



## Article

# Wavelet Analysis and the Information Cost Function Index for Selection of Calibration Events for Flood Simulation

Sheik Umar Jam-Jalloh <sup>1</sup>, Jia Liu <sup>1,\*</sup>, Yicheng Wang <sup>1</sup>, Zhijia Li <sup>2</sup> and Nyakeh-Momodu Sulaiman Jabati <sup>2</sup>

<sup>1</sup> State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing 100038, China; sheikjamjalloh@gmail.com (S.U.J.-J.); wangych@iwhr.com (Y.W.)

<sup>2</sup> College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China; zjli@hhu.edu.cn (Z.L.); msjabati@yahoo.com (N.-M.S.J.)

\* Correspondence: jia.liu@iwhr.com; Tel.: +86-10-68781656

**Abstract:** Globally, floods are a prevalent type of natural disaster. Simulating floods is a critical component in the successful implementation of flood management and mitigation strategies within a river basin or catchment area. Selecting appropriate calibration data to establish a reliable hydrological model is of great importance for flood simulation. Usually, hydrologists select the number of flood events used for calibration depending on the catchment size. Currently, there is no numerical index to help hydrologists quantitatively select flood events for calibrating the hydrological models. The question is, what is the necessary and sufficient amount (e.g., 10 events) of calibration flood events that must be selected? This study analyses the spectral characteristics of flood data in Sequences before model calibration. The absolute best set of calibration data is selected using an entropy-like function called the information cost function (ICF), which is calculated from the discrete wavelet transform (DWT) decomposition results. Given that the validation flood events have already been identified, we presume that the greater the similarity between the calibration dataset and the validation dataset, the higher the performance of the hydrological model should be after calibration. The calibration datasets for the Tunxi catchment in southeast China were derived from 21 hourly flood events, and the calibration datasets were generated by arranging 14 flood events in sequences from 3 to 14 (i.e., a Sequence of 3 with 12 sets (set 1 = flood events 1, 2, 3; set 2 = flood events 2, 3, 4, . . . , and so on)), resulting in a total of 12 sequences and 78 sets. With a predetermined validation set of 7 flood events and the hydrological model chosen as the Hydrologic Engineering Center (HEC–HMS) model, the absolute best calibration flood set was selected. The best set from the Sequence of 10 (set 4 = S10') was found to be the absolute best calibration set of flood events. The potential of the percentile energy entropy was also analyzed for the best calibration sets, but the ICF was the most consistent index to reveal the ranking based on similarity with model performance. The proposed ICF index in this study is helpful for hydrologists to use data efficiently with more hydrological data obtained in the new era of big data. This study also demonstrates the possibility of improving the effectiveness of utilizing calibration data, particularly in catchments with limited data.

**Keywords:** discrete wavelet transform; HEC–HMS model; Tunxi catchment; percentile energy



**Citation:** Jam-Jalloh, S.U.; Liu, J.; Wang, Y.; Li, Z.; Jabati, N.-M.S. Wavelet Analysis and the Information Cost Function Index for Selection of Calibration Events for Flood Simulation. *Water* **2023**, *15*, 2035. <https://doi.org/10.3390/w15112035>

Academic Editor: Marco Franchini

Received: 28 April 2023

Revised: 24 May 2023

Accepted: 25 May 2023

Published: 27 May 2023



**Copyright:** © 2023 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

Globally, floods are a prevalent type of natural disaster [1,2]. Due to their severe effects on both people and infrastructure, floods are often regarded as devastating natural disasters [3–5]. Floods pose a tremendous risk to human life and significantly damage agricultural production, buildings, and infrastructure. Flooding can have a significant effect on socioeconomic activities, human health, and death rates, in addition to having a devastating impact on physical infrastructure [6]. Simulating floods is a critical component in the successful implementation of flood management and mitigation strategies within

a river basin or catchment area. Furthermore, flood simulation can serve as a potential flood early warning system, allowing for the protection of lives and property [7]. Selecting appropriate historical data to calibrate and establish a reliable hydrological model is of great importance for flood simulation, as it directly affects the accuracy and reliability of the simulation results [8,9]. With the advent of the big data age, hydrological data are collected with higher frequency and resolution by modern telecommunication systems. In addition to resolving manual and automatic optimization-related calibration challenges which have been the focus of research efforts over the past two decades, researchers have recently placed an increasing emphasis on the significance of selecting appropriate flood data for calibration purposes. This emphasis on carefully selecting flood data is meant to make calibration procedures reliable and efficient. Generally, hydrologists try to use as many flood events as possible to select the calibration set of data that can “represent” the different phenomena observed within the study catchment. While it may appear advantageous to utilize more flood events, the information quality of the data is more important in determining how well the model performs after calibration. It should be noted that beyond a certain level, the accuracy of the parameter estimates will only slightly increase with the addition of more data [10]. Calibration flood events significantly have an impact on the determination of the model parameters, particularly those relating to surface runoff generation and concentration, and can lead to inaccurate estimates of the model parameters and increased uncertainty [11].

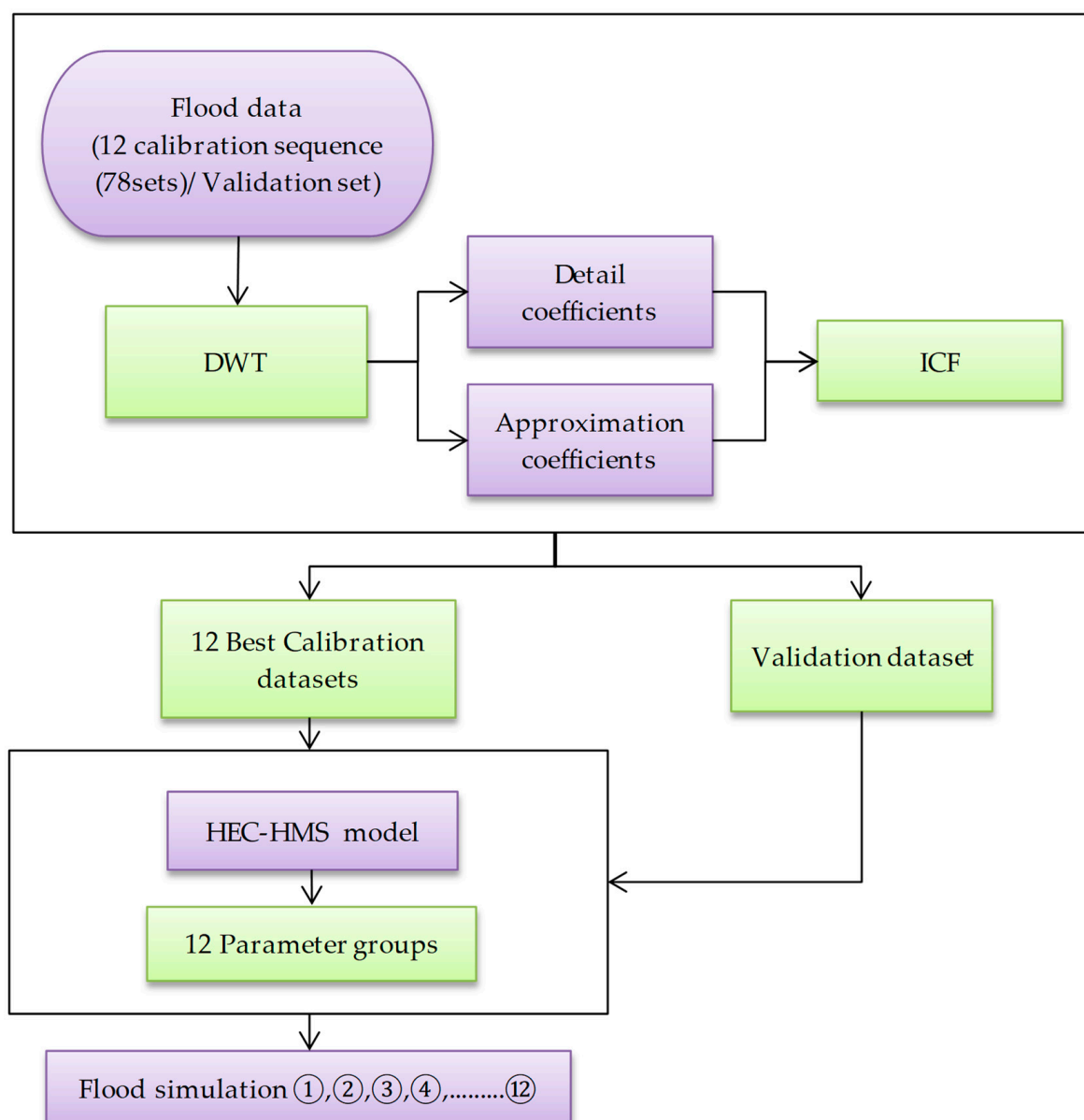
Numerous researchers have concentrated on determining the optimal amount of flood calibration data and have shown that the number of flood events used for calibration does not give a better model performance. Depending on the models employed in their research, different numbers of events are recommended for calibration [12–19]. In the 1960s, Dooge (1969) [13] proposed a method for calibrating hydrological models based on the analysis of flood frequency data. Dooge suggested that the appropriate number of flood events for the calibration should depend on the catchment size and the frequency of extreme events. Later, important contributions were made by Bruen and Dooge (1992) [14], proposing a regularization method with additional information by a split sample test of data from 30 catchments. The method was only useful when only a few flood events were available for analysis. Vieux, Cui, and Gaur (2004) [15] initially used 8 events for calibration, and the number of events was later augmented to 18 events in their study. An interesting discovery regarding the stability of calibrated parameters from the initial storm series of 8 to 18 events was made. They came to the conclusion that parameter values barely changed as more events were included to increase the number of storms. Reynolds et al. (2020) [18] investigated whether a few flood-event hydrographs in a tropical basin would be sufficient to calibrate a bucket-type rainfall-runoff model. They found that when one event was used for calibration, as opposed to using no discharge data, flood predictions were already more accurate. The results also revealed that using two to four events for calibration had a significant positive impact on both the accuracy of the flood predictions and the reduction in uncertainty, whereas using additional events produced only modest performance improvements. Gupta and Sorooshian (1985a, 1985b) [20,21] also conducted a theoretical investigation which demonstrated that data in sequences exhibiting higher “hydrologic variability” have a greater possibility to yield accurate parameter estimates, and this is expected to lead to an improvement in the performance of the model after calibration.

However, rather than examining the inherent qualities of the calibration data themselves, the conclusions of these studies were all dependent on the features of their case studies and the hydrological models used. There is increasing attention paid to selecting the most appropriate calibration flood data that are both representative and of adequate quantity in order to minimize the challenges associated with calibration and produce a reliable hydrological model. This challenge will worsen as modern telemetry systems collect more and more observed data. Currently, there are limited simple yet efficient techniques for choosing the best calibration dataset based only on the data. The question is, is it possible to select the appropriate amount of calibration flood events before the actual

calibration process? Moreover, the performance of the model remains unknown until the calibration process has been completed. Are there any existing criteria for selecting the correct calibration flood events? Let us assume that the flood events which will be used to validate the model have already been determined. By establishing a set of criteria that can be used to evaluate the similarity between several calibration flood datasets and the validation dataset, the process can be made simple. It can be assumed that the accuracy of the calibrated model is directly proportional to the degree of similarity between the calibration flood dataset and the validation set. In other words, the expected improvement in the performance of the hydrological model after calibration increases with the similarity of the calibration flood dataset to the validation set.

