variance, as a goodness-of-fit measure. Nineteen parameters of the HBV model are the decision variables of the optimization problem. The upper and lower boundaries are the constraints. The calibration approach is first used to calibrate the benchmark model (using the full dataset FD) and then implemented for the datasets in the four scenarios (calibration of 12 datasets). The initial states of the models differ from one model to another, making it necessary to obtain the initial states of each model (can be done by randomized search or simply by trial and error), at the beginning of the modeling process.

The Augmented Lagrangian Harmony Search Optimizer (ALHSO) algorithm (belonging to the class of randomized search algorithms) from the pyOpt Python library was used

to solve the optimization problem. It has been applied in complex and continuous problems, such as calibrating hydrologic models, and is efficient [61]. The ALHSO algorithm is suitable for solving an SO optimization problem and has fewer control parameters without the need to set the initial values of decision variables [10,62].

3.4.2. Formulation of MO-SB Optimization Approach

In this approach, the calibration problem is solved by evaluating the extent of signature achievement in the parameter search process executed using an optimization algorithm. Signature achievement is measured by computing the signature score function, which compares the simulated and measured signatures. The multiobjective optimization problem was solved by maximizing 16 objective functions, including 15 individual hydrological signature score functions, with a certain level of acceptability (threshold), and the NSE. Decision variables and constraints are the same as in the first approach (SO). The initial model states are known in this phase from the first experiment and can be reused. The nondominated sorting genetic algorithm (NSGA)-II algorithm [63] from the Inspired Python library was used to solve the optimization problem. NSGA-II is a multiobjective evolutionary algorithm belonging to the class of randomized search methods. NSGA-II has advantageous features compared to other multiobjective constrained optimizers regarding convergence to Pareto optimal solutions and ensuring their good spread in decision spaces and performs well in constrained problems [64].

The observed and simulated hydrological signatures were calculated for observations and model simulations. The signature deviations (Dev) between them were calculated individually (Equation (1)), consistent with past studies [29,51,65], and transferred to scores (normalizing value) using binary functions (Equation (3)). The idea of the binary score function is based on defining thresholds (+ or −) for the acceptable values of signatures (Equation (2)), meaning that if the value of deviation is within the limits, the score will equal 1 and, if not, the score will equal 0, as implemented by [29,66].

$$Dev = \frac{Signature_{Observed} - Signature_{simulated}}{Signature_{Observed}} \quad (1)$$

$$\pm Dev_{threshold} = \pm \frac{Acceptability\ threshold * Signature_{Observed}}{Signature_{Observed}} \quad (2)$$

$$Score = \begin{cases} 1, & |Dev| \leq |\pm Dev_{threshold}| \\ 0, & |Dev| > |\pm Dev_{threshold}| \end{cases} \quad (3)$$

In this study, two acceptability thresholds were used (10% and 20% deviation), similar to previous research [29]. In addition, to explore the algorithm convergence (speed) and diversity in Pareto optimal sets, two crossover types were implemented in setting the NSGA-II (the blend crossover (BC) and uniform crossover (UC)). In total, four parameterization algorithms were formulated and coded in Python:

- 10% acceptability threshold and BC, MO-BC (10%);
- 10% acceptability threshold and UC, MO-UC (10%);
- 20% acceptability threshold and BC, MO-BC (20%);
- 20% acceptability threshold and UC, MO-UC (20%);

3.5. The Diagnostic Model Evaluation Approach

The diagnostic model evaluation approach is based on validating (testing) both model performance and consistency. The performance was evaluated by calculating the performance measures for each model (Table 4). Consistency was evaluated by calculating the difference between the simulated hydrological signatures and those calculated from observed measurements (error). In the MO-SB calibration approach, the solution is the Pareto set that contains a large number of solutions (100 solutions), making it difficult to evaluate them all. We propose herein the idea of choosing and further evaluating a single

best solution using a single aggregated criterion (score) after exploring the composition of the optimal Pareto set. We adopted the method of ideal point, i.e., choosing a solution closest to the ideal point (for the considered problem, it is the point where all objective functions have a value of 1). In this study, NSE was used without normalization because for the considered models, it was always between 0 and 1. Minimizing the distance to the ideal point is equivalent to maximizing the distance to 0; thus, the aggregated score can be written as:

$$AggregatedScore = \sqrt{NSE^2 + \sum_{i=1}^N Score(Signature)_i^2} \quad (4)$$

Table 4. Performance measures matrix.

Symbol	Description	Formula	Optimal Value	References
NSE	Nash–Sutcliffe efficiency	$NSE = 1 - \frac{\sum_{i=1}^n (Y_i^{obs} - Y_i^{sim})^2}{\sum_{i=1}^n (Y_i^{obs} - Y_{mean})^2}$	1	[17,49,67,68]
RMSE	Root mean square error	$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_{obs}^t - Y_{sim}^t)^2}$	0	[15,49,69,70]
PBIAS	Percent bias (relative volume error)	$PBIAS = \frac{\sum_{i=1}^n (Y_i^{obs} - Y_i^{sim})}{\sum_{i=1}^n Y_i^{obs}} \times 100$	0	[17,67,71]

4. Results

One validation dataset (Table 2) with hourly data spanning one year (1 June 1997 01:00–30 June 1998 23:00) was used in all experiments, whereas the FD was used for calibration. Calibration was run under four scenarios. Scenarios 1 and 2 have the same number of partial datasets with different combinations, whereas scenarios 3 and 4 are limited to dry and wet periods, respectively (Table 2). Figure 4 shows the number of records of partial datasets in the four scenarios models.

Figure 4. Record numbers per (partial) dataset.

