Evaluating Hydrological Models for Deriving Water Resources in Peninsular Spain

**Julio Pérez-Sánchez [1],* , Javier Senent-Aparicio [1] , Francisco Segura-Méndez [1],
David Pulido-Velazquez [1] and Raghavan Srinivasan [2]**

[1] Department of Civil Engineering, Catholic University of San Antonio, Campus de los Jerónimos, s/n, Guadalupe, 30107 Murcia, Spain; jsenent@ucam.edu (J.S.-A.); fjsegura7@gmail.com (F.S.-M.); d.pulido@igme.es (D.P.-V.)

[2] Spatial Science Laboratory, Ecosystem Science and Management Department, Texas A & M University, College Station, TX 77843, USA; r-srinivasan@tamu.edu

* Correspondence: jperez058@ucam.edu

check for updates

Abstract: Water availability is essential for the appropriate analysis of its sustainable management. We performed a comparative study of six hydrological balance models (Témez, ABCD, GR2M, AWBM, GUO-5p, and Thornthwaite-Mather) in several basins with different climatic conditions within Spain in the 1977–2010 period. We applied six statistical indices to compare the results of the models: the Akaike information criterion (AIC), the Bayesian information criterion (BIC), Nash–Sutcliffe model efficiency coefficient (NSE), coefficient of determination (R^2), percent bias (PBIAS), and the relative error between observed and simulated run-off volumes (REV). Furthermore, we applied the FITEVAL software to determine the uncertainty of the model. The results show that when the catchments are more humid the obtained results are better. The GR2M model gave the best fit in peninsular Spain in a UNEP aridity index framework above 1, and NSE values above 0.75 in a 95% confidence interval classify GR2M as very good for humid watersheds. The use of REV is also a key index in the assessment of the margin of error. Flow duration curves show good performance in the probabilities of exceedance lower than 80% in wet watersheds and deviations in low streamflows account for less than 5% of the total streamflow.

1. Introduction

Water resources assessment is key to the analysis of catchment management [1]. The cost and irregular distribution of worldwide water resources are evident. Even in parts of the planet where water resources are abundant, the problems of availability or scarcity are common due to, largely, weak water management practices and anthropic activities [2]. The development of models to study water availability is a complex task that presents a fundamental scientific challenge. Complex cases, in particular, occur in arid and semi-arid regions, where precipitation is limited or irregular and evapotranspiration (ET) rates are high. Hydrological balance models are used to reconstruct historical series and predict future ones [3]. They are based on the principle of mass conservation or the continuity equation [4], which considers that the difference of inputs and outputs will be reflected in water storage in the catchment [5,6].

The concept of hydrological balance models was first introduced by Thornthwaite [7] and Thornthwaite and Mather [8]. They proposed two different conceptual models based on two parameters: soil moisture capacity and water excess above the maximum soil moisture storage capacity. These models

demonstrated a good fit to estimate monthly run-off [9] and have formed the basis of many other two-parameter hydrological models [9–14]. There are also water balance models that comprise more than two parameters [15–17]. However, Xiong and Guo [10] showed that their proposed two-parameter model in China performed as well as a five-parameter model. To date, several studies have shown that many models produce similar results to previous ones [18–21]. In a lumped water balance model, catchment parameters and variables are averaged in space, so hydrological processes are approached by conceptual solutions formulated using semi-empirical equations. The system is described using different reservoirs, the moisture content of which depends on the relationships (physical and empirical) between them [22]. A lumped hydrological balance model may have only three or four parameters [9,22,23] and can be implemented with several lines of computer code, whereas a complex model may have more than 20 parameters [24]. Some examples of lumped models are the ABCD model [14,25–27], GR2M [28,29], Sacramento [30], Guo-5p [10,31], Témez [32–36], Thornwaite-Mather [37–39], IHACRES [40], SIMHYD [41], GR4J [42], AWBM [43–47], and SMAR [48]. More examples of rainfall run-off models can be found in Singh [49] and Singh and Frevert [50].

With the development of computer aided tools and more detailed information, there is an increasing trend toward using distributed or semi-distributed models [51,52]. They provide more detailed distributed results on a catchment scale approximating heterogeneities of the system. Following these assumptions, many complex models have been developed that are assumed to be capable of simulating environmental change. These spatially explicit and physically-based model approaches are often criticized because the necessary a priori estimation of model parameters is difficult [53,54], and uncertainty at high resolution may diminish potential gains in prediction accuracy [55]. Thus, efficient calibration of the models is difficult due to the spatially distributed nature of those models, showing, sometimes, a decrease in efficiency [56], and lumped models can provide a more appropriate alternative [57]. Moreover, lumped models do not need as much data as distributed models, and the complexity and requirements to process them are lower. The calibration of the lumped parameter models is much less time consuming and produces higher overall model performance in comparison to the more complex distributed models [58]. Notwithstanding, recent studies in distributed hydrologic models have developed methodology approaches that improve the limitations mentioned above. Samaniego et al. [59] used a multiscale parameter regionalization (MPR) technique in a fully spatially distributed conceptual hydrologic model, resulting in the limitation of a number of model parameters and the transferability of the global parameters at a coarser scale to a finer scale, meaning that there is a substantial shortcoming in the calibration procedure. Beck et al. [60] developed a scheme for the regionalization of model parameters at the global scale, resulting in HBV parameter maps, such as ancillary data that are available via www.gloh2o.org.

Notwithstanding, despite the simplicity of lumped models, they performed well in many studies [61–64]. Other studies that have compared lumped and distributed models confirmed that both of them lead to similar accuracy [65–75] and that lumped models can be calibrated more efficiently [76]. Vansteenkiste et al. [77] learned that in a Belgian catchment, the lumped models perform better than the distributed ones in seasonal events and in terms of overall water balance, with very low discrepancies in the subflow volumes in comparison to the distributed models. Furthermore, lumped models provide a valuable integrated view of the basin outlet response, as concluded by the Distributed Model Intercomparison Project for the Oklahoma region by Reed et al. [72] and Smith et al. [66]. Martínez-Santos and Andreu [78] used lumped and distributed approaches to model natural recharge in semi-arid aquifers in Spain, and even though both approaches performed similarly, the lumped models exhibited a better agreement with field records.

Therefore, spatial discretization is not the only determinant of the simulation's quality. The choice of model is dictated by the modelling purpose. When flow at the catchment outlet is the main required goal in water resources management, lumped models may be the best choice, but when spatially explicit predictions or land use change predictions are required, a (semi-)distributed model would be more appropriate, although land use changes can also be carried out in a lumped model by means of transfer

functions [68]. Recently, lumped models have been used, among other purposes, to obtain detailed assessments of surface flow, water balance components, and the impact of climate change [79,80] to estimate catchment discharge [81–83], and to explore transferability under contrasting climate conditions [84,85].