The primary objective of this study is to employ wavelet analysis and the information cost function (ICF) index to determine the best set of calibration flood events that exhibit the highest level of similarity to the validation dataset. Furthermore, this study aims to evaluate whether the hydrological model performance is consistent with the ranking of similarity indicated by the ICF index after calibration, as well as to explore the potential of another entropy index to determine the relationship between the similarity and performance of the model. Various hydrology and water resource-related disciplines have recently shown an increased interest in using wavelet analysis. Particularly in terms of periodicity, numerous research studies have demonstrated that wavelet analysis is a useful tool for characterizing and analyzing climatic and hydrological data [22–31], and, also, the frequency domain variable structures of various climatic or hydrological variables can be examined and synthesized using wavelet analysis, which is also an effective method for exploring relationships between them [32–41]. Recent research has shown that wavelet analysis is an effective technique for identifying irregularly distributed multiscale characteristics in hydrometeorological data, and, also, the ability to establish quantitative correlations between various observation series using wavelet-based expressions has been demonstrated [42–51]. The wavelet transform technique can also be used to improve the accuracy of machine learning models' prediction ability of groundwater level and qanat water flow [52,53].

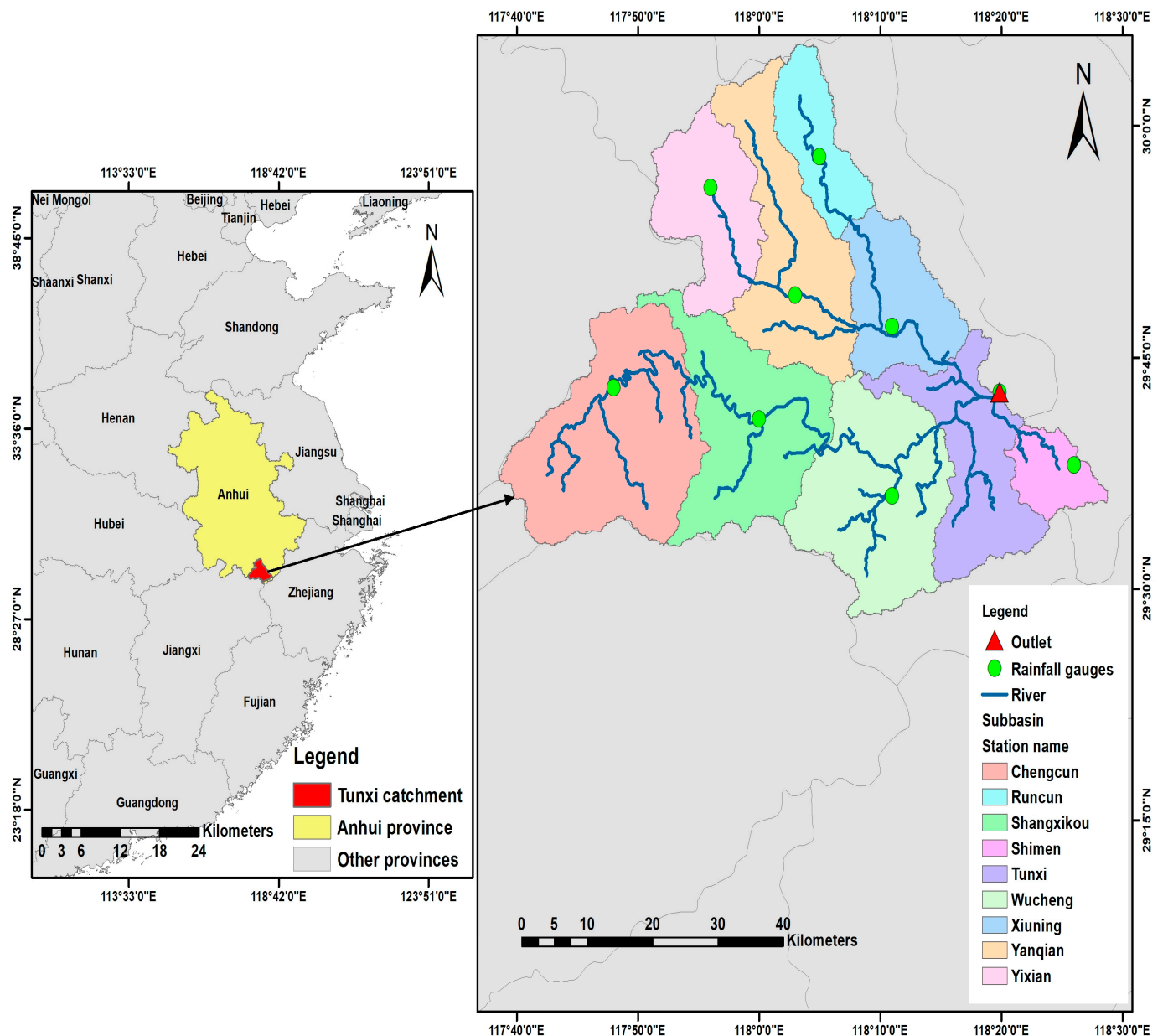
The discrete wavelet transform (DWT) method and a constructed entropy-like metric called the information cost function (ICF) were employed in this study. The ICF metric value, calculated from wavelet analysis results, was used to evaluate the spectral characteristics of the calibration and validation flood datasets and to determine their degree of similarity. The observed flow data are employed in this process. The Hydrologic Engineering Center (HEC) hydrological model performance verified the similarity ranking after model calibration. From the 21 flood events, 7 flood events were selected for validation, and 14 events were used for calibration. The 14 flood events were arranged in sequences containing from 3 to 14 flood events (i.e., a sequence of 3 flood events resulting in 12 calibration datasets (set 1 = flood events 1, 2, 3; set 2 = flood events 2, 3, 4, . . . , and so on). This resulted in a total of 12 sequences and 78 datasets. The primary objective of the DWT analysis was to identify the best calibration set from each Sequence, as well as the absolute best set from all the identified best sets. Furthermore, this study aimed to investigate whether the degree of similarity indicated by the wavelet analysis and ICF index aligned with the results obtained from the calibrated model's performance after utilizing the best datasets from the sequences. The potential of the percentile energy entropy was also analyzed for the best calibration sets. The flow chart of this study is shown in Figure 1.



**Figure 1.** Flowchart of the study.

## 2. Study Area and Events

In this study, the Tunxi (TX) catchment, an inland catchment near the southeast of China's Anhui province, is used as the study area. The Tunxi catchment is mesoscale, and the catchment area is around 2754 km<sup>2</sup>. The mean elevation of this catchment is about 380 m a.s.l., with the lowest point at 116 m and the highest at 1398 m. The Tunxi catchment has a subtropical monsoon climate with a mean annual temperature of 17 °C. This catchment is a typical humid region with an annual rainfall of 1600 mm, with 50% of the precipitation occurring from April to June, which is the period most prone to flooding. The vegetation in the study area is in good condition, with predominant species including evergreen coniferous forest, deciduous broad-leaved forest, and mixed forest. The soil type in the area is primarily characterized as clay loam. The watershed is divided into 9 sub-basins, with each sub-basin having its respective hydrological station: Yanqian, Chengcun, Shangxikou, Wucheng, Yixian, Runcun, Tunxi, Xiuning, and Shimen. Tunxi is the outlet station of the watershed. The division, locations of drainages, and stations of the study catchment are shown in Figure 2.



**Figure 2.** Location and division of the Tunxi catchment with the hydrological stations in each subbasin.

Based on the 6-year hourly observational data from 2008 to 2013 in the Tunxi catchment, 21 flood events were selected for model calibration and validation in this study. Of these, 14 flood events that occurred during the first four (4) years, from 2008 to 2011, were used for model calibration, and 7 flood events during the last two (2) years, from 2012 to 2013, were used for model validation. With the 7 flood events selected for validation, the remaining 14 flood events were arranged in calibration sequences containing 3 to 14 events. That is, a Sequence of 3 with 12 sets (set 1 contains flood events 1, 2, 3; set 2 contains flood events 2, 3, 4; . . . , and so on), a Sequence of 4 with 11 sets, a Sequence of 5 with 10 sets, a Sequence of 6 with 9 sets, a Sequence of 7 with 8 sets, a Sequence of 8 with 7 sets, a Sequence of 9 with 6 sets, a Sequence of 10 with 5 sets, a Sequence of 11 with 4 sets, a Sequence of 12 with 3 sets, a Sequence of 13 with 2 sets, and a Sequence of 14 with 1 set, resulting in a total of 12 sequences and 78 sets, as shown in Table 1.

**Table 1.** Calibration event arrangement.

	Calibration Events													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Sequence of 3		Set 1												
			Set 2											
				Set 3										
					Set 4									
						Set 5								
							Set 6							
								Set 7						
									Set 8					
										Set 9				
											Set 10			
												Set 11		
													Set 12	
Sequence of 4	a moving window of 4				...									
Sequence of 5	a moving window of 5					...								
Sequence of 6	a moving window of 6						...							
Sequence of 7	a moving window of 7							...						
Sequence of 8	a moving window of 8								...					
Sequence of 9	a moving window of 9									...				
Sequence of 10	a moving window of 10										...			
Sequence of 11	a moving window of 11											...		
Sequence of 12	a moving window of 12												...	
Sequence of 13	a moving window of 13													...
Sequence of 14	Set 1													

It is important to note that, while selecting validation flood events is crucial to evaluating the calibrated model, this study focuses on examining the influence of the number and Sequence of the calibration events on the calibration outcomes. Therefore, the set of validation events remained fixed throughout the investigation. The performance of the calibrated model using the best sets of calibration sequences was evaluated based on the results of the validation set of data, which served as the criteria for evaluation. The flood events in Tunxi catchment used for calibration and validation are shown in Table 2.

**Table 2.** Flood events in the Tunxi catchment used for model calibration and validation.

	Event	Start Date	Start Time	End Date	End Time	Peak Flow (m <sup>3</sup> /s)
Calibration	1	27 May 2008	3:00:00 PM	5 June 2008	12:00:00 AM	1340.0
	2	7 June 2008	8:00:00 PM	13 June 2008	4:00:00 AM	5290.0
	3	13 June 2008	5:00:00 AM	17 June 2008	12:00:00 AM	1900.0
	4	17 June 2008	1:00:00 AM	22 June 2008	12:00:00 AM	1860.0
	5	29 July 2008	8:00:00 AM	8 August 2008	12:00:00 AM	1200.0
	6	19 April 2009	8:00:00 AM	29 April 2009	12:00:00 AM	585.0
	7	26 July 2009	12:00:00 PM	5 August 2009	12:00:00 PM	1300.0
	8	10 April 2010	8:00:00 PM	25 April 2010	12:00:00 AM	1123.3
	9	16 May 2010	8:00:00 AM	27 May 2010	8:00:00 AM	1970.0
	10	6 July 2010	8:00:00 AM	13 June 2010	8:00:00 PM	1870.0
	11	13 July 2010	11:00:00 PM	28 June 2010	8:00:00 AM	1700.0
	12	10 May 2011	10:00:00 PM	19 May 2011	3:00:00 AM	475.8
	13	9 June 2011	4:00:00 PM	14 June 2011	4:00:00 AM	3400.0
	14	14 June 2011	6:00:00 AM	25 June 2011	5:00:00 PM	5230.0



Table 2. Cont.