4.1. Diagnostic Evaluation of the SO Optimization Approach

4.1.1. General Characterization of Results

Although the evaluation criteria in this study were based on performance and consistency, it is worthwhile to visually inspect the simulated flow hydrographs from the calibrated models to provide an overall idea of models' ability to estimate the observed flow (peaks, low values). Figures 5–7 show the simulated hydrographs of the FD model,

50%-FRD model (scenario 2), and 5%-FRD model (scenario 2), respectively. Figures 8 and 9 show the simulated hydrographs for the dry- and wet-period models, respectively.

Figure 5. (a) Validation of the calibrated FD model using the single-objective (SO) approach; (b) zoomed in for a clear view.

Figure 6. (a) Validation of the calibrated 50%-FRD model; (b) zoomed in for a clear view.

Figure 7. (a) Validation of the calibrated 5%-FRD model; (b) zoomed in for a clear view.

Figure 8. (a) Validation of the calibrated dry-period model; (b) zoomed in for a clear view.

Figure 9. (a) Validation of the calibrated wet-period model; (b) zoomed in for a clear view.

The simulation graphs show that the FD and 50%-FRD models show relatively good results (Figures 5 and 6); however, none of them captured any peaks. For instance, the maximum observed peak flow was $31 \text{ m}^3/\text{h}$, whereas the simulated flow of the FD and 50%-FRD models were 17.8 and $21.5 \text{ m}^3/\text{h}$, respectively, which were much lower than the observed one was. Based on the visual comparison, it is difficult to decide which model shows better results; thus, the performance metrics and signatures must be evaluated. The 5%-FRD and dry-period models (Figures 7 and 8, respectively) show poor results, representing a typical case with short-period data that do not hold enough information that could help simulate the streamflow. However, the wet-period dataset model (Figure 9) showed reasonable results with a general flow overestimation of $19.52 \text{ m}^3/\text{h}$ for the $2.6 \text{ m}^3/\text{h}$ observed flow. Some peaks were overestimated ($43.7 \text{ m}^3/\text{h}$ vs. $22.3 \text{ m}^3/\text{h}$ observed), whereas other peaks were underestimated ($7.2 \text{ m}^3/\text{h}$ vs. $30.1 \text{ m}^3/\text{h}$ observed).

4.1.2. General Characterization of Results

All models in scenario 1 showed similar NSE values in the calibration period, ranging between 0.87 and 0.96 , with the superiority of the 5%-FRD model. However, in the validation period, the 5%-FRD model showed poor performance with an NSE of approximately zero whereas the rest of the models showed good NSE values with an average deviation of 0.1 from the NSE in the calibration period (Table 5). Root mean square error (RMSE) values in the calibration and validation periods were small, ranging between 0.96 and 1.7 mm , except for the 5%-FRD model showing a 5.65-mm RMSE in the validation period (Table 5). All PBIAS values in the calibration period were positive, whereas the validation period exhibited negative values from three models (25% FRD, 10% FRD, and 5% FRD). Negative PBIAS values indicate an underestimation in the flow simulation. The PBIAS of the 25%-FRD and 10%-FRD models were acceptable but that of the 5%-FRD model was high (-106.47), which is unacceptable (Table 5), indicating that this model was built using data that were insufficient to simulate the flow.

Table 5. Results of performance measures for the four scenarios.

Dataset	NSE		RMSE		PBIAS	
	Calibration	Validation	Calibration	Validation	Calibration	Validation
Scenario 1						
FD	0.87	0.77	1.14	1.7	9.16	10.2
75% FRD	0.87	0.77	1.25	1.6	8.9	10.73
50% FRD	0.89	0.81	1.31	1.52	8.42	9.61
25% FRD	0.89	0.85	1.28	1.5	8.81	−3.16
10% FRD	0.94	0.82	1.14	1.51	10.03	−7.78
5% FRD	0.96	0.01	0.96	5.65	1.2	−106.47
Scenario 2						
FD	0.87	0.7	1.12	1.35	7.5	11.2
75% FRD	0.87	0.72	1.09	1.3	7.33	10.24
50% FRD	0.9	0.69	1.02	1.8	7.63	15.98
25% FRD	0.84	0.78	1.3	1.34	15.44	5.96
10% FRD	0.86	0.69	0.41	1.6	6.92	−22.64
5% FRD	0.52	0.1	0.25	2.8	−0.24	−37.1
Scenario 3						
Dry-period dataset	0.72	0.18	0.1	2.65	6.58	−36
Scenario 4						
Wet-period dataset	0.86	0.57	2	1.75	6.19	25.6

Similarly, in scenario 2 (Table 5), the 5%-FRD model showed poor NSE in the calibration and validation periods (0.52 and 0.1, respectively). NSE fluctuated without a clear pattern in the calibration and validation periods. Overall, the NSE values of scenario 2 were lower than those of scenario 1, ranging between 0.69 and 0.78 (excluding the 5%-FRD model) in the validation period. The 5%-FRD model showed the lowest performance in terms of the RMSE (2.8 mm), whereas the rest of the models showed acceptable RMSE values, ranging between 0.41 and 1.8 mm in the calibration and validation periods. The 75%-FRD and 25%-FRD models had similar RMSEs (with an average of 0.3) in the calibration and validation periods, whereas the RMSE of the 50%-FRD and 10%-FRD models increased slightly in the validation period, reaching 1.6 and 1.8 mm, respectively. According to the PBIAS values, the minimum limit to acquire acceptable performance was 25% FRD as models of the short-term data (10%-FRD and 5%-FRD models) resulted in high underestimations, as indicated by the PBIAS values (−22.64 and −37.1).

Scenarios 3 and 4 (Table 5) represent short-term data for the dry- and wet-period models. The wet-period model performed better than the dry-period model in terms of the NSE, RMSE, and PBIAS. Specifically, the wet-period model showed higher NSE in the validation period (0.57) than the dry-period model (0.18), whereas the RMSE of the wet-period model was 1 mm less than that of the dry-period model (2.65 mm). Both models were inaccurate in the validation period, with either high overestimation (wet-period model) or high underestimation (dry-period model), as indicated by the PBIAS values for both scenarios. The consistency evaluation of the SO calibrated models is discussed in Section 4.2.2, with the MO-SB calibrated models to compare them.