In this study, choosing the models is mainly based on known performance in different climatic regions, both in Spain and Europe, and structural diversity (i.e., 3–6 free parameters and 2–4 storage units). However, an inadequate complexity often results in over-parameterization [19,86,87], so the models with too many parameters (more than 6) are excluded from this study. Thus, the Témez model has been widely used in Spanish catchments [88–91] and by the Spanish government in water management [92]. The widely used ABCD model has proven to have reasonable predictability and has been compared with numerous monthly water balance models, thus leading to its recommendation [27,93–96]. Vandewiele and Xu [17] also found that the ABCD model compares favorably with other more recent monthly water balance models in Belgium. Currently, the ABCD model has been found to have the best performance in the Segura River basin in southern Spain [97]. The GR2M model is widely used in France in all of its different versions [98]. Wriedtand and Bouraoiu [99] used this model in 492 catchments in Germany, France, Spain, and Portugal and obtained high NSE in the center and north of Spain and in Central European basins. The AWBM is one of the most widely used rainfall run-off models in Australia [46], but it has been used in different global catchments [25,100,101], both humid and dry basins, and it has proven to have performance similar to the distributed models [102]. The Thornthwaite and Mather water balance model is valid as a water accounting procedure when only reduced information about hydrologic inputs and aquifer characteristics is available, as applied by Peranginangin et al. [103]. It has been successfully used in different water balance research studies in Spain [104–107]. Finally, Guo-5p is an adaptation of Thornthwaite and Mather's model with five parameters, so it was chosen to compare with the latter. Its use is particularly recommended in humid and semi-humid regions [10,31].

Our primary objective in this study is to use model predictions for the assessment of water resources, as the first step in predicting future resources under climate change. This should predict and solve problems related to the quantity and quality of water, as well as ensure its sustainable use. Furthermore, the following must also be considered: the population growth rate, increasing demand, depleted fluvial networks, the requirements concerning ecological flow, and the reduction in the amount of groundwater. Thus, efficient water management becomes an essential issue due to global scarcity. Our work is mainly focused on the comparative analysis of hydrological models and their ability to predict and quantify outflows on the basis of different watersheds in Spain, where the wide range of climates is optimal for such hydrological research.

The first step, which is developed in this research, addresses the issue of comparing and selecting, using different metrics and graphics, the water balance model that has the best fit according to our goal: water resources assessment. Although, as previously stated, there are many studies that compare different hydrological models, we have used a full set of statistical methods in order, not only to verify the feasibility and the performance of these models in different climatic areas, but also to assess the usefulness of the combination of different goodness-of-fit tools depending on the purpose of the study. Therefore, six lumped water balance models were evaluated in 16 basins located in different climatic regions of Spain, where a long-time series of climatic and natural streamflow data are available (more than 30 years). We intended to calibrate parsimonious approaches to estimate natural streamflow series, and we intended to generate a series that represents the stochastic variability of the rainfall process. These series can be used as inputs of management models to assess the operation of water resource systems. In these cases, we need long time series of inflows (natural streamflows) to account for the influence of the hydrology's stochastic behaviors in the reliabilities of the demand supply [108,109]. Our aim is to select the models with the best fit and to assess the comparison methods used according to the region's characteristics. We structured the paper's contents as follows. Section 2 introduces the

data and methodology; Section 3 presents the results and the discussion; and Section 4 highlights the main conclusions.

2. Materials and Methods

2.1. Study Area and Data

Spain is the second largest country in Western Europe, with a territory covering 505,990 km^2, a dense network of rivers with many branches and a large number of aquifers. This disparity within the Iberian Peninsula is optimal for hydrological research. Spain features a wide range of climates due to its position between the European temperate zone and the subtropical zone. It also includes some of the rainiest areas in Europe in the northeast and the driest areas in the southeast, with a marked summer drought. To ensure the validity of the results, the 16 selected catchments are in the natural regime. They are located all over the country, as shown in Figure 1. Thus, the geographic and climatic variety in Spain are reflected wherever data were available.

Their altitudes vary from 1632 to 342 m above sea level (MASL), and catchment areas range from 29–837 km^2 with an average of 300 km^2 (Table 1). There is no southwestern catchment in this study due to the lack of data in this area, where most gauging stations have data for a period of less than 10 years. As shown in Table 1, the average temperatures range from 8–16 °C, depending on the latitude and average altitude of the catchment, with a positive gradient to the south. With regards to the rainfall regime, the highest yearly precipitation occurs in the north of Spain, where average temperatures are lower, and consequently, ET is less. However, in the southern half of the Iberian Peninsula, ET is generally higher than precipitation, especially in the lowlands, which have an average altitude of less than 600 m.
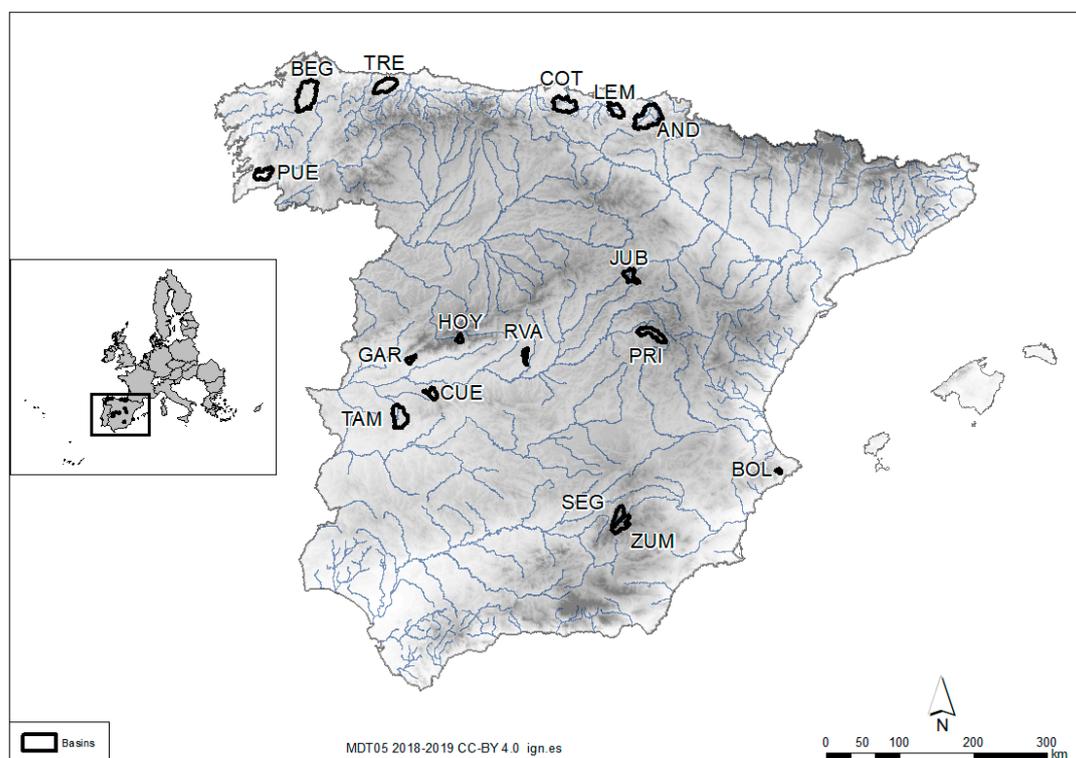
Figure 1. Position and location of catchments (abbreviations refer to catchments codes).