	Event	Start Date	Start Time	End Date	End Time	Peak Flow (m <sup>3</sup> /s)
Validation	15	27 February 2012	12:00:00 PM	15 March 2012	6:00:00 AM	1040.0
	16	21 April 2012	10:00:00 PM	28 April 2012	12:00:00 PM	3170.0
	17	22 June 2012	2:00:00 AM	8 July 2012	8:00:00 PM	1200.0
	18	6 August 2012	7:00:00 PM	14 August 2012	3:00:00 AM	2641.7
	19	28 April 2013	12:00:00 AM	4 May 2013	11:00:00 PM	2228.3
	20	5 June 2013	1:00:00 PM	14 June 2013	5:00:00 AM	3610.0
	21	24 June 2013	7:00:00 AM	4 July 2013	7:00:00 AM	4214.6

### 3. Methodology

#### 3.1. Wavelet Analysis and the Information Cost Function (ICF)

Wavelet analysis is a mathematical technique used to simultaneously analyze signals or data in both the time and frequency domains [54,55]. It decomposes a signal into small wavelets, which are well-localized waveforms scaled and translated across the time and frequency domains. Compared to other signal analysis techniques such as Fourier analysis, wavelet analysis offers several advantages, such as the ability to accurately deconstruct and reconstruct finite, nonstationary signals and accurately represent functions with sharp peaks and discontinuities [56]. Also, it can capture transient events or sudden changes in a signal more accurately, whereas Fourier analysis can only provide information about the overall frequency content of a signal.

Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT) are the two main types of wavelet transforms. The main difference between Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT) is in the way they utilize wavelets. CWT applies wavelets of varying scales and locations, resulting in continuous variations in scale parameters and the translation of the wavelets. Unlike CWT, which uses a continuous set of wavelets, DWT employs a finite set of wavelets that are sampled discretely. The most commonly used form of DWT has scales and locations arranged in a dyad structure, meaning that the scales and locations are powers of two. More detailed information on the methodology of the CWT and DWT wavelet transforms is found in Meyer (1993) [57]. DWT is simpler and easier to use than CWT. The algorithm provided by Mallat is an efficient approach for implementing the Discrete Wavelet Transform (DWT) [58]. The original signal  $f$  is decomposed into a sequence of approximations and details by the algorithm using several successive filtering steps shown below:

$$S_n^0 = f[n], n \in \mathbb{N} \quad (1)$$

$$S_k^j = \sum_{n=0}^{L-1} h[n] S_{n+2k}^{j-1}, j = 1, 2, \dots, J \quad (2)$$

$$C_k^j = \sum_{n=0}^{L-1} g[n] S_{n+2k}^{j-1}, j = 1, 2, \dots, J, \quad (3)$$

where  $f[n]$  is the original signal;  $\mathbb{N}$  denotes the sum of the data points in the signal  $f$ ;  $S_k^j$  and  $C_k^j$  denote the approximation and detail coefficients, respectively; the low-pass filter  $H$  and the high-pass filter  $G$  impulse responses are denoted as  $h[n]$  and  $g[n]$ , respectively;  $J$  denotes the largest scale that can be achieved by the Mallat decomposition algorithm, where  $J \leq [\log_2(N - L) + 1]$ ; and  $L$  denotes the number of impulse responses in  $h[n]$  and  $g[n]$  not equal to zero [59]. The initial step of the Mallat decomposition algorithm involves decomposing the original signal  $f$  into an approximation and its corresponding detail. This is achieved when the signal is convolved with the decomposition low-pass filter  $H$  to obtain the approximation coefficients  $S_k^j$  and the high-pass filter  $G$  to obtain the detail coefficients

$C_k^j$ . This procedure is repeated iteratively by breaking down each successive approximation, resulting in the original signal decomposing into numerous lower-resolution components. The details represent the low-scale, high-frequency components of the signal, while the approximations represent the high-scale, low-frequency components. The number of approximation and detail coefficients at each level depends on the length of the original signal  $f$  and the lengths of the  $h[n]$  and  $g[n]$  impulse responses.

Wavelets and the information cost function are related in that wavelets can be used to analyze the frequency components of a signal, which can, in turn, be used to estimate the information content or complexity of the signal. After decomposing the signal into wavelet coefficients  $C_k^j$  and  $S_k^j$ , the total  $E_j = \sum_k C_k^{j2}$  (or  $E_j = \sum_k S_k^{j2}$ ) provides the approximate (or detailed) energy at level  $j$  of the signal  $f$ . If  $E_{tot} = \sum_j E_j$  represents the total energy, then, at level  $j$ , the resulting percentile energy is as follows:

$$P_j = \frac{E_j}{E_{tot}} \quad (4)$$

Each level  $j$  is linked with a frequency band,  $\Delta F$ , which is determined as below:

$$2^{-j-1}F_s \leq \Delta F \leq 2^{-j}F_s, \quad (5)$$

where  $F_s$  represents the frequency of the samples, and  $j = 1, 2, \dots, J$ .

The energy probability distribution for each level  $j$  is provided by the Sequence  $P_j$ . The Shannon entropy of this distribution is a measure of the order within the system, and this is the information cost function [60]:

$$ICF = - \sum_j P_j \ln P_j, \quad (6)$$

where the sum is considered to be zero for any level  $j$  with  $P_j = 0$ . This entropy-like function, ICF, which is easy to calculate, provides a straightforward way to estimate the level of disorder in a system [61]. This study utilized the ICF as a metric for evaluating the similarity between the calibration and validation datasets in the frequency domain.

### 3.2. The HEC–HMS Model

The hydrological model used in this study is the HEC–HMS (Hydrologic Engineering Center’s Hydrologic Modeling System) model (Version 4.7.1), which was developed by the United States Army Corps of Engineers and is designed to simulate many hydrological processes of dendritic watershed systems, such as investigating urban flooding, the frequency of flooding, the planning of flood warning systems, the capacity of reservoir spillways, stream restoration, etc. [62].

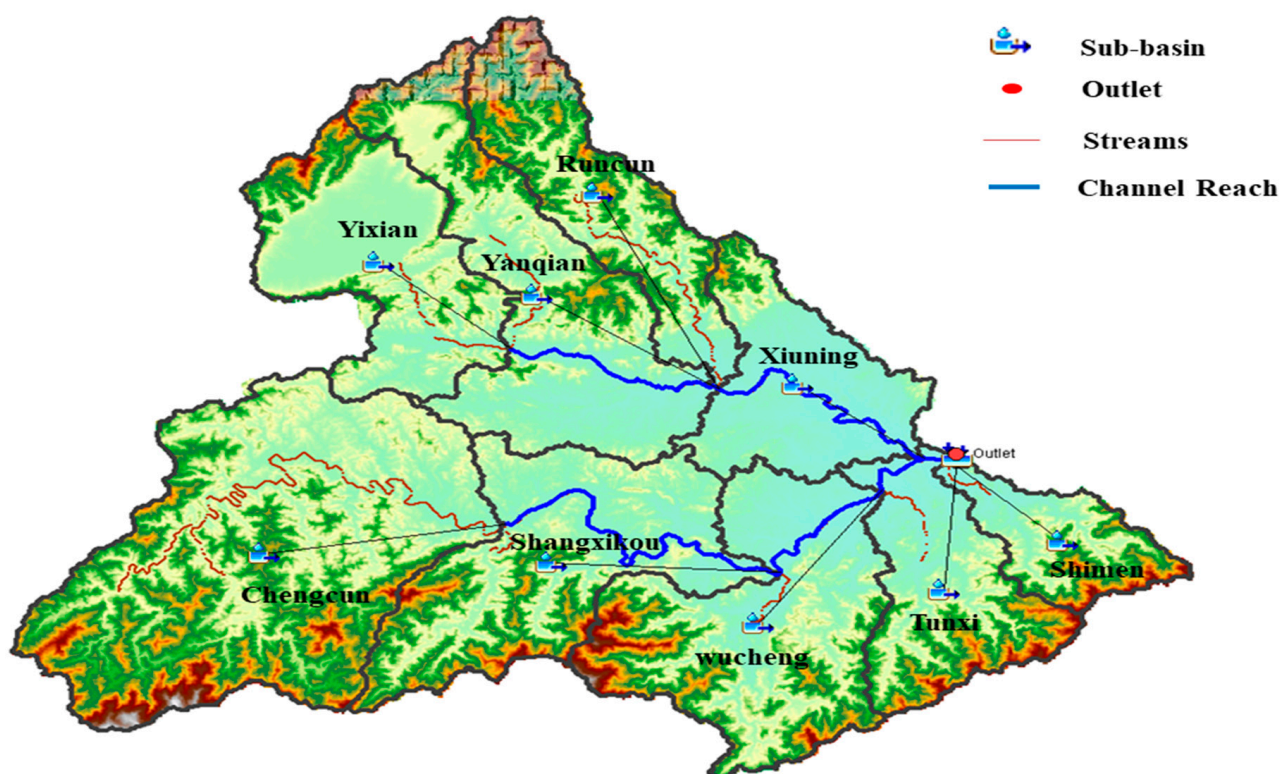
The structure of the HEC–HMS model generally includes the following components:

- Watershed delineation: The first step in creating an HEC–HMS model is to delineate the boundaries of the watershed;
- Meteorological data: HEC–HMS requires rainfall data as input. These data can be obtained from various sources, including weather stations, radar, and satellite data;
- Hydrologic data: HEC–HMS also requires hydrologic data such as streamflow data and soil properties;
- Model parameters: The HEC–HMS model requires input parameters such as infiltration parameters, routing coefficients, and curve numbers. These parameters can be obtained from literature or calibrated using observed data;
- Hydrologic models: HEC–HMS offers a variety of hydrologic models to simulate the rainfall-runoff process, including the SCS (Soil Conservation Service) Curve Number method, the Green–Ampt infiltration method, and the Muskingum–Cunge routing method;



- **Simulation and analysis:** After input data and model parameters are provided, the HEC–HMS model simulates the rainfall–runoff process and provides output data such as hydrographs and flood volumes.