4.2. Diagnostic Evaluation of the MO-SB Optimization Approach

In this section, a comparison between the four multiobjective optimization algorithms (MO-BC (10%), MO-BC (20%), MO-UC (10%), and MO-UC (20%)) and the SO optimization is provided. First, the performances of the models were evaluated; then, the consistency

of the models was evaluated by comparing the difference between the observed and simulated values of each signature. The results presented in this section focus on scenarios 2, 3 (dry-period data), and 4 (wet-period data) of the dataset sectioning (Table 2). Scenario 1 is not presented because its results pertain to the simple case of a gradually decreasing dataset size.

4.2.1. Performance Evaluation of MO-SB

The evaluation was based on the closest solution to the ideal point in the Pareto set, which has a maximum aggregated score according to Equation (4). Figure 9 shows the NSE values of the five models in the validation period. The MO-BC (20%) algorithm parameterization gave the highest NSE in all models. The NSE obtained from the other three algorithm parameterizations varied from one model to another, but in most cases, the MO-BC led to higher NSE than the MO-UC algorithm parameterization did. The NSE, 5%-FRD, and dry-period models resulted in low NSE, indicating poorer performance than the rest of the models; however, the wet-period model showed an acceptable NSE (average: 0.564). The RMSE values ranged between 1.18 and 2.8 mm for all models, which is relatively low. MO-BC (20%) yielded the lowest RMSEs in all models, with different dataset sizes (Figure 10). Moreover, the highest RMSEs were observed for the 5%-FRD and dry-period models (Figure 11). A noticeable decrease in the PBIAS value occurred in all models after implementing MO-SB (Figure 12) with the different parameterization algorithms. The poor performance was also observed for models of short-term data (5%-FRD, dry-period, and wet-period models). The wet-period model showed negative PBIAS as the streamflow values were overestimated because the same dataset was used as the validation dataset for all models instead of using a different dataset for the wet-period model.

Figure 10. NSE values of different models using five datasets via two main calibration approaches (SO & 4 algorithms parameterization of MO-SB) for each dataset.

Figure 11. RMSE values of different models using five datasets via two main calibration approaches (SO & 4 algorithms parameterization of MO-SB) for each dataset.

Figure 12. PBIAS values of different models using five datasets via two main calibration approaches (SO & 4 algorithms parameterization of MO-SB) for each dataset.

4.2.2. Behavioral Consistency Evaluation

The differences between the observed and simulated signatures were calculated for each signature to evaluate the consistency of the output from the optimized calibrated models and their ability to simulate the catchment's behavior. This section presents the results of the consistency evaluation for each signature.

Baseflow index (I_{BF}): The observations and simulation results revealed high I_{BF} values (0.84–0.98), indicating a high baseflow in the catchment and a high groundwater contribution. I_{BF} values for the wet-period model were lower than those for others, confirming that wet-period data contain high flows, therefore having more direct streamflow and consequently, less baseflow than in other datasets. Furthermore, using the MO-SB approach with different parameterization algorithms did not improve the results significantly; however, MO-BC (20%) yielded the lowest errors for all models (Figure 13). The simulated I_{BF} values for all models were close to the observed values, meaning that all models were consistent with the baseflow index.

Figure 13. I_{BF} errors based on observations and simulations of five datasets using two main calibration approaches (SO & 4 algorithms parameterization of MO-SB).

Streamflow elasticity (E_{QP}): The value of the E_{QP} calculated from observations was high (127.7), indicating that the streamflow is sensitive to precipitation and that the catchment is elastic. The results obtained from simulated flow after implementing different calibration algorithm parameterizations dramatically varied for each model, indicating signature sensitivity to the length of the records and information held by the data. The 25%-FRD model was the most accurate at reflecting the streamflow's elasticity. The performances of SO and MO-SB were similar, with errors ranging between −2.7 and −11.3. The 5%-FRD and dry-period models showed small E_{QP} values, resulting in high errors (Figure 14). Wet-period model simulations also resulted in larger errors than other models but in the

opposite direction. MO-BC (20%) is the best calibration parameterization approach as it enhanced the results in all models, especially in the 10%-FRD and wet-period models.

Figure 14. EQP errors based on observations and simulations of five datasets using two main calibration approaches (SO & 4 algorithms parameterization of MO-SB).

Rising limb density (R_{LD}): The values of the observed and simulated R_{LD} were small (0.02–0.04), indicating the smoothness of the flow hydrograph. The results after implementing the MO-SB algorithms were similar to those obtained using the SO approach; however, MO-BC (20%) reduced the errors marginally in the FD, 75%-FRD, 50%-FRD, 25%-FRD, and wet-period models (Figure 15).

Figure 15. R_{LD} errors based on observations and simulations of five datasets using two main calibration approaches (SO & 4 algorithms parameterization of MO-SB).

Runoff ratio (R_{QP}): The R_{QP} values were high for all models (10.9–21.6), indicating the domination of blue water in the catchment. Therefore, the streamflow is larger than evapotranspiration in the context of water balance if we assume no change in the storage of the catchment. The dry-period model showed the lowest simulated R_{QP} and consequently, the highest errors among the other models (Figure 16), whereas the wet-period model showed a high R_{QP} but in the opposite direction. The results confirm that data containing frequent and high events will result in a large R_{QP} and vice versa. The 5%-FRD and 10%-FRD models also resulted in low R_{QP} . However, using MO-BC (20%) lowered the errors significantly from 6.5 and 5.3 (obtained via SO) to 1.5 and 0.4, respectively, which are acceptable values compared to the errors of the rest of the models. MO-BC (10%) enhanced the results for R_{QP} of the 5%-FRD and 10%-FRD models and ranked second after MO-BC (20%). Furthermore, the MO-BC (20%) significantly improved the R_{QP} values in all models.