The study area comprises the most common climate groups in the Iberian Peninsula, according to Köppen's [110–112] classification: Bsk. (Cold semi-arid), Csb. (Warm-summer Mediterranean),

Csa. (Hot-summer Mediterranean climate), Cfb. (Temperate oceanic), and UNEP aridity index [113]. This index (AI$_U$) is defined by Equation (1).

$$AI_U = \frac{P}{PET} \tag{1}$$

where P is the average annual precipitation and PET is the potential evapotranspiration. The thresholds that define the various degrees of aridity depend on the value of AI$_U$: hyper-arid (AI$_U$ < 0.05), arid (0.05 < AI$_U$ < 0.20), semi-arid (0.20 < AI$_U$ < 0.50), dry sub-humid (0.50 < AI$_U$ < 0.65), humid sub-humid (1.00 < AI$_U$ < 0.65), and humid (AI$_U$ > 1.00). Among the 16 catchments studied, nine are considered humid; thus, their AIU exceeds one, and five of them even have an aridity index near to or above 2. All are located in northern Spain. Despite the rainfall gradient from the northwest to the southeast, the SEG and ZUM catchments are classified as humid sub-humid due to their altitude above 1000 MASL. The BOL catchment, which is close to the Mediterranean Sea, is the only eastern dry sub-humid region of the regions studied; thus, the other dry sub-humid catchments (TAM and CUE) are located in the center of the Iberian Peninsula. The only semi-arid region in the studied catchments is RVA, the average altitude of which is approximately 600 MASL.

We carried out an analysis of land cover variations in the period 1990–2012 in the watersheds (Table 1) in order to consider their influence in models both in the calibration and the validation. Notwithstanding, nearly all watersheds show a variation below 5%, which will not affect the main conclusions. However, RVA and LEM land cover have changed 9.18% and 19.12%, respectively, in the available data period, and will be further discussed.

Precipitation and PET data series in each basin are from a 33-year period (1977–2010). We obtained the data from the official monthly series provided by the CEDEX (Centre of Studies and Experimentation of Civil Works) for the Spanish government [114] at a spatial resolution of 500 × 500 m^2 [115]. The model has been validated at more than 100 control points [91] and used in Spain for water resources assessment, in the White Paper Book of Waters [89], and in several studies [116–118]. Natural streamflow data in each catchment are available for the same period. They come from measurements at gauging stations in the official Spanish network. Missing streamflow values range from 2–8% in the stations considered. We obtained digital elevation models and land cover variations [119] from the National Geographical Institute of Spain [120]. Furthermore, in order to assess hierarchical order of the main river courses of the basins, the Pfastetter code is considered. As shown in Table 2, the range of the code of the basins studied varies between 913685 and 104080, indicating a wide variety of topological location within the watershed.

Table 1. Summary of catchments characteristics (1977–2010) (Pfafstetter Code is the river Pfafstetter code; X and Y coordinates refer to the centroid of the basin; CLC, Corine Land Cover, 1990–2012 period).

Code	Name	Area (km²)	Pfafstetter Code	X ETRS89 UTM 30N	Y ETRS89 UTM 30N	MASL (m)	Köppen Class.	UNEP Aridity Index	Average Temperature (°C)	Average Yearly Precipitation (mm)	Average Yearly ETP (mm)	CLC Variation (%)
AND	Andoain	778.49	172988	573,531.02	4,769,234.41	486.01	Cfb	2.15	11.62	1563.47	727.86	5.23
BEG	Begonte	836.89	104080	111,096.28	4,799,111.26	504.01	Csb	2.07	11.44	1332.62	632.50	1.89
BOL	Bolulla	29.23	806014	749,970.13	4,286,756.78	600.31	Csa	0.54	16.56	579.71	1080.30	0.00
COT	Coterillo	488.22	184926	460,081.35	4,786,345.03	559.51	Cfb	1.65	11.48	1311.12	793.16	2.47
CUE	Cuernacabras	139.86	308769	280,997.89	4,392,588.21	610.63	Csa	0.52	15.33	570.56	1098.46	1.32
GAR	Gargüera	69.92	301389	251,972.85	4,439,204.60	689.98	Csa	1.02	14.73	1060.18	1043.32	5.31
HOY	Hoyos	66.15	211914	318,316.11	4,466,995.19	1632.12	Csb	1.01	8.58	777.22	770.35	0.87
JUB	Jubera	207.66	913685	549,719.20	4,553,001.74	1150.05	Csb	0.65	10.84	509.76	783.22	0.12
LEM	Lemona	252.58	172552	530,214.27	4,779,163.90	342.18	Cfb	1.96	12.35	1393.18	709.20	19.12
PRI	Priego	328.16	309656	578,643.32	4,473,678.12	1255.05	Csb	1.19	10.96	763.04	642.86	0.02
PUE	Puenteareas	263.85	104192	52,975.07	4,692,077.36	400.05	Csb	2.20	14.07	1662.05	756.42	5.48
RVA	Vallehermoso	85.68	304523	407,367.20	4,444,846.50	607.70	Bsk	0.40	14.69	396.53	998.49	9.18
SEG	Segura	232.89	702180	533,459.57	4,225,568.22	1416.46	Csb	0.88	11.53	807.66	915.83	0.94
TAM	Tamuja	458.12	309755	237,068.72	4360580.97	447.46	Csa	0.54	15.93	596.36	1112.84	0.02
TRE	Trevias	413.54	185314	217,492.08	4,812,225.19	526.64	Cfb	1.84	12.31	1220.91	663.77	5.00
ZUM	Zumeta	266.03	702184	536,758.71	4,213,732.23	1549.95	Csb	0.79	11.35	750.28	951.88	0.04

2.2. Methodology

The methods we used in this investigation are based on the identification of the hydrological models that best fit each catchment considered. Six monthly water balance models were used: Témez, ABCD, GR2M-1994, Australian Water Balance Model (AVBM), Guo-5p, and Thornwaite-Mather. All these models use precipitation and PET as input data. We assessed the goodness of fit for the six models using various model selection criteria for the purpose of establishing a robust methodology that can validate the conclusions obtained and its application in other different regions. Accepting the fact that models only approximate reality, our objective was to determine which of the candidate models best approximates the data. Because the field of information theory is used to quantify or measure the expected value of the information, we used the information theoretic approach to derive the two most-commonly-used criteria in model selection: the Akaike information criterion (AIC) [121,122] and the Bayesian information criterion (BIC) [123]. The difference between the BIC and the AIC is the greater penalty imposed for the number of parameters by the former than the latter. Alternatively, and based on Moriasi et al. [124] and Bressiani et al. [125], we established a grading method to evaluate the good performance of the model based on the NSE, R^2, and PBIAS values. We also analyzed the relative error between observed and simulated run-off volumes and the flow duration curves (FDCs) in the studied period to assess both low-flows and high-flows. Finally, we carried out the uncertainty of the model selected in each watershed with the FITEVAL software [126]. All of these hydrological models are defined with four parameters, except Thornwaite-Mather (three parameters) and Guo-5p (five parameters). Through all phases of the hydrological cycle, the models conduct different moisture balances according to the different processes in a hydrological system. The processes are governed by the continuity principle and mass balance and remain regulated by the specific laws of division and the transfer between the reservoirs of each model [90].