Overall, the HEC–HMS model provides a comprehensive tool for studying the hydrologic response of a watershed to rainfall events and for planning and managing water resources [62]. The studied watershed was delineated into nine sub-basins, as shown in Figure 3. Channel Reach represents a segment of the stream or river with similar or varying hydrologic conditions between two streamgages. The transformation of excess precipitation into direct surface runoff was modeled using the SCS (Soil Conservation Service) unit hydrograph method. More detailed information on the SCS method can be found in Ara and Zakwan 2018 [63]. For model infiltration loss, the initial and constant method was used. For model baseflow, the exponential recession model was employed. The Muskingum routing model was used for river routing. More detailed information on the Muskingum routing model can be found in Niazkar and Zakwan (2022) [64]. In normal cases, a subjective adjustment of parameters is employed to calibrate the model by the trial-and-error method. Even though the model can be calibrated manually, the HEC–HMS also provides an automatic built-in optimization procedure that can be used to verify the appropriateness and feasibility of the parameter values and their ranges for their intended use in the model. The objective function in this study is the Nash–Sutcliffe efficiency coefficient (NSE) [65].



**Figure 3.** HEC–HMS schematic DEM map of the Tunxi catchment.

## 4. Results

### 4.1. Flow Similarity Identified by ICF

The Information Cost Function (ICF) values were calculated to investigate the relationship between the spectrum similarity of the calibration flood datasets and the validation set. This study calculated the ICF value by applying the Discrete Wavelet Transform (DWT) to detailed frequency domain subdivisions. Specifically, the calibration flood datasets and the validation set were decomposed into 6 levels of detail (d1–d6) and approximation (a1–a6),

using the simple Daubechies wavelet for decomposition with an order of 10 (db10). More detailed information on the Daubechies wavelets is found in Daubechies (1990) [54]. The ICF value of each calibration set, which contained different numbers of flood events in each Sequence, was calculated to identify the best calibration sets, with the assumption that a closer ICF value of the calibration flood dataset to the validation set ICF value will produce an improved hydrological model performance after calibration. The calibration flood datasets of each of the 12 Sequences (Sequences of 3 to 14) are plotted against the ICF values in Figure 4. It can be seen that the best datasets in each subfigure (a–l) are set 6, set 4, set 7, set 4, set 7, set 4, set 5, set 4, set 3, set 2, set 1, and set 1, respectively, for the calibration sequences containing 3 to 14 flood events. Since there were 14 flood events for calibration, the Sequence of 14 has only 1 dataset. In the following analyses, we used the best calibration set from each Sequence to represent the Sequences with different numbers of flood events; that is, S3' referred to the best calibration set from Sequence of 3, S4' meant the best set from Sequence of 4, and so on, up to S14' meant the best set from Sequence of 14.

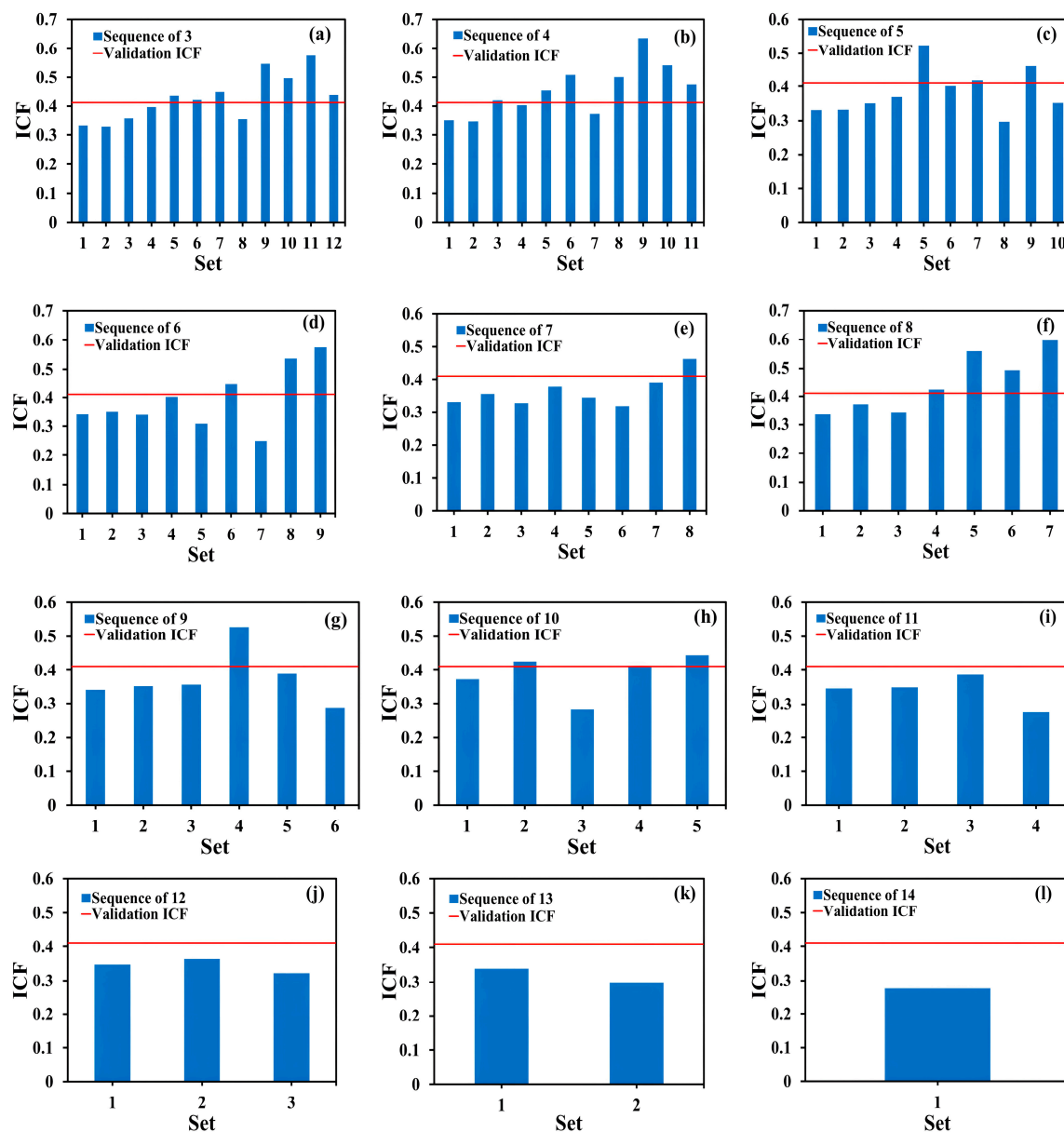
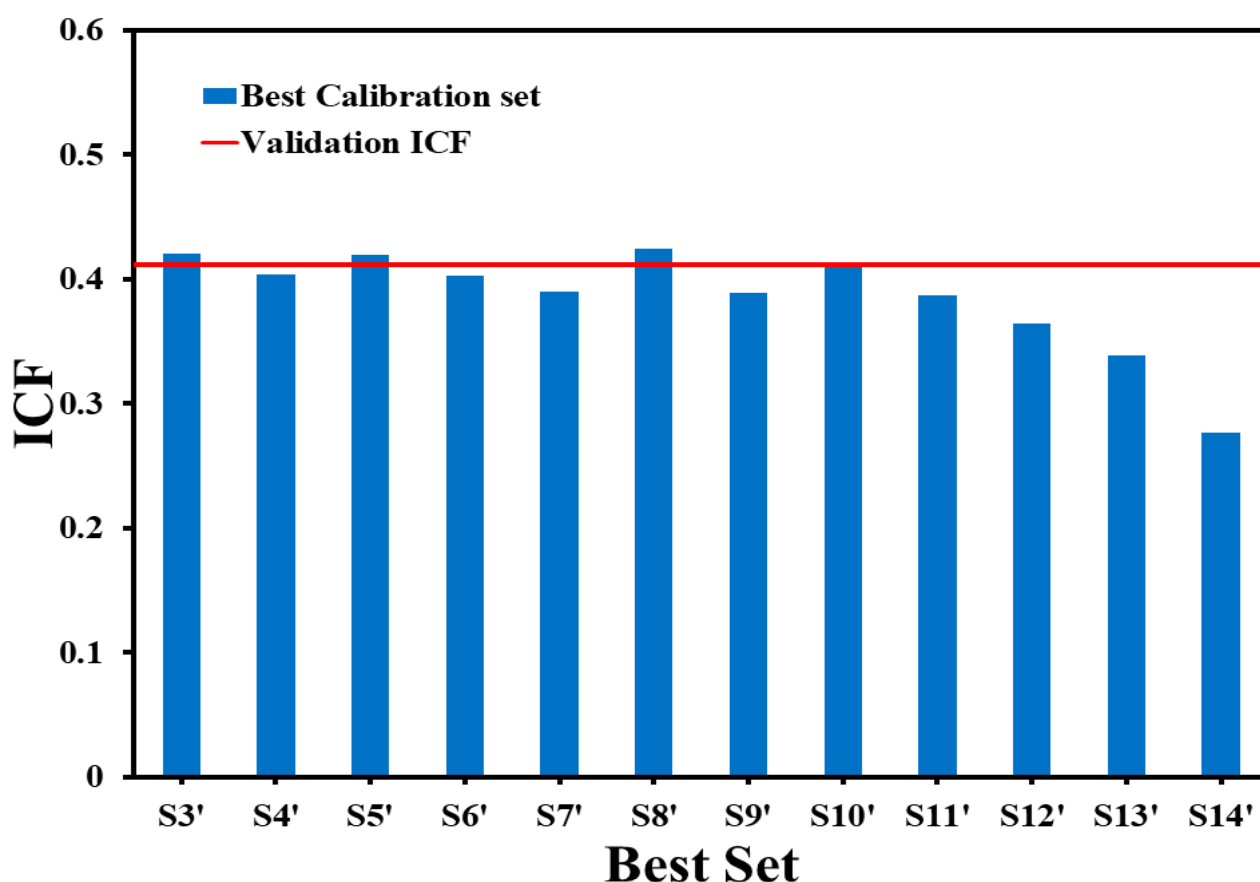


Figure 4. The ICF values of all calibration datasets for each Sequence.