Figure 16. RQP errors based on observations and simulations of five datasets using two main calibration approaches (SO & 4 algorithms parameterization of MO-SB).

High-flow segment volume of the FDC (FHV (FDC)): The observed FHV (FDC) was high (2365.1), indicating that the catchment faces frequent flooding because of high stream-flow. Simulated FHV (FDC) using long-period models (FD, 75% FRD, and 50% FRD) showed lower values than the observed FHV (FDC), although they were still acceptable. However, short-period models failed to reflect this signature as their simulated values were too far from the observed values—either very small values, as simulated using the 5%-FRD and dry-period models, or very high values, as simulated using the wet-period model. The results confirm that short-term and dry-period data lack high-flow events, underestimating the volume of the very high flows, with the opposite being true for the wet-period models. Although the errors were high for short-period models, MO-BC (20%) was the best approach according to simulations of this signature (Figure 17).

Figure 17. FHV (FDC) errors based on observations and simulations of five datasets using two main calibration approaches (SO & 4 algorithms parameterization of MO-SB).

Low-flow segment volume of the FDC (FLV (FDC)): From the high errors (overestimation) in Figure 18, we found that the FLV (FDC) volume cannot be simulated by any of the models; all models failed to reflect the FLV (FDC) in the Brue catchment. However, the wet-period models yielded the lowest errors. MO-BC (20%) reduced the errors significantly in the FD, 75%-FRD, and 50%-FRD models. Nevertheless, according to this signature, no model was consistent.

Figure 18. FLV (FDC) errors based on observations and simulations of five datasets using two main calibration approaches (SO & 4 algorithms parameterization of MO-SB).

Mid-flow segment slope of the FDC (FMS (FDC)): The FMS (FDC) was well simulated (Figure 19). The errors ranged between -0.2 and 0.5 , with the 5%-FRD model exhibiting the best performance with an error of 0.1 for all calibration approaches. The observed FDC slope was steep (0.8), indicating flashy runoff, indicating that moderate flows do not remain in the catchment. Signature-based calibration improved the results slightly. Overall, all models were consistent for this signature.

Figure 19. FMS (FDC) errors based on observations and simulations of five datasets using two main calibration approaches (SO & 4 algorithms parameterization of MO-SB).

Mean discharge (Q_{mean}): The observed and simulated values of mean streamflow resulting from different calibration approaches ranged between 1.2 and 2.5 , whereas the observed mean was 1.9 . The range of the simulated mean flows was small and acceptable. However, most simulated mean flows were lower than the observed mean flow for all models, except the wet-period model because of a lack of low flows in the wet-period data. The MO-SB algorithms provided more consistent models than the SO approach. The MO-BC (20%) enhanced the results, particularly in the FD, 75%-FRD, 50%-FRD, and 25%-FRD models (errors were almost zero; see Figure 20).

Figure 20. Mean streamflow errors based on observations and simulations of five datasets using two main calibration approaches (SO & 4 algorithms parameterization of MO-SB).

Median discharge (Q_{median}): The results of the simulated median streamflow resulting from the long and short-term period models using different calibration approaches converged with those of the observed median flow of the catchment (0.77–1.4) and the observed value was 1. The wet-period model was the most consistent according to the Q_{median} as it showed zero errors for all parameterization approaches (Figure 21). Overall, the MO-SB algorithms improved the results for most models, except the 5%-FRD, dry-period, and wet-period models as the SO approach resulted in smaller errors than the MO-SB optimization algorithms.

Figure 21. Median discharge errors based on observations and simulations of five datasets using two main calibration approaches (SO & 4 algorithms parameterization of MO-SB).

Discharge variance ($DV(Q)$): The observed discharge variance was high, indicating varying streamflow. The simulated $DV(Q)$ was close to the observed value, except for the 10%-FRD, 5%-FRD, and dry-period models, where the differences were slightly higher than for other models (Figure 22). The 5%-FRD and dry-period models indicated no variability in streamflow as their variances were small because their data were in the same range and close to the mean.

Figure 22. Discharge variance errors based on observations and simulations of five datasets using two main calibration approaches (SO & 4 algorithms parameterization of MO-SB).

Peak discharge (Q_{PEAK}): This signature was associated with the maximum peak observed in the catchment. The results presented in Figure 23 show a significant decrease in the simulated peaks of the 5%-FRD and dry-period models, whereas there was a significant increase in the simulated peaks of the wet-period model. The 25%-FRD model was the most consistent in simulating the peak discharge. The simulated peaks obtained using the SO and MO-SB algorithms were the same. Overall, the MO-SB improved the results slightly, especially in the 10%-FRD model.

Figure 23. Peak discharge errors based on observations and simulations of five datasets using two main calibration approaches (SO & 4 algorithms parameterization of MO-SB).

5. Discussion

Overall, the results showed that the HBV model was successfully calibrated using SO and MO-SB optimization approaches using short-term datasets with a lower limit of approximately four months of data (10%-FRD model). These results correlate with those obtained by Brath et al. (2004) and Perrin et al. (2007), indicating that calibrated models can generate reasonable results using data for less than one year [38,41].

It is difficult to compare the results of this study to previous investigations of model performance under continuous short-term data because no previous study incorporated hydrological signatures in the process of parameters' search using short-term data. Thus, the comparison is restricted to the general results of the models' performance regardless of the calibration method. According to the performance measurements, the performance of MO-SB was the best for MO-BC (20%) whereas both MO-UC (10% and 20%) optimization algorithms showed the lowest performance (less than the SO approach), indicating the ineffectiveness of using this setting (uniform crossover). This finding contradicts that of a study by Shafii and Tolson (2015), who concluded that the model performance depends

more on the formulation of the optimization problem than on the choice of the optimization algorithm [29].