2.2.1. Water Balance Models

Témez Model

Témez [127] developed the Témez model, which has been widely used in Spanish catchments [88–91] and by the Spanish government in water management [92]. This model considers the system to be divided into two zones (Figure 2): the upper or non-saturated zone (S) and the lower or saturated zone (G). Some of the precipitation (P) drains directly into the river or through the aquifer, while the remainder is converted into ET. Excess is divided into run-off (Qs) through river networks at the present time, and infiltration to aquifers, draining one part (Qg) at the present time, with the rest remaining in the groundwater storage tank (G) for drainage at a later date (Figure 2).
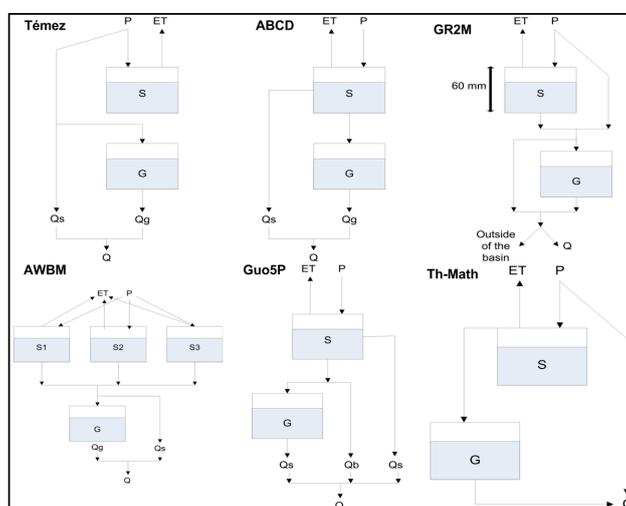

Figure 2. Conceptual representation of the six water balance models studied.

ABCD Model

The ABCD 4-parameter model [9] introduces a different formulation in the ET process and allows for a water surplus even though the soil moisture tank (S) is not full yet. As shown in Figure 2, this model also considers two storage tanks: upper storage (S), or soil, and groundwater storage (G). The upper storage tank (S) has two outputs: run-off (Qs) and infiltration. Thus, the model has two inputs, precipitation (P) and PET, and their outputs are soil moisture content at the end of the month (S), monthly available water, ET, run-off (Qs), infiltration, groundwater run-off (Qb), monthly groundwater storage (G), and total run-off (Q) (Figure 2).

GR2M-1994 Model (GR4-1994)

This model [11] was developed in the 1990s by the CEMAGREF (Centre of Agricultural and Environmental Research of France). It is also based on monthly precipitation and ET [128]. Afterwards, the model evolved into different versions, such as GR1A, GR2M, GR3J, and GR4J, denominated by the number of required parameters and with the last letter denominating the period considered: J (daily), M (monthly), or A (yearly). GR2M transforms precipitation into run-off through the implementation of two equations: production and transfer functions [128,129]. Initially, P and ET are balanced and precipitation is distributed between the upper storage tank (S) with a limited capacity and groundwater storage tank (G) [130]. Nief et al. [131] showed that model parameters are robust to non-stationary rainfall series, and calibrated parameter values are highly correlated with land use. Like previous models, monthly P and ET are the inputs, and the operating diagram is shown in Figure 2.

AWBM Model

The AWBM was also developed in the 1990s and is the most commonly used water balance method in Australia [46]. This model has three surface water-storage tanks (S1, S2, and S3). The water balance of each is estimated independently, resulting in three surpluses. One part of these surpluses is transformed into run-off (Qs), and the other part percolates to a groundwater storage tank or aquifer (G), which in turn goes to groundwater run-off (Qg). Total flow (Q) is obtained by adding both run-offs (Figure 2).

Guo Model (Five Parameters)

This model was developed to estimate the run-off in 70 catchments in southern China. It has a similar performance to the two-parameter Guo model, and its use is particularly recommended in humid and semi-humid regions [10,31]. Precipitation and evapotranspiration (P and ET) are the input data, on the basis of which the remaining parameters are estimated: ET, soil water storage (S), water surpluses, surface run-off (Qs), subsurface run-off (Qb), aquifer recharge, groundwater storage (G), groundwater run-off (Qg), and total flow (Q).

Thornthwaite-Mather Model

This model was developed by Thornwaite and Mather [8,132] in the early 1940s for the Delaware River, and many water balance models are based on it. The model distinguishes two water storage tanks: surface (S) and groundwater (G), which lead to the output flow (Q) through different calculations.

More detailed information about model performance and governing model equations is provided in Appendix A.

2.2.2. Goodness-of-Fit Tests

The study period is 34 years (1977–2010). We employed three years for warming up, and the other 30 years, starting October 1980, for calibration and validation, applying the split sample test proposed by Klemeš [133]. Therefore, the data series from 1980–2010 (30 years) was divided into two sets. We used the first 15 years (1980–1995) for calibration, and the remaining 15 years (1995–2010) for

model validation, namely as a security measure of calibration. As suggested by Kirchner [87], both series (calibration and validation) should exhibit different revealing behaviors with regards to the representativeness and reliability of modelling results and their deployment. Thus, we provided a comparison of precipitation values in both calibration and validation periods. We conducted calibration of all models' parameters by comparing the predicted data with the observed data. The optimal value for each parameter is the value that minimizes the differences between both flow series and the objective function that minimizes the sum of the square of deviations. We used the generalized reduced gradient as the optimization algorithm [134], which searches for the extreme values of the functions by the generalized reduced gradient algorithm method (GRG2) [135–138]. We used the multi-start method to find a globally optimal solution [139,140]. The multi-start method operates by generating candidate starting point values randomly selected between the bounds specified for the variables. These points are then grouped into clusters that are run repeatedly to capture the locally optimal solution. We used a Bayesian test to determine whether the process should continue or stop. As the number of runs of the non-linear iterations increases, the probability that the globally optimal solution has been found also increases 100%. For most non-linear problems, this method will at least yield very good solutions [140]. In this study, 1000 starting points have been used for each model and watershed. Once the model was calibrated and the parameters validated, we executed the predictive stage.