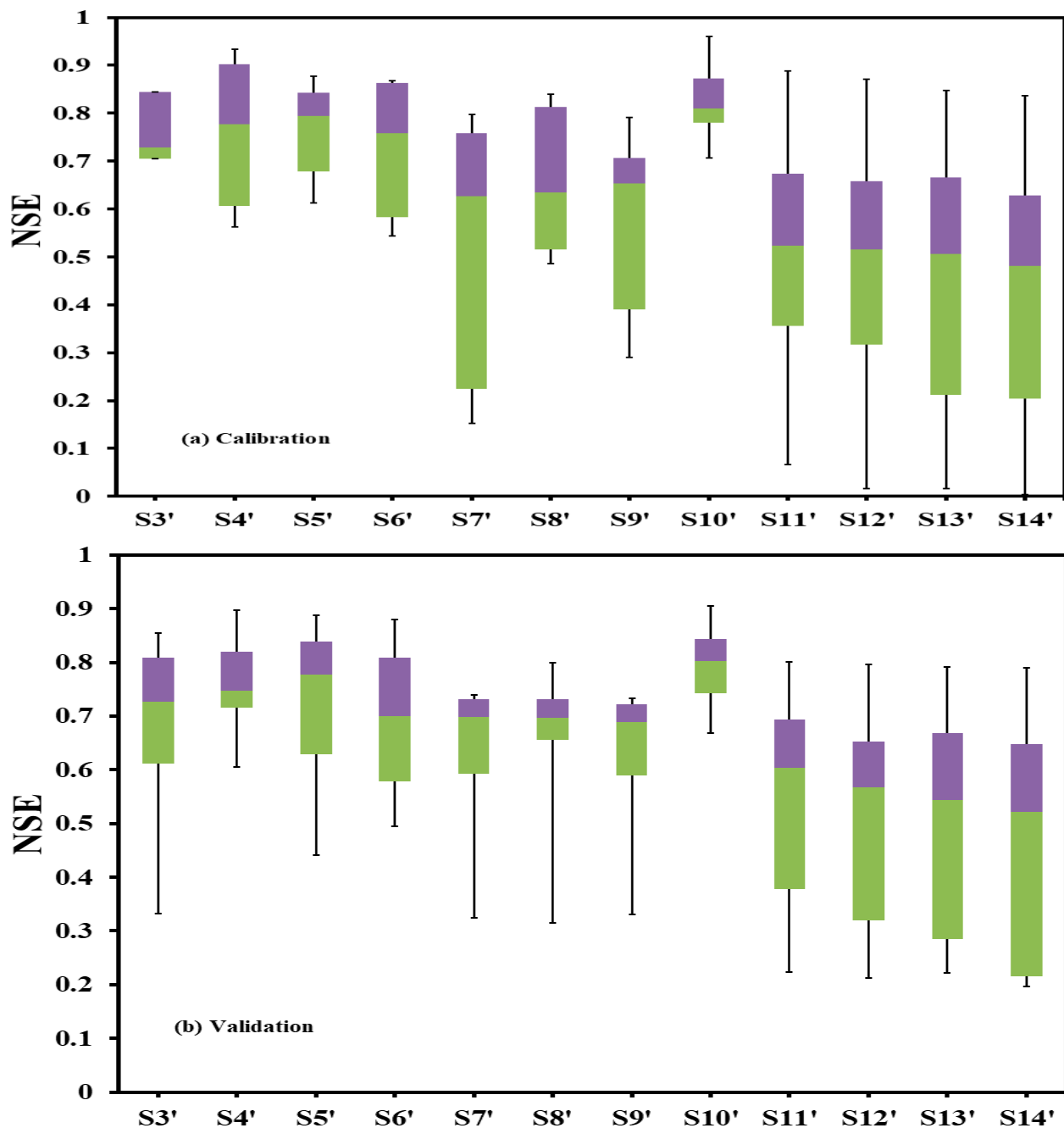
According to the assumption, the closest ICF value of the calibration flood datasets to the validation set ICF value results in the most-improved hydrological model performance after calibration. We investigated the absolute best calibration set from all the best sets from each Sequence. The ICF values of all the best calibration datasets are plotted against the validation ICF in Figure 5. For all 12 best calibration sets from the different sequences, the ICF value of S10' was the closest to the validation ICF, followed by S4', S5', S6', S3', S8', S7', S9', S11', S12', S13', and S14'.



**Figure 5.** The ICF values of all the best calibration sets.

#### 4.2. Model Performances of the Best Calibration Datasets

Calibration runs were conducted for the 12 best calibration datasets from each Sequence with various numbers of flood events. For each set, manual calibration was performed with the help of the optimization trail runs to help obtain the best fit between the observed and simulated flood events. All the parameters calibrated using the 12 calibration sets were then applied to the validation dataset, which contained 7 flood events to evaluate their performance. In addition to the Nash–Sutcliffe Efficiency (NSE), various statistical measures were used to evaluate the performance of the model based on the validation results. These measures included the percent relative error of peak flow and runoff volume, the root-mean-square error, the mean bias error, and the correlation coefficient. However, due to the consistent results obtained from all the evaluation metrics, the NSE was selected as the only indicator for analyzing the model's performance and, hence, is the only measure presented in this paper. The calibration and validation results of the model using the best calibration dataset from each Sequence are presented in a boxplot in Figure 6 that compares the Nash–Sutcliffe Efficiency (NSE) Coefficient Statistics. The height represents the range of NSEs, the top cap represents the maximum NSE value, the bottom cap represents the minimum NSE value, the green and purple boxes represent the 25th to 75th percentile, and the line between the boxes represents the median NSE value.

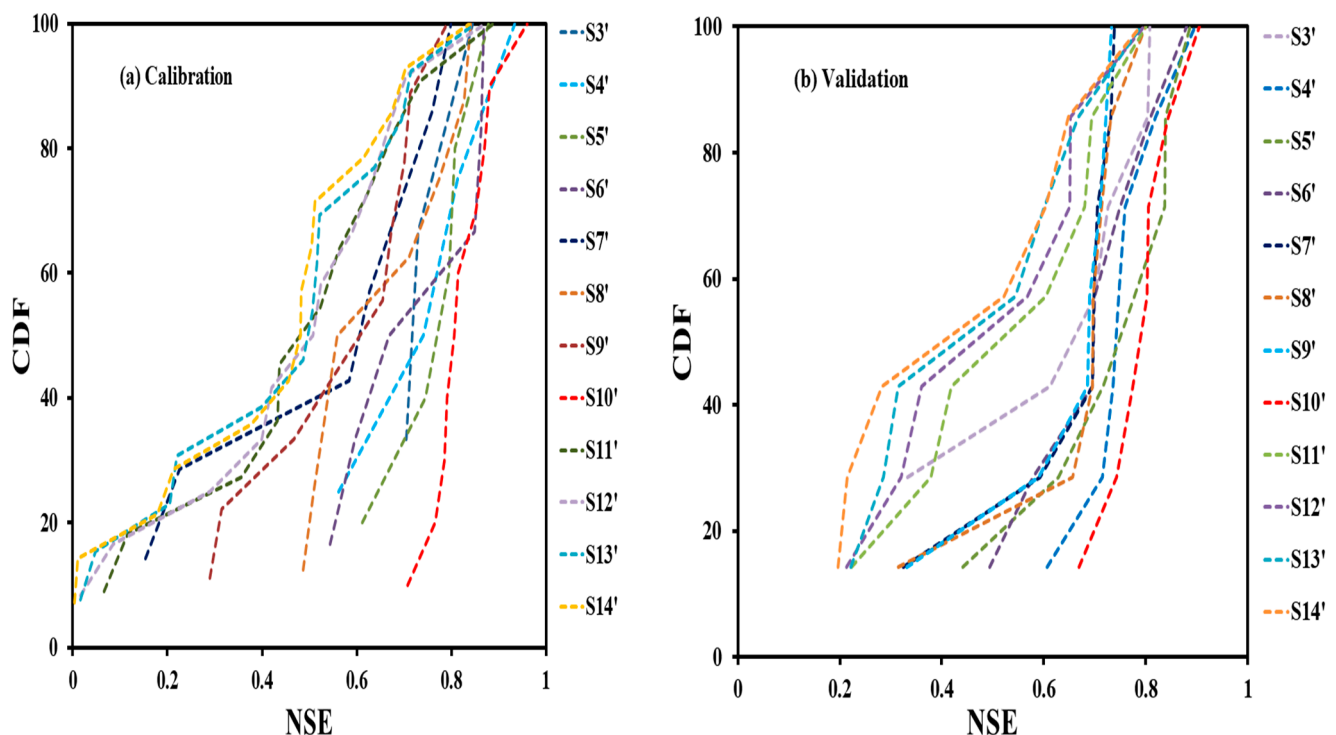


**Figure 6.** Boxplot of the NSE values of the model performances for (a) calibration and (b) validation results of the best calibration datasets for each Sequence.

In general, it can be observed, from the comparison of the model performances of the best calibration datasets from the sequences shown in Figure 6a,b, that the validation results exhibited slightly lower performance than the calibration results. Mostly, well-calibrated models have better NSE values than validation. The S10' calibration set performed best, with the highest maximum NSE value and a better range of NSE values in both calibration and validation, followed by the S4', S5', S6', S3', S8', S7', S9', S11', S12', S13', and S14' datasets. The S7' and S9' datasets in both calibration and validation did not have higher maximum NSE values compared to the S11', S12', S13', and S14' datasets, but their NSE ranges were better. In both calibration and verification stages of the sequences, some flood events were underestimated, as not all the flood events had NSE values  $\geq 0.70$ , especially for floods with high peaks. This may be due to the runoff generation mechanism, since the

model assumes complete saturation of the unsaturated soil layer before the overland flow generation and can only account for the portion near the peak. This assumption might not be accurate for the steep ground topography of the study basin, especially when heavy rainstorms happen. The topographic features influence the runoff generation mechanism and the convergence process of the basin runoff [66].

To further verify the model performance, empirical cumulative distribution functions (CDFs) were constructed for the NSE statistic results to represent the performance of the model in both calibration and validation stages, as shown in Figure 7. Let us suppose that a calibration flood dataset from the sequences is randomly selected. In that case, the CDF of each dataset represents the probability of obtaining an NSE value that is less than or equal to a specific value. In Figure 7a,b, it can be seen that CDFs became increasingly steep and narrow as the Sequence progressed through S10', S4', S5', S6', S3', S8', S7', S9', S11', S12', S13', and S14'. An increasing steepness in the CDFs indicates that the model performance was less sensitive to the selection of Sequences with varying amounts of flood events [67], which means that the S14' set could yield more stable model performances than the other sequences, in descending order according to the CDF list order above (S14' to S10'). The CDF also indicates how fast the Sequence can reach the best NSE. The CDF also verified that the S7' and S9' sets did not have higher maximum NSE values compared to the S11', S12', S13', and S14' sets, but their NSE ranges were better. Similar results can be found in Table 3, showing the NSE statistics results of the best calibration datasets. The validation results of the S11', S12', S13', and S14' datasets had higher maximum NSE values, even though the S7' and S9' datasets produced an improved model performance after calibration.



**Figure 7.** Empirical cumulative distribution functions (CDFs) of the NSE values of (a) calibration and (b) validation results of the calibration best datasets from each Sequence.

**Table 3.** Nash–Sutcliffe Efficiency Coefficient Statistics (NSE) of the validation results when comparing the average model performances produced by the best calibration datasets.