In terms of the impact of dataset size on performance, the FD, 75%-FRD, and 50%-FRD models exhibited similar performance. The performance of the 10%-FRD model was lower but still acceptable. The obtained results are consistent with those presented in a study by Brath et al. (2004), who proved that data of less than one year can contribute to acceptable model performances [41]. The 5%-FRD and dry-period models exhibited the lowest performance under all calibration approaches, indicating that in both cases, short hourly records and scarcity of events make the dataset insufficient to build a model that provides acceptable performance, even when using signature-based calibration. The results of the dry-period model correlate with previous findings, such as those of Li et al. (2010) who showed that dry catchments require a longer data-collection period for calibration to obtain stable parameters [36]. Perrin et al. (2007) concluded the difficulty of estimating robust model parameters in dry catchments and recommended longer periods of calibration to achieve stable parameters [38]. Pool et al. (2017) found that dry runoff periods, defined by mean and minimal flow samples, convey less information for hydrograph prediction than wet periods [43]. The situation is different when limited data, as in this study, are selected from a wet period because the availability of multiple events results in good performance. This finding is consistent with previous research, confirming that data containing sufficient high flows leads to better calibration and improved model performance [31,32,38].

Using the signatures in model evaluation allowed for developing a better understanding of the catchment processes. For example, the results provided insight into the catchment's baseflow. The baseflow index results show a high baseflow in the catchment, correlating with results from a study on the same basin [47]. Additionally, E_{QP} values demonstrated the streamflow sensitivity to the observed precipitation. The R_{LD} values indicate the smoothness of the flow hydrograph. However, the R_{QP} values indicate the domination of blue water, meaning that the streamflow is larger than evapotranspiration in the context of water balance if we assume no change in the storage of the catchment. The volume of the high segment of the FDC indicates a frequent flood in the catchment because of the high streamflow, which is consistent with previous investigation of the Brue catchment [72]. The FDC slope was steep, indicating a flashy runoff; therefore, moderate flows do not remain in the catchment. Also, note that all models (including the long-term data models) failed to reflect the FLV (FDC) in the Brue catchment with significant errors. This result could be interpreted by the effects of vegetation growth in the Brue catchment, leading to poor simulations of low flows, as reported in previous studies [47]. Overall, the results show that five of the hydrological signatures were sensitive to dataset size, namely, the elasticity of streamflow, FHV (FDC), FLV (FDC), R_{QP} , and the peak of the flow.

Finally, the study provides a quantitative estimation of the impact of dataset size on model performance and consistency for several metrics and signatures. Incorporating signatures in the model calibration process produced a consistent model with higher performance, improving the results in the case of limited data compared to only using goodness-of-fit measures. However, there is a lower limit on the length of the observation records to build a successful signature-based calibrated model. This is a matter of both dataset length and the hydrological information contained within these limited observations. Note that the signature-based diagnostic evaluation approach enables the selection of consistent models reflecting critical characteristics of the catchment behavior and allows for identifying the deficiencies of various models.

6. Conclusions

In this study, the usefulness of incorporating hydrological signatures in calibrating a rainfall-runoff model (HBV model) under different data-availability scenarios was assessed. In contrast to previous studies, the effect of dataset size on both performance and consistency of HBV calibrated models was investigated in this study. The results showed that a limited number of records can ensure a reasonably good performance in

calibration because the modeled hydrograph can fit the observations to an extent sufficient for hydrological practice. Contrarily, for the validation period, using limited data resulted in poor performance and consistency, as illustrated by the 5%-FRD model in both scenarios 1 and 2.

Nevertheless, the progressive reduction in dataset size deteriorates the model's performance and the dry periods do not allow for feeding the models with enough information. Therefore, model performance depends on whether the data contains enough climatic information and some extreme events that could help make the model more representative of the watershed; however, more accurate quantification of such influence would require more studies. The MO-SB calibration improved the model results better than the SO calibration approach. The diagnostic evaluation approach provides a powerful and meaningful basis for interpreting model results. However, for the considered case study, the improvement was not as much as expected, and more experiments are needed with various types of catchments.

This study has some limitations, which also allow to formulate the future directions of additional research. We have not considered the uncertainty resulting from the level of confidence in the data, which is particularly important in areas where the data might not accurately represent reality. The set of signatures could be extended, and various combinations could be considered. The results of multiobjective calibration allow for a wider interpretation and provide many possibilities; however, only a single final solution was evaluated in this study. Further insights into the impact of dataset size on model performance would have been obtained if more data-partitioning approaches were explored and investigated. To extend the conclusions and validate the findings of this study, so that they would have more universal applicability, it is recommended to repeat the same investigation on more catchments with different hydrological regimes.

Author Contributions: Conceptualization, D.P.S.; methodology, D.P.S., M.H., and S.A.M.; software, S.A.M.; validation, D.P.S., M.H., and S.A.M.; formal analysis, S.A.M.; investigation, S.A.M.; resources, D.P.S.; data curation, D.P.S. and S.A.M.; writing—original draft preparation, S.A.M., and M.A.H.; writing—review and editing, S.A.M., D.P.S., M.H., and M.A.H.; visualization, S.A.M., D.P.S. and M.A.H.; supervision, D.P.S. and M.H.; project administration, D.P.S., and M.H.; funding acquisition, D.P.S. and M.A.H. All authors have read and agreed to the published version of the manuscript.

Funding: The APC was funded by the National Water and Energy Center at UAE University Strategic Grant number 12R019.

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: The data used in this study are available from the second author upon reasonable request.