To evaluate model accuracy, six statistic indices were obtained: AIC [121,122], BIC [123], NSE [141], R^2 [142], PBIAS [143], and REV [144]. The formulas used are presented in Equations (2)–(11).

$$\text{AIC} = -2\log\ L\big(\check{O}\big) + 2k \tag{2}$$

$$\text{BIC} = -2\log\ L\big(\check{O}\big) + k\log n \tag{3}$$

where $\check{O}$ is the set (vector) of model parameters

$L(\check{O})$ is the likelihood of the candidate model given the data when evaluated at the maximum likelihood estimate of θ. Root mean squared error of the model output (RMSE) was used in this study; k is the number of estimated parameters in the candidate model, and n is the sample size

$$\text{NSE} = 1 - \frac{\sum_{i=1}^{n}\big(Q_{obs,i} - Q_{sim,i}\big)^2}{\sum_{i=1}^{n}\big(Q_{obs,i} - \overline{Q_{obs}}\big)^2} \tag{4}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\big(Q_{sim,i} - Q_{obs,i}\big)^2}{n}} \tag{5}$$

$$R^2 = \left(\frac{S_{obs,sim}}{\sqrt{S_{obs} * S_{sim}}}\right)^2 \tag{6}$$

$$S_{obs,sim} = \frac{1}{n-1} * \sum_{i=1}^{n}\big(Q_{obs,i} - \overline{Q_{obs}}\big) * \big(Q_{sim,i} - \overline{Q_{sim}}\big) \tag{7}$$

$$S_{obs} = \frac{1}{n-1} * \sum_{i=1}^{n}\big(Q_{obs,i} - \overline{Q_{obs}}\big)^2 \tag{8}$$

$$S_{sim} = \frac{1}{n-1} * \sum_{i=1}^{n}\big(Q_{sim,i} - \overline{Q_{sim}}\big)^2 \tag{9}$$

$$PBIAS = \frac{\sum_{i=1}^{n}\big(Q_{obs,i} - Q_{sim,i}\big) * 100}{\sum_{i=1}^{n} Q_{obs,i}} \tag{10}$$

$$REV = \frac{\sum_{i=1}^{n}(V_{obs,i} - V_{sim,i})}{\sum_{i=1}^{n} V_{obs,i}} \tag{11}$$

where Q_{obs}, V_{obs} and Q_{sim}, V_{sim} are the observed and simulated streamflow and run-off volumes, respectively.

According to Moriasi et al. [124], a grading method was established to evaluate the good performance of the model based on the NSE, R^2, and PBIAS values. A model was considered unsatisfactory if one of the previous tests is set as unsatisfactory, according to Moriasi et al. [124]. In any other case, the grading system proposed based on Moriasi et al. [124] and Bressiani et al. [125] was applied to classify the models in four categories—very good, good, satisfactory, and unsatisfactory—calculated by adding 3, 2, or 1 for each goodness-of-fit result according to Table 2.

Table 2. Classification criteria for hydrological models.

Goodness-of-Fit	NSE	PBIAS (%)	R^2	Grading	Classification-Sum
Very Good (V)	$0.75 < NSE \leq 1.00$	$PBIAS < \pm10$	$R^2 \geq 0.85$	3	$7 < E \leq 9$
Good (G)	$0.65 < NSE \leq 0.75$	$\pm10 \leq PBIAS < \pm15$	$0.75 < R^2 \leq 0.85$	2	$5 < E \leq 7$
Satisfactory (S)	$0.50 < NSE \leq 0.65$	$\pm 5 \leq PBIAS < \pm25$	$0.60 < R^2 \leq 0.75$	1	$3 < E \leq 4$
Unsatisfactory (U)	$NSE \leq 0.50$	$PBIAS \geq \pm25$	$R^2 < 0.60$	Unsatisfactory	Unsatisfactory

Furthermore, as our ultimate goal in this study is to use model predictions for water management, especially regarding the assessment of interannual water volumes, flow duration curves (FDCs) have been analyzed for the selected water balance models, and we evaluated specific hydrological metrics according to Yilmaz et al. [145] and Shafi and Tolson [146]. The signatures (Si) that will assess low-flows, high-flows, and mid-flows, in both observed and simulated ones, are shown in Equations (12)–(14).

FDC mid-segment slope (MS):

$$\log (Q_{m1}) - \log (Q_{m2}) \tag{12}$$

where m_1 and m_2 are the lowest and the highest flow exceedance probabilities within the mid-segment of the FDC (0.2 and 0.7, respectively)

FDC high-segment volume (HV):

$$\sum_{h=1}^{H} Q_h \tag{13}$$

where $h = 1, 2, \ldots , H$ are flow indices located within the high-flow segment probabilities of exceedance lower than 0.02; H is the index of the maximum flow.

FDC low-segment volume (LV):

$$\sum_{l=1}^{L}[\log(Q_l) - \log(Q_L)] \tag{14}$$

where $l = 1, 2, \ldots , L$ are the flow indices located within the flow-segment (0.7–1.0 flow exceedance probabilities); L is the index of the minimum flow.

The score for each signature will be calculated as Equation (15) [146]:

$$D_I = \frac{S_i^{obs} - S_i^{sim}}{S_i^{obs}} \times 100 \tag{15}$$

where Di is the deviation between the signatures (S$_i$) of the observed data (obs) and simulated model result (sim).

Model uncertainty for each watershed was assessed using FITEVAL software [126]. FITEVAL uses the general formulation of the coefficient of efficiency E_j (Equation (16)).

$$E_j = 1 - I \tag{16}$$

where Bi is a benchmark series, which may be a single number (such as the mean of observations), seasonally varying values (such as seasonal means), or predicted benchmark values using a function of other variables. For j = 2 and Bi = $\overline{Q_{obs}}$, E_j yields E_2 = NSE. Threshold values used for delimiting model efficiency classes are denoted as Unsatisfactory (NSE < 0.50), Acceptable (0.50 ≤ NSE < 0.65), Good (0.65 ≤ NSE < 0.75), and Very good (NSE ≥ 075). We studied two methods to account for model uncertainties and then implemented them within FITEVAL: Probable Error Range (PER) and Correction Factor (CF) [126].

3. Results

3.1. Precipitation Data Series Assessment

Firstly, we carried out an analysis of the monthly precipitation data series in Figure 3 using the boxplot graphics, distinguishing calibration (blue boxes) and validation (red boxes) periods, as the form of the evapotranspiration equation has limited influence on model performance in monthly hydrological models [147]. As can be seen in Figure 3, patterns of precipitation series in both periods are usually different, especially during the rainiest months that occur, mainly, from October to February in the more humid catchments (PUE, AND, BEG), thus decreasing (until May, even June) as the watershed becomes less humid (LEM, TRE, COT, PRI). The humid sub-humid ones (SEG, ZUM, JUB) shifted the rainiest period to spring, although maximums were also reached in winter. When the watershed is more arid, the distribution of precipitation over the year is more irregular, reaching similar values in December and June (RVA). This heterogeneity, not only between regions but also within the same watershed in both calibration and validation periods, represents different conditions and behaviors, which provides an optimal framework for hydrological purposes [87]. The driest period for all the watersheds occurs in summer months reaching, obviously, the lowest values in semi-arid and dry sub-humid watersheds. Likewise, July and August show the lowest variations, and consequently, the smallest boxplots. On the contrary, the highest differences are shown in the rainy months, varying in a wide range, reaching up to three times the median precipitation in nearly all watersheds. Concerning outliers in precipitation values (points and asterisks in Figure 3), the most humid watersheds, with aridity indices above 2 (PUE, AND, BEG), do not usually show numerous cases outside the bounds of the boxplots compared with the rest of the studied watersheds. When the aridity index ranges between 1 and 2, the monthly outliers in the boxplots occur more often and they are especially far out in the late spring or at the beginning of autumn. This trend levels out in humid sub-humid watersheds but dry sub-humid and semi-arid regions show extreme values for practically the whole year.