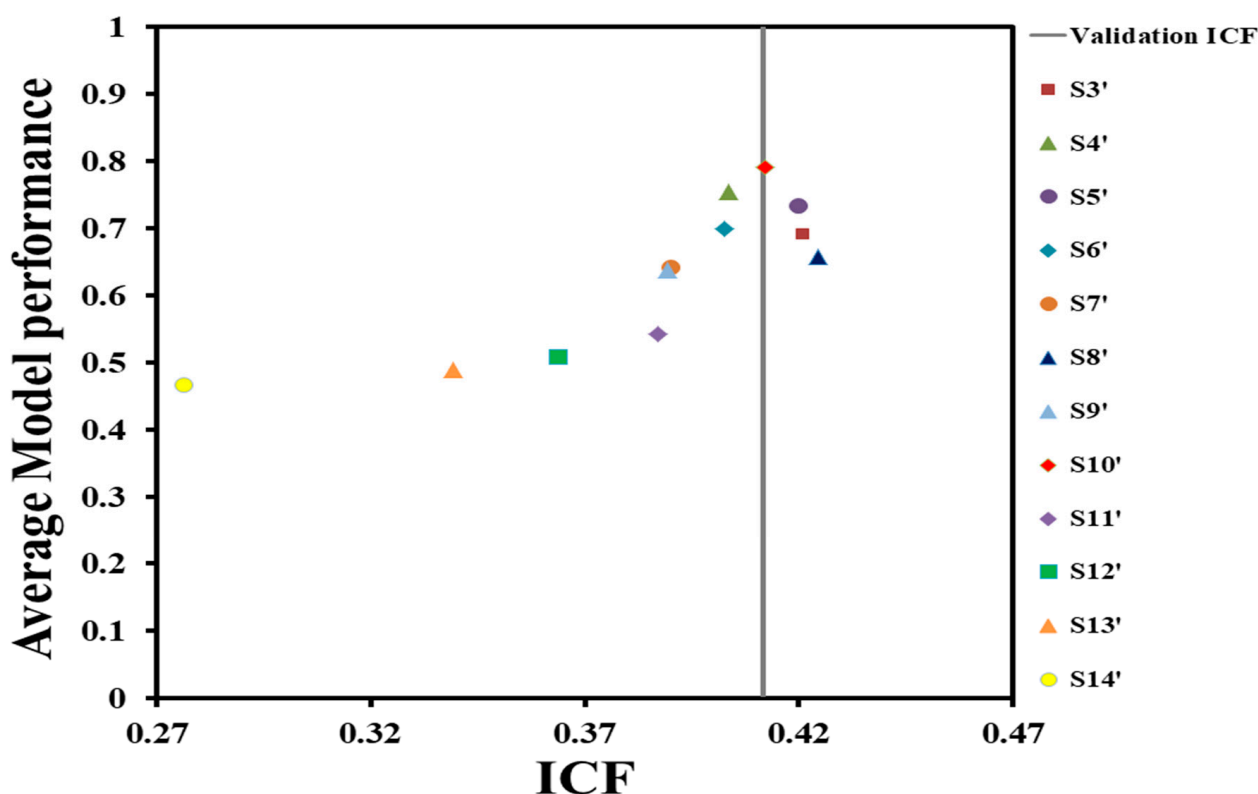
Validation Nash–Sutcliffe Efficiency Coefficient Statistics			
Best Calibration Set	Average Value	Maximum Value (Max)	Minimum Value (Min)
S3'	0.69	0.85	0.33
S4'	0.75	0.90	0.61
S5'	0.73	0.89	0.44
S6'	0.70	0.88	0.49
S7'	0.64	0.74	0.33
S8'	0.66	0.80	0.32
S9'	0.64	0.73	0.33
S10'	0.79	0.91	0.67
S11'	0.54	0.80	0.22
S12'	0.51	0.80	0.21
S13'	0.49	0.79	0.22
S14'	0.47	0.79	0.20

#### 4.3. Consistency of the ICF Selection with the Model Performance

The index of Information Cost Function (ICF) was evaluated and investigated in choosing the best set of calibration flood events in different sequences for flood simulation. Since the ICF is an entropy-like function, it is based on the decomposition result of DWT. The ICF value of each Sequence was calculated to determine the degree of similarity between the calibration and validation sets. The validation ICF was chosen as the threshold to select the best calibration set from each Sequence. Based on the ICF values, the best set from the Sequence of 10 (set 4 = S10') was found to be the absolute best one, followed by the best from Sequence of 4 (set 4 = S4'), the best from Sequence of 5 (set 7 = S5'), the best from Sequence of 6 (set 4 = S6'), the best from Sequence of 3 (set 6 = S3'), the best from Sequence of 8 (set 4 = S8'), the best from Sequence of 7 (set 7 = S7'), the best from Sequence of 9 (set 5 = S9'), the best from Sequence of 11 (set 3 = S11'), the best from Sequence of 12 (set 2 = S12'), the best from Sequence of 13 (set 1 = S13'), and the best from Sequence of 14 (set 1 = S14'). The HEC–HMS model was calibrated and validated with the best calibration datasets to verify these analyses. Figure 8 shows the ICF values of the best calibration dataset from each Sequence versus the average model performances of the validation results.

The calibration results of the HEC–HMS model demonstrate the importance of selecting calibration flood events with the most appropriate amount of calibration data. The S10' calibration dataset with the closest ICF value to the validation dataset resulted in the highest model performance. The S14' calibration dataset with the most distant ICF value resulted in the poorest model performance. These are good findings, which imply that, by simply checking the ICF values of sets of data, we can decide which one to use to calibrate the model.



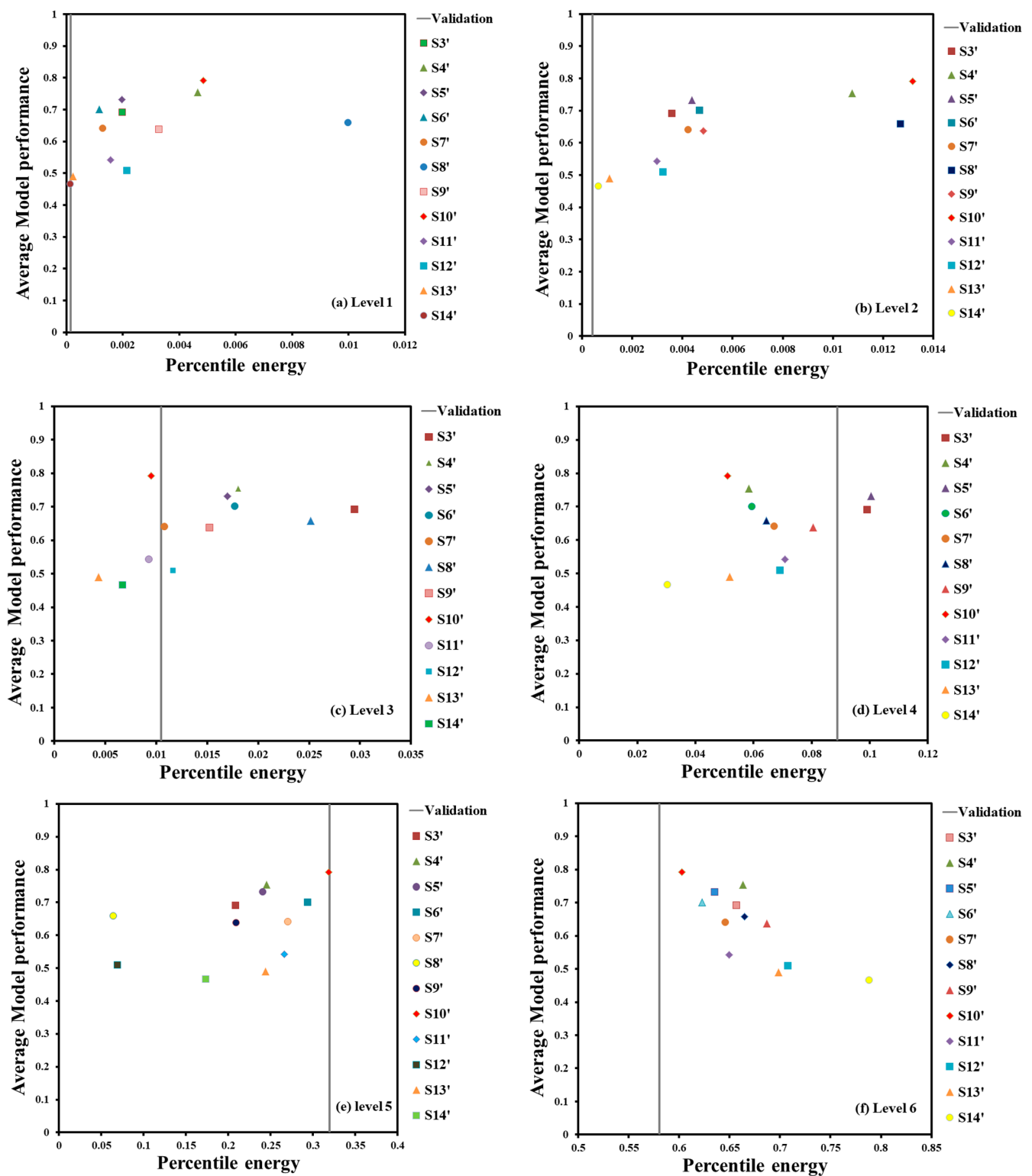


**Figure 8.** Relationship between the average model performance and the similarity of the information cost function (ICF).

#### 4.4. The Potential of Other Entropy-Based Indices

The percentile energy can also be a valuable metric to evaluate the spectral similarity between the calibration flood datasets and validation sets. The percentile energy measure provides information on energy distribution across different levels of the wavelet decomposition. It indicates the proportion of energy distributed at each decomposition level in the corresponding frequency domains (Equation (5)). After the Discrete Wavelet Transform (DWT) decomposition, the details, which contain high-frequency information, are considered to be more important than approximations because they capture the relish and more intricate characteristics of a signal. On the other hand, approximations contain low-frequency information that fundamentally indicates a signal identity. This study only presents results of the percentile energy of the details, since the approximation is a more abstract representation of the original signal as the wavelet decomposition progresses. In Figure 9, the average model performances are plotted against the percentile energy of the details for each of the decomposing levels of the best calibration flood datasets of the sequences.

It can be seen that the ranking of the best calibration sets based on the similarity of the percentile energy was not consistent with the order of the model performance on the decomposition levels of details (d1–d6), except for the S10' and the S14' datasets in decomposition levels 5 and 6 (Figure 9e–f). The S10' dataset with the best model performance was closest to the percentile energy of details of the validation set and the S14' dataset with the poorest model performance was most distant from the percentile energy of details of the validation set in levels 5 and 6, respectively.



**Figure 9.** Percentile energies of details on different wavelet decomposition levels of the best sets of calibration data sequences.

## 5. Discussion

The HEC-HMS model performance verifies the Information Cost Function (ICF) analysis, and these results reveal that the quality of the information in the calibration flood data, instead of the quantity, is more important in determining how well the model performs after calibration. As the ICF of each calibration dataset is calculated from the

Discrete Wavelet Transform (DWT) and compared with the validation sets' ICF to identify the similarities, the relationship between the model performances and ICF analyses justifies that a simple index such as ICF can help ease and improve calibration work. The potential of the percentile energy entropy index was also analyzed for the best calibration sets. The percentile energy results are not consistent, as the ICF index for all the best calibration sets in showing that model performance improved as similarity increased, especially in most decomposition levels which represented particular frequency domain ranges. Based on the results, the ICF index proves to be the most appropriate indicator for evaluating the similarity between the calibration flood datasets and the validation set. According to this perspective, the information in the "good" S10' dataset is of higher quality than the information in the other best Sequences with different numbers of flood events. One can deduce that, if the calibration set contains higher "quality of information", then there is a certain level of similarity between the calibration flood dataset and the validation set, given that the validation results of the calibrated models using seven flood events were chosen as the evaluation criteria.