Acknowledgments: The first author is grateful to the Netherlands Fellowship Programme (NFP) for providing the fellowship for the Master study in Hydroinformatics at the IHE Delft Institute for Water Education, in the framework of which this research has been carried out.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gupta, H.V.; Sorooshian, S.; Yapo, P.O. Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resour. Res.* **1998**, *34*, 751–763. [\[CrossRef\]](#)
2. Solomatine, D.P.; Dibike, Y.B.; Kukuric, N. Calage automatique de modèles d'écoulement souterrain utilisant des techniques d'optimisation globales. *Hydrol. Sci. J.* **1999**, *44*, 879–894. [\[CrossRef\]](#)
3. Lawrence, D.; Haddeland, I.; Langsholt, E. *Calibration of HBV Hydrological Models Using PEST Parameter Estimation*; Norwegian Water Resources and Energy Directorate: Oslo, Norway, 2009; ISBN 9788241006807.
4. Patil, S.D.; Stieglitz, M. Comparing spatial and temporal transferability of hydrological model parameters. *J. Hydrol.* **2015**, *525*, 409–417. [\[CrossRef\]](#)
5. Wöhling, T.; Samaniego, L.; Kumar, R. Evaluating multiple performance criteria to calibrate the distributed hydrological model of the upper Neckar catchment. *Environ. Earth Sci.* **2013**, *69*, 453–468. [\[CrossRef\]](#)

6. Fenicia, F.; Solomatine, D.P.; Savenije, H.H.G.; Matgen, P. Soft combination of local models in a multi-objective framework. *Hydrol. Earth Syst. Sci.* **2007**, *11*, 1797–1809. [\[CrossRef\]](#)
7. Sahraei, S.; Asadzadeh, M.; Unduche, F. Signature-based multi-modelling and multi-objective calibration of hydrologic models: Application in flood forecasting for Canadian Prairies. *J. Hydrol.* **2020**, *588*, 125095. [\[CrossRef\]](#)
8. Reed, P.M.; Hadka, D.; Herman, J.D.; Kasprzyk, J.R.; Kollat, J.B. Evolutionary multiobjective optimization in water resources: The past, present, and future. *Adv. Water Resour.* **2013**, *51*, 438–456. [\[CrossRef\]](#)
9. Zhou, A.; Qu, B.Y.; Li, H.; Zhao, S.Z.; Suganthan, P.N.; Zhangd, Q. Multiobjective evolutionary algorithms: A survey of the state of the art. *Swarm Evol. Comput.* **2011**, *1*, 32–49. [\[CrossRef\]](#)
10. Efstratiadis, A.; Koutsoyiannis, D. Une décennie d’approches de calage multi-objectifs en modélisation hydrologique: Une revue. *Hydrol. Sci. J.* **2010**, *55*, 58–78. [\[CrossRef\]](#)
11. Kollat, J.B.; Reed, P.M.; Wagener, T. When are multiobjective calibration trade-offs in hydrologic models meaningful? *Water Resour. Res.* **2012**, *48*, 3520. [\[CrossRef\]](#)
12. Asadzadeh, M.; Tolson, B.A.; Burn, D.H. A new selection metric for multiobjective hydrologic model calibration. *Water Resour. Res.* **2014**, *50*, 7082–7099. [\[CrossRef\]](#)
13. Euser, T.; Winsemius, H.C.; Hrachowitz, M.; Fenicia, F.; Uhlenbrook, S.; Savenije, H.H.G. A framework to assess the realism of model structures using hydrological signatures. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 1893–1912. [\[CrossRef\]](#)
14. Martinez, G.F.; Gupta, H.V. Hydrologic consistency as a basis for assessing complexity of monthly water balance models for the continental United States. *Water Resour. Res.* **2011**, *47*, W12540. [\[CrossRef\]](#)
15. van Werkhoven, K.; Wagener, T.; Reed, P.; Tang, Y. Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models. *Adv. Water Resour.* **2009**, *32*, 1154–1169. [\[CrossRef\]](#)
16. Pokhrel, P.; Yilmaz, K.K.; Gupta, H.V. Multiple-criteria calibration of a distributed watershed model using spatial regularization and response signatures. *J. Hydrol.* **2012**, *418–419*, 49–60. [\[CrossRef\]](#)
17. Pfannerstill, M.; Guse, B.; Fohrer, N. Smart low flow signature metrics for an improved overall performance evaluation of hydrological models. *J. Hydrol.* **2014**, *510*, 447–458. [\[CrossRef\]](#)
18. Asadzadeh, M.; Leon, L.; McCrimmon, C.; Yang, W.; Liu, Y.; Wong, I.; Fong, P.; Bowen, G. Watershed derived nutrients for Lake Ontario inflows: Model calibration considering typical land operations in Southern Ontario. *J. Great Lakes Res.* **2015**, *41*, 1037–1051. [\[CrossRef\]](#)
19. Chilkoti, V.; Bolisetti, T.; Balachandar, R. Multi-objective autocalibration of SWAT model for improved low flow performance for a small snowfed catchment. *Hydrol. Sci. J.* **2018**, *63*, 1482–1501. [\[CrossRef\]](#)
20. Sawicz, K.; Wagener, T.; Sivapalan, M.; Troch, P.A.; Carrillo, G. Catchment classification: Empirical analysis of hydrologic similarity based on catchment function in the eastern USA. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 2895–2911. [\[CrossRef\]](#)
21. Seibert, J.; McDonnell, J.J. On the dialog between experimentalist and modeler in catchment hydrology: Use of soft data for multicriteria model calibration. *Water Resour. Res.* **2002**, *38*, 1241–1252. [\[CrossRef\]](#)
22. Hrachowitz, M.; Fovet, O.; Ruiz, L.; Euser, T.; Gharari, S.; Nijzink, R.; Freer, J.; Savenije, H.H.G.; Gascuel-Oudou, C. Process consistency in models: The importance of system signatures, expert knowledge, and process complexity. *Water Resour. Res.* **2014**, *50*, 7445–7469. [\[CrossRef\]](#)
23. McMillan, H.K.; Clark, M.P.; Bowden, W.B.; Duncan, M.; Woods, R.A. Hydrological field data from a modeller’s perspective: Part 1. Diagnostic tests for model structure. *Hydrol. Process.* **2011**, *25*, 511–522. [\[CrossRef\]](#)
24. Clark, M.P.; McMillan, H.K.; Collins, D.B.G.; Kavetski, D.; Woods, R.A. Hydrological field data from a modeller’s perspective: Part 2: Process-based evaluation of model hypotheses. *Hydrol. Process.* **2011**, *25*, 523–543. [\[CrossRef\]](#)
25. Reusser, D.E.; Zehe, E. Inferring model structural deficits by analyzing temporal dynamics of model performance and parameter sensitivity. *Water Resour. Res.* **2011**, *47*, W07550. [\[CrossRef\]](#)
26. Wagener, T.; Montanari, A. Convergence of approaches toward reducing uncertainty in predictions in ungauged basins. *Water Resour. Res.* **2011**, *47*, 453–460. [\[CrossRef\]](#)
27. Westerberg, I.K.; Guerrero, J.L.; Younger, P.M.; Beven, K.J.; Seibert, J.; Halldin, S.; Freer, J.E.; Xu, C.Y. Calibration of hydrological models using flow-duration curves. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 2205–2227. [\[CrossRef\]](#)
28. Schaeffli, B. Snow hydrology signatures for model identification within a limits-of-acceptability approach. *Hydrol. Process.* **2016**, *30*, 4019–4035. [\[CrossRef\]](#)
29. Shafii, M.; Tolson, B.A. Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. *Water Resour. Res.* **2015**, *51*, 3796–3814. [\[CrossRef\]](#)
30. Yap, P.O.; Gupta, H.V.; Sorooshian, S. Automatic calibration of conceptual rainfall-runoff models: Sensitivity to calibration data. *J. Hydrol.* **1996**, *181*, 23–48. [\[CrossRef\]](#)
31. Kim, U.; Kaluarachchi, J.J. Hydrologic model calibration using discontinuous data: An example from the upper Blue Nile River Basin of Ethiopia. *Hydrol. Process.* **2009**, *23*, 3705–3717. [\[CrossRef\]](#)
32. Sun, W.; Wang, Y.; Wang, G.; Cui, X.; Yu, J.; Zuo, D.; Xu, Z. Physically based distributed hydrological model calibration based on a short period of streamflow data: Case studies in four Chinese basins. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 251–265. [\[CrossRef\]](#)
33. McIntyre, N.R.; Wheeler, H.S. Calibration of an in-river phosphorus model: Prior evaluation of data needs and model uncertainty. *J. Hydrol.* **2004**, *290*, 100–116. [\[CrossRef\]](#)