3.2. Models' Parameters

Table 3 shows the main characteristics of the models' structure, besides the range of parameters to be calibrated and the optimal ones found in the models and watersheds studied. Nearly all the optimal parameters of the models vary over the entire possible range. This circumstance is probably caused by the adaptation of the different hydrological processes embedded in the models' structure to the wide diversity of the climatic conditions in the regions analyzed. Parameters related to storage capacity of the superficial tanks (H in Témez; a in GR2M; ϕ in Thornthwaite-Mather; A1, A2, A3 in AWBM, etc.) have a strong decreasing value trend when the watershed is drier, due to the average soil moisture throughout the year and the specific climatic conditions of each watershed. The finding is further confirmed in other parameters, such as the maximum soil moisture included in the Guo-5p model in the S parameter, whose values are around 500–1000 mm in humid and sub-humid regions and drop out at 50–70 mm in the driest watersheds, but we found no tendency when PET is summed, as the ABCD model takes parameter b into account. However, underground storage is only regionally sensitive when the aquifer capacity is important, regardless of the climatic location. Moreover, the

ABCD model fails to perform the groundwater processes in semi-arid regions (d = 0), which is of high importance in the watersheds' hydrological water balance. No correlations were found between the parameters and catchment area.
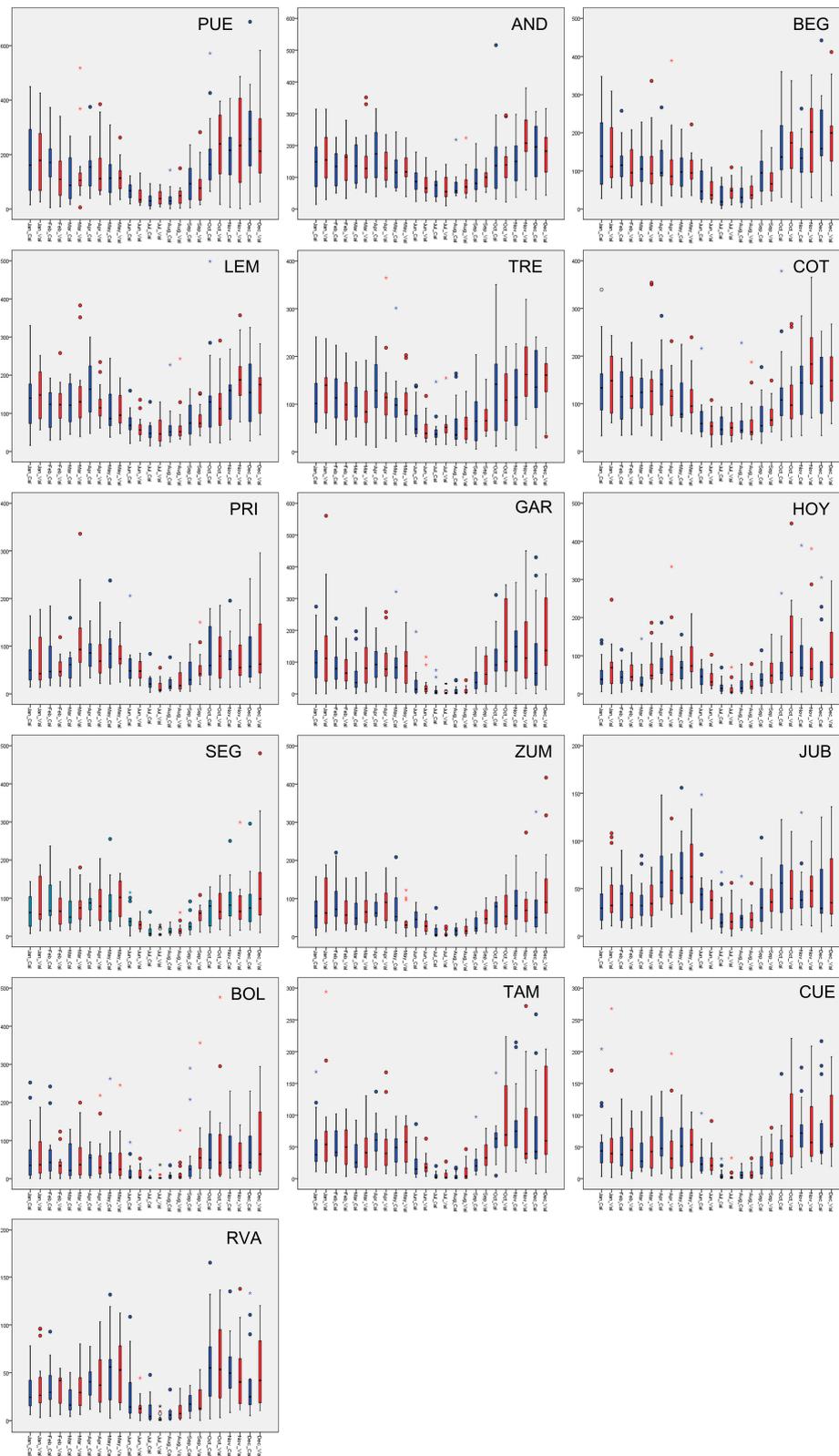

Figure 3. Boxplot for the calibration (blue) and validation (red) data series precipitation.

Table 3. Models' characteristics, parameters' value range, and optimal values obtained by calibration.

Model	Number of Storages	Number of Optimized Parameters	Parameters Value Range	Optimal Value Range
Témez	2	4	$50 < H < 250$ $0.2 < C < 1$ $10 < I < 150$ $0.001 < \alpha < 0.9$	$50 < H < 140$ $0.2 < C < 1$ $13 < I < 150$ $0.2 < \alpha < 0.9$
ABCD	2	4	$0 < a < 1$ $5 < b$ $0 < c < 1$ $0 < d < 1$	$a = 1$ $157 < b < 554$ $0.35 < c < 0.83$ $0.015 < d < 1$
GR2M-GR4	2	4	$0.6 < X1 < 1.9$ $0.03 < X2 < 18.2$ $100 < a$ $0.2 < \alpha < 0.5$	$0.46 < X1 < 1.87$ $0.07 < X2 < 0.94$ $126 < a < 462$ $0.2 < \alpha < 0.41$
AWBM	4	6	$50 < C < 200$ $0 < B < 1$ $0 < K < 1$ $0.5 < A1 < 1.5$ $0.5 < A2 < 1.5$ $0.5 < A3 < 1.5$	$42 < C < 92$ $0.4 < B < 0.5$ $0.3 < K < 0.61$ $0.2 < A1 < 0.3$ $0.35 < A2 < 0.4$ $0.35 < A3 < 0.4$
Guo 5p	2	5	$0 < K0 < 2$ $0 < K1 < 1$ $0 < K2 < 1$ $0 < C < 1$ $0 < S$	$0.6 < K0 < 1.5$ $0 < K1 < 0.5$ $0 < K2 < 0.6$ $0 < C < 0.7$ $250 < S < 1000$
Thorn Thwaite-Mather	2	3	$0 < \alpha < 1$ $0 < \phi$ $0 < \lambda < 1$	$0.02 < \alpha < 1$ $0.001 < \phi < 256$ $0.001 < \lambda < 0.91$

3.3. Goodness-of-Fit Tests

For a better understanding of the research and subsequent discussion, we ranked the results in tables and figures according to the AI_U values.