More work is needed to establish the wavelet transform as an effective tool for hydrograph analysis. Allowing hydrologists to evaluate the "similarity" or the information of all the potential calibration flood datasets in specific sequences would greatly reduce the amount of calibration work needed to find the most suitable dataset. Except in circumstances where there is a poor performance with extremely high peak flood events, it is difficult to directly assess the quality of the information in the calibration data or to visually compare the similarity between the calibration flood datasets and the validation set. Therefore, it is essential to use an appropriate index to evaluate the similarity between the calibration flood datasets and the validation set. In research carried out by Liu and Han (2010) [68], they applied "the flow-duration curve," "the Fourier transform," and "the wavelet analysis and ICF" indices to analyze the similarities of the validation and the calibration sets and concluded that wavelet analysis with ICF was the most appropriate index. It should be noted that the ICF presented in this paper was calculated based on the detail coefficients, as the ICF from the approximations did not show much difference.

It should be mentioned here that other Daubechies wavelets (db2, db8, etc.) of a different order were also chosen for the decomposition. There was little difference in similarities found between the calibration flood dataset and validation set. Also, the wavelet and ICF index were implemented for the observed rainfall data and the observed flow in this research, but the observed rainfall data results were inconsistent because there was no rain during some periods of the flood events. Nevertheless, it should be noted that this might only be for the Tunxi catchment, so it might be interesting to investigate whether rainfall data could be useful for other study areas and case studies with varying calibration sequence designs.

## 6. Conclusions

Selecting calibration flood events is an essential task for hydrologists in the simulation of flood events using hydrological models. Although numerous studies have been conducted on flood simulation, there is a lack of information on how to select a suitable set of historical floods within a given study area for calibrating hydrologic models. The usual rule of choosing the number of flood events used for calibration depending on the catchment size is inadequate for catchments with different climatic or hydrological characteristics. Over time, hydrologists have gradually recognized that the quality of the information in the calibration flood dataset, instead of the quantity, is the most crucial factor influencing the performance of a hydrological model after calibration. The importance of selecting a sufficient amount of calibration data with a correct Sequence is increasing as telemetry systems continue to gather observed data with high resolution. This study has proposed a practical approach to selecting calibration data when using hydrological models for flood simulation. As the validation flood events have already been determined, it can be assumed that the accuracy of the calibrated model is directly proportional to the degree of similarity

between the calibration flood dataset and the validation set. The wavelet analysis and Information Cost Function (ICF) index were applied to describe the similarities between the calibration and validation sets of data. For the analysis of selecting the absolute best set of calibration flood data, it is impressive to note that the best dataset from the Sequence of 10 (set 4 = S10') in this study performed better than the other best sets from the calibration data Sequence with more and fewer flood events. This result corroborates the notion that the quality of the information in the calibration data is more important than the quantity. The potential of the percentile energy entropy was also analyzed for the best calibration sets, but the ICF was the most consistent index to reveal the ranking based on similarity with model performance. This study also demonstrated the possibility of improving the effectiveness of utilizing calibration data, particularly in catchments with limited data.

The findings presented in this paper are context-specific and depend on the specifics of the case study, including the catchment, the calibration sequence design, and the hydrological model employed. One potential limitation of this study is the predetermined set of validation flood events used. An important consideration in assessing how well a calibrated model performs is the choice of validation flood events. The practical selection of the validation flood events involves a combination of factors, including the magnitude and frequency of flood events, the availability and quality of streamflow data, and the representativeness of flood events across a range of hydrologic conditions in the study area. These factors can be appropriately evaluated and decided upon before selecting the calibration flood event set of data in standard cases. Therefore, we hope this study will encourage further research to investigate the index's effectiveness in diverse catchment conditions and with various flood events and hydrological models, particularly spatially distributed models with more complex input requirements.

**Author Contributions:** Conceptualization, S.U.J.-J., J.L. and Y.W.; methodology, S.U.J.-J. and J.L.; software, S.U.J.-J. and J.L.; validation, S.U.J.-J., J.L. and Y.W.; formal analysis, S.U.J.-J. and J.L.; investigation, S.U.J.-J. and J.L.; resources, S.U.J.-J. and J.L.; data curation, S.U.J.-J. and Z.L.; writing—original draft preparation, S.U.J.-J. and J.L.; writing—review and editing, S.U.J.-J., J.L. and N.-M.S.J.; visualization, S.U.J.-J.; supervision, J.L. and Y.W.; project administration, J.L. and Y.W.; funding acquisition, J.L. and Y.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (51822906).

**Data Availability Statement:** The hydrological data used in this study are provided by the State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing, and the College of Hydrology and Water Resources, Hohai University, Nanjing. Access to the 30 m digital elevation model (DEM) can be requested through the website: <http://www.gscloud.cn/sources/?cdatid=302&pdid=10>. (last access: 26 December 2022) (Geospatial Data cloud, 2022).

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

## References

1. Devitt, L.; Neal, J.; Coxon, G.; Savage, J.; Wagener, T. Flood hazard potential reveals global floodplain settlement patterns. *Nat. Commun.* **2023**, *14*, 2801. [CrossRef]
2. Sivakumar, H.K.B. Assessment of change in design flood frequency under climate change using a multivariate downscaling model and a precipitation-runoff model. *Stoch. Environ. Res. Risk Assess.* **2011**, *25*, 567–581. [CrossRef]
3. Messner, F.; Penning-rowsell, E.; Green, C.; Tunstall, S.; Van Der Veen, A.; Tapsell, S.; Wilson, T.; Krywkow, J.; Logtmeijer, C.; Fernández-bilbao, A.; et al. Evaluating flood damages: Guidance and recommendations on principles and methods. *Risk Manag. Hazards Vulnerabil. Mitig. Meas.* **2007**, 1–189. Available online: [https://floodsite.net/html/partner\\_area/project\\_doccd/T09\\_06\)01\\_Flood\\_damage\\_guidelines](https://floodsite.net/html/partner_area/project_doccd/T09_06)01_Flood_damage_guidelines) (accessed on 26 December 2022).
4. Eleutério, J. Flood Risk Analysis: Impact of Uncertainty in Hazard Modelling and Vulnerability Assessments on Damage Estimations. Ph.D. Thesis, University of Strasbourg, Strasbourg, France, 2013.
5. Yu, H.; Hatzivassiloglou, V.; Rzhetsky, A.; Wilbur, W.J. Automatically identifying gene/protein terms in MEDLINE abstracts. *J. Biomed. Inform.* **2002**, *35*, 322–330. [CrossRef]
6. Romali, N.S.; Yusop, Z.; Sulaiman, M.; Ismail, Z. Flood risk assessment: A review of flood damage estimation model for Malaysia. *J. Teknol.* **2018**, *80*, 145–153. [CrossRef]

7. Hao, F.; Sun, M.; Geng, X.; Huang, W.; Ouyang, W. Coupling the Xinanjiang model with geomorphologic instantaneous unit hydrograph for flood forecasting in northeast China. *Int. Soil Water Conserv. Res.* **2015**, *3*, 66–76. [\[CrossRef\]](#)
8. Bouadila, A.; Bouizrou, I.; Aqnouy, M.; En-nagre, K.; El Yousfi, Y.; Khafouri, A.; Hilal, I.; Abdelrahman, K.; Benaabidate, L.; Abu-Alam, T.; et al. Streamflow Simulation in Semiarid Data-Scarce Regions: A Comparative Study of Distributed and Lumped Models at Aguenza Watershed (Morocco). *Water* **2023**, *15*, 1602. [\[CrossRef\]](#)
9. Ali, M.H.; Popescu, I.; Jonoski, A.; Solomatine, D.P. Remote Sensed and/or Global Datasets for Distributed Hydrological Modelling: A Review. *Remote Sens.* **2023**, *15*, 1642. [\[CrossRef\]](#)
10. 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. [\[CrossRef\]](#)
11. Huang, Y.; Bárdossy, A. Impacts of Data Quantity and Quality on Model Calibration: Implications for Model Parameterization in Data-Scarce Catchments. *Water* **2020**, *12*, 2352. [\[CrossRef\]](#)
12. Jodhani, K.H.; Patel, D.; Madhavan, N. A review on analysis of flood modelling using different numerical models. *Mater. Today Proc.* **2023**, *80*, 3867–3876. [\[CrossRef\]](#)
13. Dooge, J.C.I. Frequency analysis of hydrologic data for design of drainage structures. *Water Resour. Res.* **1969**, *5*, 1273–1290. [\[CrossRef\]](#)
14. Bruen, M.; DOOGE, J. Unit hydrograph estimation with multiple events and prior information: II. Evaluation of the method. *Hydrol. Sci. J.* **1992**, *37*, 445–462. [\[CrossRef\]](#)
15. Vieux, B.E.; Cui, Z.; Gaur, A. Evaluation of a physics-based distributed hydrologic model for flood forecasting. *J. Hydrol.* **2004**, *298*, 155–177. [\[CrossRef\]](#)
16. Garambois, P.A.; Roux, H.; Larnier, K.; Labat, D.; Dartus, D. Characterization of catchment behaviour and rainfall selection for flash flood hydrological model calibration: Catchments of the eastern Pyrenees. *Hydrol. Sci. J.* **2015**, *60*, 424–447. [\[CrossRef\]](#)
17. Shafii, M.; Tolson, B.A. Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. *Water Resour. Res.* **2015**, *51*, 3796–3814. [\[CrossRef\]](#)
18. 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. [\[CrossRef\]](#)
19. Reynolds, E.; Halldin, S.; Seibert, J.; Xu, C.-Y.; Grabs, T. Flood prediction using parameters calibrated on limited discharge data and uncertain rainfall scenarios. *Hydrol. Sci. J.* **2020**, *65*, 1512–1524. [\[CrossRef\]](#)
20. Gupta, V.K.; Sorooshian, S. The relationship between data and the precision of estimated parameters. *J. Hydrol.* **1985**, *81*, 55–77. [\[CrossRef\]](#)
21. Gupta, V.K.; Sorooshian, S. The Automatic Calibration of Conceptual Catchment Models Using Derivative-Based Optimization Algorithms. *Water Resour. Res.* **1985**, *21*, 473–485. [\[CrossRef\]](#)
22. Smith, L.C.; Turcotte, D.L.; Isacks, B.L. Stream flow characterization and feature detection using a discrete wavelet transform. *Hydrol. Process.* **1998**, *12*, 233–249. [\[CrossRef\]](#)
23. Partal, T.; Küçük, M. Long-term trend analysis using discrete wavelet components of annual precipitations measurements in Marmara region (Turkey). *Phys. Chem. Earth Parts A/B/C* **2006**, *31*, 1189–1200. [\[CrossRef\]](#)
24. Beecham, S.; Chowdhury, R.K. Temporal characteristics and variability of point rainfall: A statistical and wavelet analysis. *Int. J. Climatol.* **2010**, *30*, 458–473. [\[CrossRef\]](#)
25. Li, M.; Xia, J.; Chen, Z.; Meng, D.; Xu, C. Variation analysis of precipitation during past 286 years in Beijing area, China, using non-parametric test and wavelet analysis. *Hydrol. Process.* **2013**, *27*, 2934–2943. [\[CrossRef\]](#)
26. Zhou, Z.; Shi, H.; Fu, Q.; Ding, Y.; Li, T.; Wang, Y.; Liu, S. Characteristics of Propagation From Meteorological Drought to Hydrological Drought in the Pearl River Basin. *J. Geophys. Res. Atmos.* **2021**, *126*, e2020JD033959. [\[CrossRef\]](#)
27. Das, J.; Mandal, T.; Rahman, A.T.M.S.; Saha, P. Spatio-temporal characterization of rainfall in Bangladesh: An innovative trend and discrete wavelet transformation approaches. *Theor. Appl. Climatol.* **2021**, *143*, 1557–1579. [\[CrossRef\]](#)
28. Yang, T.; Wang, G. Periodic variations of rainfall, groundwater level and dissolved radon from the perspective of wavelet analysis: A case study in Tengchong, southwest China. *Environ. Earth Sci.* **2021**, *80*, 492. [\[CrossRef\]](#)
29. Yue, Y.; Liu, H.; Mu, X.; Qin, M.; Wang, T.; Wang, Q.; Yan, Y. Spatial and temporal characteristics of drought and its correlation with climate indices in Northeast China. *PLoS ONE* **2021**, *16*, e0259774. [\[CrossRef\]](#)
30. Zerouali, B.; Chettih, M.; Abda, Z.; Mesbah, M.; Santos, C.A.G.; Brasil Neto, R.M. A new regionalization of rainfall patterns based on wavelet transform information and hierarchical cluster analysis in northeastern Algeria. *Theor. Appl. Climatol.* **2022**, *147*, 1489–1510. [\[CrossRef\]](#)
31. Wu, L.; Wang, S.; Bai, X.; Chen, F.; Li, C.; Ran, C.; Zhang, S. Identifying the Multi-Scale Influences of Climate Factors on Runoff Changes in a Typical Karst Watershed Using Wavelet Analysis. *Land* **2022**, *11*, 1284. [\[CrossRef\]](#)
32. Kumar, P.; Foufoula-Georgiou, E. Wavelet analysis for geophysical applications. *Rev. Geophys.* **1997**, *35*, 385–412. [\[CrossRef\]](#)
33. Labat, D. Recent advances in wavelet analyses: Part 1. A review of concepts. *J. Hydrol.* **2005**, *314*, 275–288. [\[CrossRef\]](#)
34. Rouyer, T.; Fromentin, J.-M.; Stenseth, N.C.; Cazelles, B. Analysing multiple time series and extending significance testing in wavelet analysis. *Mar. Ecol. Prog. Ser.* **2008**, *359*, 11–23. [\[CrossRef\]](#)
35. Li, C.H.; Yang, Z.F.; Huang, G.H.; Li, Y.P. Identification of relationship between sunspots and natural runoff in the Yellow River based on discrete wavelet analysis. *Expert Syst. Appl.* **2009**, *36*, 3309–3318. [\[CrossRef\]](#)