34. Tan, S.B.; Chua, L.H.; Shuy, E.B.; Lo, E.Y.-M.; Lim, L.W. Performances of Rainfall-Runoff Models Calibrated over Single and Continuous Storm Flow Events. *J. Hydrol. Eng.* **2008**, *13*, 597–607. [\[CrossRef\]](#)
35. Sorooshian, S.; Gupta, V.K.; Fulton, J.L. Evaluation of Maximum Likelihood Parameter estimation techniques for conceptual rainfall-runoff models: Influence of calibration data variability and length on model credibility. *Water Resour. Res.* **1983**, *19*, 251–259. [\[CrossRef\]](#)
36. Li, C.Z.; Wang, H.; Liu, J.; Yan, D.H.; Yu, F.L.; Zhang, L. Effect of calibration data series length on performance and optimal parameters of hydrological model. *Water Sci. Eng.* **2010**, *3*, 378–393.
37. Tada, T.; Beven, K.J. Hydrological model calibration using a short period of observations. *Hydrol. Process.* **2012**, *26*, 883–892. [\[CrossRef\]](#)
38. Perrin, C.; Oudin, L.; Andreassian, V.; Rojas-Serna, C.; Michel, C.; Mathevet, T. Impact of limited streamflow data on the efficiency and the parameters of rainfall-runoff models. *Hydrol. Sci. J.* **2007**, *52*, 131–151. [\[CrossRef\]](#)
39. Reynolds, J.E.; Halldin, S.; Seibert, J.; Xu, C.Y.; Grabs, T. Robustness of flood-model calibration using single and multiple events. *Hydrol. Sci. J.* **2020**, *65*, 842–853. [\[CrossRef\]](#)
40. Seibert, J.; McDonnell, J.J. Gauging the Ungauged Basin: Relative Value of Soft and Hard Data. *J. Hydrol. Eng.* **2015**, *20*, A4014004. [\[CrossRef\]](#)
41. Brath, A.; Montanari, A.; Toth, E. Analysis of the effects of different scenarios of historical data availability on the calibration of a spatially-distributed hydrological model. *J. Hydrol.* **2004**, *291*, 232–253. [\[CrossRef\]](#)
42. Seibert, J.; Beven, K.J. Gauging the ungauged basin: How many discharge measurements are needed? *Hydrol. Earth Syst. Sci.* **2009**, *13*, 883–892. [\[CrossRef\]](#)
43. Pool, S.; Viviroli, D.; Seibert, J. Prediction of hydrographs and flow-duration curves in almost ungauged catchments: Which runoff measurements are most informative for model calibration? *J. Hydrol.* **2017**, *554*, 613–622. [\[CrossRef\]](#)
44. Gharari, S.; Shafiei, M.; Hrachowitz, M.; Kumar, R.; Fenicia, F.; Gupta, H.V.; Savenije, H.H.G. A constraint-based search algorithm for parameter identification of environmental models. *Hydrol. Earth Syst. Sci.* **2014**, *18*, 4861–4870. [\[CrossRef\]](#)
45. Bell, V.A.; Moore, R.J. The sensitivity of catchment runoff models to rainfall data at different spatial scales. *Hydrol. Earth Syst. Sci.* **2000**, *4*, 653–667. [\[CrossRef\]](#)
46. Shrestha, D.L.; Solomatine, D.P. Data-driven approaches for estimating uncertainty in rainfall-runoff modelling. *Int. J. River Basin Manag.* **2008**, *6*, 109–122. [\[CrossRef\]](#)
47. Westerberg, I.K.; McMillan, H.K. Uncertainty in hydrological signatures. *Hydrol. Earth Syst. Sci.* **2015**, *19*, 3951–3968. [\[CrossRef\]](#)
48. Allen, R.G.; Pereira, L.S.; Raes, D.; Smith, M. Crop evapotranspiration—Guidelines for computing crop water requirements. In *FAO Irrigation and Drainage*; FAO: Rome, Italy, 1998.
49. Yadav, M.; Wagener, T.; Gupta, H. Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Adv. Water Resour.* **2007**, *30*, 1756–1774. [\[CrossRef\]](#)
50. Mcmillan, H.; Westerberg, I.; Branger, F. Five guidelines for selecting hydrological signatures. *Hydrol. Process.* **2017**, *31*, 4757–4761. [\[CrossRef\]](#)
51. Yilmaz, K.K.; Gupta, H.V.; Wagener, T. A process-based diagnostic approach to model evaluation Application to the NWS distributed hydrologic model. *Water Resour. Res.* **2008**, *44*. [\[CrossRef\]](#)
52. Arnold, J.G.; Allen, P.M. Automated methods for estimating baseflow and ground water recharge from streamflow records. *J. Am. Water Resour. Assoc.* **1999**, *35*, 411–424. [\[CrossRef\]](#)
53. Shamir, E.; Imam, B.; Gupta, H.V.; Sorooshian, S. Application of temporal streamflow descriptors in hydrologic model parameter estimation. *Water Resour. Res.* **2005**, *41*, 1–16. [\[CrossRef\]](#)
54. Sankarasubramanian, A.; Vogel, R.M.; Limbrunner, J.F. Climate elasticity of streamflow in the United States. *Water Resour. Res.* **2001**, *37*, 1771–1781. [\[CrossRef\]](#)
55. Donnelly, C.; Andersson, J.C.M.; Arheimer, B. Using flow signatures and catchment similarities to evaluate the E-HYPE multi-basin model across Europe. *Hydrol. Sci. J.* **2016**, *61*, 255–273. [\[CrossRef\]](#)
56. Westerberg, I.K.; Wagener, T.; Coxon, G.; McMillan, H.K.; Castellarin, A.; Montanari, A.; Freer, J. Uncertainty in hydrological signatures for gauged and ungauged catchments. *Water Resour. Res.* **2016**, *52*, 1847–1865. [\[CrossRef\]](#)
57. Bergström, S. *Development and Application of a Conceptual Runoff Model for Scandinavian Catchments*; Report RHO 7; Swedish Meteorological and Hydrological Institute: Norrköping, Sweden, 1976; 134p.
58. Lindström, G.; Bergström, S. Improving the HBV and PULSE-models by use of temperature anomalies. *Vannet i Norden* **1992**, *25*, 16–23.
59. Seibert, J. Estimation of Parameter Uncertainty in the HBV Model. *Nord. Hydrol.* **1997**, *28*, 247–262. [\[CrossRef\]](#)
60. Lindström, G.; Johansson, B.; Persson, M.; Gardelin, M.; Bergström, S. Development and test of the distributed HBV-96 hydrological model. *J. Hydrol.* **1997**, *201*, 272–288. [\[CrossRef\]](#)
61. Geem, Z.W.; Kim, J.H.; Loganathan, G.V. A New Heuristic Optimization Algorithm: Harmony Search. *Simulation* **2001**, *76*, 60–68. [\[CrossRef\]](#)
62. Dai, X.; Yuan, X.; Zhang, Z. A self-adaptive multi-objective harmony search algorithm based on harmony memory variance. *Appl. Soft Comput. J.* **2015**, *35*, 541–557. [\[CrossRef\]](#)
63. Deb, K.; Pratap, A.; Agarwal, S.; Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **2002**, *6*, 182–197. [\[CrossRef\]](#)