3.3.1. AIC and BIC Criteria

AIC and BIC criteria classify the best model when the value achieved is lower. Although they penalize the number of estimated parameters, neither of them provide priority to the Thornthwaite-Mather model in this study (Tables 4 and 5), as it has the fewest parameters to be estimated compared to the others used. In fact, it achieves less value in either of the watersheds. Both AIC (Table 4) and BIC (Table 5) show better results globally when using GR2M, proving to be the best model in 12 of the 16 catchments for AIC and 8 of the 16 for BIC, both in humid and dry sub-humid catchments. Témez, widely used in Spain, achieves the best result only in CUE for AIC and BIC, as well as for LEM and PRI, located in humid sub-humid areas. AWBM and Guo-5p reach the lowest values in humid sub-humid HOY, SEG, and ZUM. Focusing on the coefficient of variation (C.V.) in the last column of Tables 4 and 5, it is clear that when the watershed is less humid, the C.V. obtained is higher, which means a clear difference between the used models. When AIU is higher than 1.5, the C.V. is below 10% but in the LEM watershed, the cover land changes are highlighted, for both AIC and BIC. Such minor differences indicate the similar performance of nearly all models studied in the most humid catchments. Nonetheless, if AIU is below 1.5, the C.V. becomes greater, which means a huge difference between the results achieved with the best model and with the rest of the models, reaching more than 130% in ZUM. Furthermore, in the driest watersheds, the C.V. is between 20% and 75%, which means that the AIC and BIC criteria are also the basis for the hydrological model selection in the dry and semi-arid ones, although the results may not differ in the humid ones.

Table 4. AIC criterion values for calibration (Calib.) and validation (Valid.) periods for the selected models (best model in bold; C.V. (%) is the coefficient of variation).

|	Témez		ABCD		GR2M		AWBM		GUO5P		TH-MTH		Average	CV (%)
	Calib.	Valid.	Calib.	Valid.	Calib.	Valid.	Calib.	Valid.	Calib.	Valid.	Calib.	Valid.		
PUE	1311	1268	1293	1119	**1189**	**1058**	1320	1146	1293	1135	1289	1131	1213	7.60
AND	1095	1213	1093	1208	**1060**	**1182**	1123	1218	1190	1148	1112	1208	1154	4.83
BEG	1150	1152	1138	1164	**1114**	**1130**	1145	1173	1139	1169	1137	1165	1148	1.54
LEM	**1053**	**718**	1042	799	1015	877	1064	786	1041	828	1048	790	922	14.41
TRE	869	783	855	753	**796**	**758**	889	790	773	833	876	772	812	6.06
COT	1376	1231	1286	1163	**1154**	**1184**	1389	1259	1282	1166	1278	1162	1244	6.60
PRI	365	470	357	505	**326**	**482**	371	499	357	500	355	501	424	17.24
GAR	22	255	35	286	**−2**	**205**	27	230	138	151	−10	245	132	84.95
HOY	352	477	**256**	**430**	216	504	326	470	231	456	306	498	377	28.72
SEG	206	609	170	454	**84**	**461**	344	482	172	431	161	458	336	50.37
ZUM	221	499	−12	207	**−146**	**163**	−89	196	−87	197	30	342	127	151.57
JUB	−319	−132	−296	−144	**−397**	**−196**	−391	−191	−366	−152	−411	−172	−264	−41.57
BOL	−50	−30	−72	−27	−118	−27	−52	2	**−61**	**−88**	−55	−23	−50	−64.86
TAM	348	609	375	478	423	433	353	545	**444**	**417**	400	671	458	22.15
CUE	**447**	**292**	440	352	431	387	444	326	191	503	433	321	381	22.92
RVA	−1050	−316	**−1036**	**−985**	−1089	−350	−1056	−325	−429	−601	−980	−301	−710	−48.87

Table 5. BIC criterion values for calibration (Calib.) and validation (Valid.) periods for the selected models (best model in bold; C.V. (%) is the coefficient of variation).

	Témez		ABCD		GR2M		AWBM		GUO5P		TH-MTH		Average	CV (%)
	Calib.	Valid.	Calib.	Valid.	Calib.	Valid.	Calib.	Valid.	Calib.	Valid.	Calib.	Valid.		
PUE	1324	1280	1306	1132	**1213**	**1071**	1339	1159	1309	1151	1299	1140	1227	7.57
AND	1108	1226	1106	1221	**1083**	**1195**	1143	1231	1206	1164	1122	1217	1168	4.63
BEG	1163	1165	1150	1177	**1137**	**1143**	1164	1186	1155	1184	1147	1174	1162	1.39
LEM	**1066**	**730**	1055	812	1038	890	1083	799	1057	843	1058	800	936	14.34
TRE	881	796	868	765	**820**	**770**	908	802	789	848	885	781	826	6.02
COT	1388	1243	1299	1176	1177	1197	1408	1272	1298	1182	**1287**	**1171**	1258	6.54
PRI	**378**	**482**	370	518	349	495	390	512	373	516	364	511	438	16.36
GAR	34	268	48	299	**20**	**218**	46	242	153	167	−1	255	146	75.84
HOY	364	489	**268**	**443**	237	517	344	482	246	472	315	508	390	27.37
SEG	218	622	183	467	107	474	**215**	**430**	187	447	171	468	332	50.47
ZUM	233	511	1	220	**−123**	**176**	−70	209	−71	213	40	352	141	134.65
JUB	−306	−120	−283	−131	**−374**	**−196**	−372	−178	−350	−136	−401	−163	−251	−42.82
BOL	−38	−17	−59	−27	−95	−15	−33	15	**−46**	**−72**	−45	−14	−37	−78.51
TAM	360	621	388	490	**423**	**445**	372	557	460	432	410	680	470	21.42
CUE	**460**	**304**	452	365	454	400	463	338	207	519	443	331	395	22.31
RVA	−1037	−304	**−1036**	**−972**	−1066	−337	−1037	−312	−413	−585	−970	−292	−697	−49.76