36. Krishna, B.; Satyaji Rao, Y.; Nayak, P.C. Time Series Modeling of River Flow Using Wavelet Neural Networks. *J. Water Resour. Prot.* **2011**, *3*, 50–59. [\[CrossRef\]](#)
37. Arora, B.; Dwivedi, D.; Hubbard, S.S.; Steefel, C.I.; Williams, K.H. Identifying geochemical hot moments and their controls on a contaminated river floodplain system using wavelet and entropy approaches. *Environ. Model. Softw.* **2016**, *85*, 27–41. [\[CrossRef\]](#)
38. Kumarasamy, K.; Belmont, P. Calibration Parameter Selection and Watershed Hydrology Model Evaluation in Time and Frequency Domains. *Water* **2018**, *10*, 710. [\[CrossRef\]](#)
39. Duran, L.; Massei, N.; Lecoq, N.; Fournier, M.; Labat, D. Analyzing multi-scale hydrodynamic processes in karst with a coupled conceptual modeling and signal decomposition approach. *J. Hydrol.* **2020**, *583*, 124625. [\[CrossRef\]](#)
40. Chong, K.L.; Huang, Y.F.; Koo, C.H.; Najah Ahmed, A.; El-Shafie, A. Spatiotemporal variability analysis of standardized precipitation indexed droughts using wavelet transform. *J. Hydrol.* **2022**, *605*, 127299. [\[CrossRef\]](#)
41. Mares, I.; Mares, C.; Dobrica, V.; Demetrescu, C. Selection of Optimal Palmer Predictors for Increasing the Predictability of the Danube Discharge: New Findings Based on Information Theory and Partial Wavelet Coherence Analysis. *Entropy* **2022**, *24*, 1375. [\[CrossRef\]](#)
42. Kisi, O.; Cimen, M. Precipitation forecasting by using wavelet-support vector machine conjunction model. *Eng. Appl. Artif. Intell.* **2012**, *25*, 783–792. [\[CrossRef\]](#)
43. Nayak, P.C.; Venkatesh, B.; Krishna, B.; Jain, S.K. Rainfall-runoff modeling using conceptual, data driven, and wavelet based computing approach. *J. Hydrol.* **2013**, *493*, 57–67. [\[CrossRef\]](#)
44. Zhou, F.; Liu, B.; Duan, K. Coupling wavelet transform and artificial neural network for forecasting estuarine salinity. *J. Hydrol.* **2020**, *588*, 125127. [\[CrossRef\]](#)
45. Kisi, O.; Shiri, J. Precipitation Forecasting Using Wavelet-Genetic Programming and Wavelet-Neuro-Fuzzy Conjunction Models. *Water Resour. Manag.* **2011**, *25*, 3135–3152. [\[CrossRef\]](#)
46. Partal, T. Wavelet analysis and multi-scale characteristics of the runoff and precipitation series of the Aegean region (Turkey). *Int. J. Climatol.* **2012**, *32*, 108–120. [\[CrossRef\]](#)
47. Nalley, D.; Adamowski, J.; Khalil, B. Using discrete wavelet transforms to analyze trends in streamflow and precipitation in Quebec and Ontario (1954–2008). *J. Hydrol.* **2012**, *475*, 204–228. [\[CrossRef\]](#)
48. Özgen-Xian, I.; Kesserwani, G.; Caviedes-Voullième, D.; Molins, S.; Xu, Z.; Dwivedi, D.; Moulton, J.D.; Steefel, C.I. Wavelet-based local mesh refinement for rainfall-runoff simulations. *J. Hydroinform.* **2020**, *22*, 1059–1077. [\[CrossRef\]](#)
49. Liu, Q.; Dai, H.; Gui, D.; Hu, B.X.; Ye, M.; Wei, G.; Qin, J.; Zhang, J. Evaluation and optimization of the water diversion system of ecohydrological restoration megaproject of Tarim River, China, through wavelet analysis and a neural network. *J. Hydrol.* **2022**, *608*, 127586. [\[CrossRef\]](#)
50. Wang, D.; Dong, Z.; Jiang, F.; Zhu, S.; Ling, Z.; Ma, J. Spatiotemporal variability of drought/flood and its teleconnection with large-scale climate indices based on standard precipitation index: A case study of Taihu Basin, China. *Environ. Sci. Pollut. Res.* **2022**, *29*, 50117–50134. [\[CrossRef\]](#)
51. Abebe, S.A.; Qin, T.; Zhang, X.; Yan, D. Wavelet transform-based trend analysis of streamflow and precipitation in Upper Blue Nile River basin. *J. Hydrol. Reg. Stud.* **2022**, *44*, 101251. [\[CrossRef\]](#)
52. Samani, S.; Vadiati, M.; Nejatjahromi, Z.; Etebari, B.; Kisi, O. Groundwater level response identification by hybrid wavelet-machine learning conjunction models using meteorological data. *Environ. Sci. Pollut. Res. Int.* **2023**, *30*, 22863–22884. [\[CrossRef\]](#)
53. Samani, S.; Vadiati, M.; Delkash, M.; Bonakdari, H. A hybrid wavelet-machine learning model for qanat water flow prediction. *Acta Geophys.* **2022**. [\[CrossRef\]](#)
54. Daubechies, I. The wavelet transform, time-frequency localization and signal analysis. *IEEE Trans. Inf. Theory* **1990**, *36*, 961–1005. [\[CrossRef\]](#)
55. Polikar, R. *The Story of Wavelets*; World Scientific and Engineering Academy and Society: Sitia, Greece, 1999; ISBN 9608052106.
56. Zhang, H.; Zhang, S.; Wang, P.; Qin, Y.; Wang, H. Forecasting of particulate matter time series using wavelet analysis and wavelet-ARMA/ARIMA model in Taiyuan, China. *J. Air Waste Manag. Assoc.* **2017**, *67*, 776–788. [\[CrossRef\]](#)
57. Meyer, Y. *Wavelets: Algorithms & Applications*; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 1993; ISBN 0-8971-309-9.
58. Mallat, S.G. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1989**, *11*, 674–693. [\[CrossRef\]](#)
59. Li, X.; Li, H.; Wang, F.; Ding, J. A remark on the mallat pyramidal algorithm of wavelet analysis wavelet analysis. *Commun. Nonlinear Sci. Numer. Simul.* **1997**, *2*, 240–243. [\[CrossRef\]](#)
60. Blanco, S.; Figliola, A.; Quiroga, R.Q.; Rosso, O.A.; Serrano, E. Time-frequency analysis of electroencephalogram series. III. Wavelet packets and information cost function. *Phys. Rev. E* **1998**, *57*, 932–940. [\[CrossRef\]](#)
61. Figliola, A.; Serrano, E. Analysis of physiological time series using wavelet transforms. *IEEE Eng. Med. Biol. Mag.* **1997**, *16*, 74–79. [\[CrossRef\]](#) [\[PubMed\]](#)
62. Bartles, M.; Brauer, T.; Ho, D.; Fleming, M.; Karlovits, G.; Pak, J.; Van, N.; Willis, J.O. Hydrologic Modeling System HEC-HMS User's Manual. Available online: <https://www.hec.usace.army.mil/confluence/hmsdocs/hmum/latest> (accessed on 19 January 2022).
63. Ara, Z.; Zakwan, M. Rainfall Runoff Modelling for Eastern Canal Basin. *Water Energy Int.* **2018**, *61*, 63–67.



64. Niazkar, M.; Zakwan, M. Parameter estimation of a new four-parameter Muskingum flood routing model. In *Computers in Earth and Environmental Sciences*; Elsevier: Amsterdam, The Netherlands, 2022; pp. 337–349. ISBN 9780323898614.
65. Nash, J.E.; Sutcliffe, J. V River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290. [[CrossRef](#)]
66. Quinn, P.; Beven, K.; Chevallier, P.; Planchon, O. The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models. *Hydrol. Process.* **1991**, *5*, 59–79. [[CrossRef](#)]
67. Yapo, P.O.; Gupta, H.V.; Sorooshian, S. Automatic calibration of conceptual rainfall-runoff models: Sensitivity to calibration data. *J. Hydrol.* **1995**, *181*, 23–48. [[CrossRef](#)]
68. Liu, J.; Han, D. Indices for Calibration Data Selection of the Rainfall-Runoff Model. *Water Resour. Res.* **2010**, *46*, W04512. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.