-
64. Wang, H.; Jiao, L.; Yao, X. Two_Arch2: An Improved Two-Archive Algorithm for Many-Objective Optimization. *IEEE Trans. Evol. Comput.* **2014**, *19*, 524–541. [[CrossRef](#)]
  65. Blazkova, S.; Beven, K. A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic. *Water Resour. Res.* **2009**, *45*, W00B16. [[CrossRef](#)]
  66. Komuro, R.; Ford, E.D.; Reynolds, J.H. The use of multi-criteria assessment in developing a process model. *Ecol. Modell.* **2006**, *197*, 320–330. [[CrossRef](#)]
  67. Zhang, H.; Huang, G.H.; Wang, D.; Zhang, X. Multi-period calibration of a semi-distributed hydrological model based on hydroclimatic clustering. *Adv. Water Resour.* **2011**, *34*, 1292–1303. [[CrossRef](#)]
  68. Krause, P.; Boyle, D.P.; Bäse, F. Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.* **2005**, *5*, 89–97. [[CrossRef](#)]
  69. Madsen, H. Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. *J. Hydrol.* **2000**, *235*, 276–288. [[CrossRef](#)]
  70. Boyle, D.P.; Gupta, H.V.; Sorooshian, S. Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. *Water Resour. Res.* **2000**, *36*, 3663–3674. [[CrossRef](#)]
  71. Gupta, H.V.; Sorooshian, S.; Yapo, P.O. Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. *J. Hydrol. Eng.* **1999**, *4*, 135–143. [[CrossRef](#)]
  72. Dogulu, N.; López López, P.; Solomatine, D.P.; Weerts, A.H.; Shrestha, D.L. Estimation of predictive hydrologic uncertainty using the quantile regression and UNEEC methods and their comparison on contrasting catchments. *Hydrol. Earth Syst. Sci.* **2015**, *19*, 3181–3201. [[CrossRef](#)]