3.3.2. Grading Classification (NSE, PBIAS, R^2)

In addition to the BIC and AIC criteria, we calculated NSE, PBIAS, and R^2 and assessed a grading classification, according to Table 2, of the observed and simulated streamflow data (Q_{obs}-Q_{sim}) for all the evaluated approaches (Figure 4). The NSE for each model considering the 16 basins (Figure 4a) showed good results for all models except Témez, for which the average NSE (0.44) was below 0.50. Likewise, the most extreme outliers are shown for the Témez model in both the calibration and validation periods, reaching an absolute minimum below −1 in ZUM. AWBM and Guo-5p also show outliers in the validation period, although closer to the minimum values in respective boxplots, which correspond with the semi-arid watershed (RVA). GR2M and ABCD seem to be the best models, in general, in peninsular Spain according to the NSE values, although NSE for GR2M in semi-arid watersheds (RVA) is close to zero (0.18), and it is considered unsatisfactory. Despite being a five-parameter model, the Guo-5p would only be satisfactory according to the NSE criterion. However, a three-parameter model, such as the Thornthwaite-Mather, also achieved good results when considering only NSE (except in RVA), accounting for both wet and dry sub-humid Spain, which confirms the absence of the need for

very complex water balance models, as suggested by Clark et al. [148], Perrin et al. [19], Jakeman and Hornberger [149], Michaud and Sorooshian [150], or Beven [151].
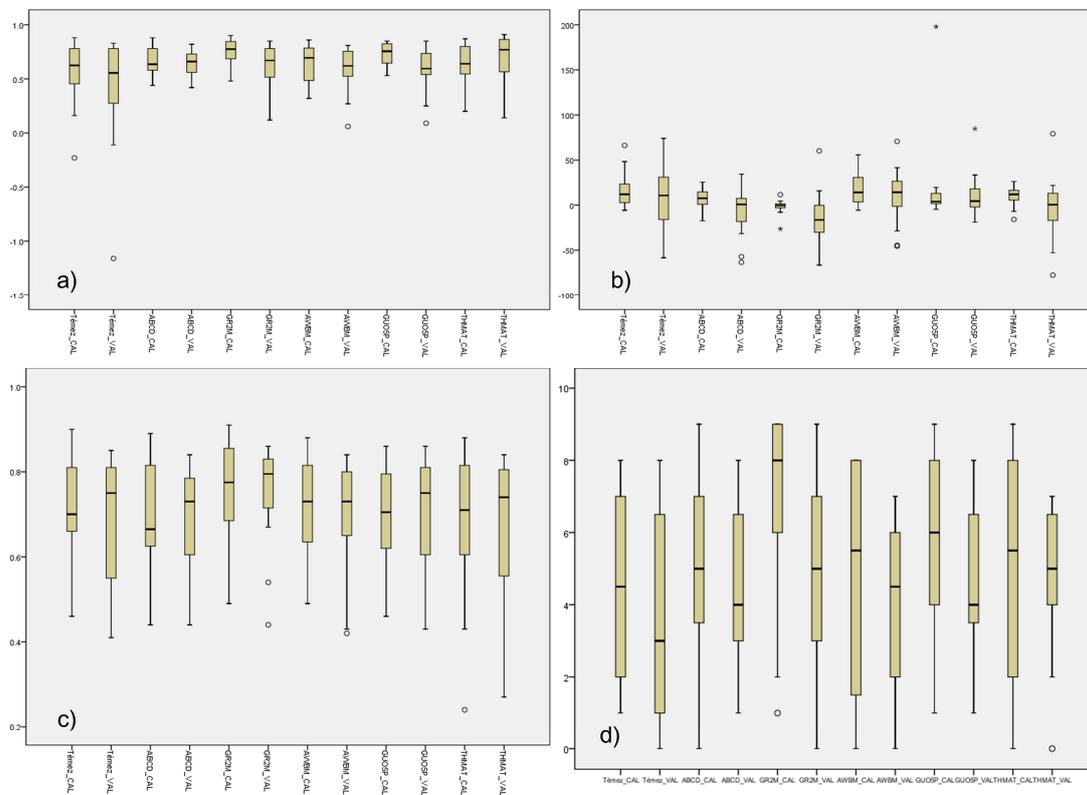
Figure 4. Boxplot for the calibration (CAL) and validation (VAL) values of NSE (**a**), PBIAS (**b**), R^2 (**c**) and grading classification (**d**) of assessed models.

However, NSE cannot help identify model bias [124], and it should be complemented by other measures, such as PBIAS, in order to consider how well the model simulates the average magnitudes for the outputs and to identify the average model simulation bias (overprediction versus underprediction). The PBIAS boxplot for the models analyzed (Figure 4b) shows that this statistic technique provides an irregular rating of model performance without correlation with the aridity, as it provides high absolute values in the calibration and validation periods for all the models and watersheds. This may happen because PBIAS is very sensitive to extreme values [124] and its use is not particularly recommended to identify differences in the timing and magnitude of peak flows, such as in the case of the watersheds in peninsular Spain, as shown in the previous section. Notwithstanding, the best results are achieved with GR2M in calibration, but the validation period is more varied and ABCD and AWBM obtain similar results in PBIAS values.

Although the GR2M showed the best fit according R^2 results, in nearly all catchments in the calibration (16 out of 16) and validation periods (14 out of 16), both semi-arid and sub-humid, we found an average correlation coefficient of close to 0.90. Boxplots for the rest of the models are not very different in either the median or amplitude (Figure 4a). The ABCD and AWBM models did not give the best fit in the validation period for any of the studied catchments, although their values were similar to those using the GR2M, especially for humid and sub-humid catchments. Increasing the number of parameters does not guarantee a better performance, as previous studies have shown [19,21,147,150,152], as the AWBM (6 parameters) did not show better results than the rest, even when it was compared to the lowest parameter model, the Thornthwaite-Mather (3 parameters), which had better results for nearly all watersheds, both humid and semi-arid. All water balance models showed correlation coefficients above 0.70 for humid and sub-humid catchments. In the dry

sub-humid HOY and semi-arid RVA catchments, some values fell below 0.60, especially with the Thornthwaite-Mather and Témez models, despite the latter being widely used in Spain. It achieved the best fit only in the GAR and PRI catchments, taking similar values as the Guo-5p and GR2M coefficient. When the catchment was more arid, the R^2 obtained was lower, which is in line with other studies [82,147]. It is essential to consider that R^2 is oversensitive to extreme values [153] and insensitive to additive and proportional differences between model predictions and measured data [154], and it should be used with other goodness-of-fit tests and different criteria, both graphical and metrics, as we considered in this study, to provide consistency in the selection of models.

According to the grading classification criteria (Table 2) shown in Figure 4d, only the Témez model was unsatisfactory, whereas the rest were, on average, between good and satisfactory. This may be because the surpluses law of Témez is asymptotic to the one proposed by Thornthwaite-Mather for the highest precipitation values, but it differs on the lowest side of the curve, without requiring both the PET and soil moisture deficit to be used up completely. Furthermore, despite being a widely used hydrological model in Spain, underground Témez modeling is oversimplified and is invalid for karstic aquifers where there is more than one curve of discharge in the depletion of the aquifer [106].

With regards to the performance of the models' results in each basin (Figure 5), we highlighted the influence of the aridity index, as mentioned before, which indicates that the climatic characteristics of the watershed are the most important issues in a model's performance. When the watershed is drier, the worst performance is achieved, as the grading classification value decreases from humid to semi-arid regions. As might be expected, validation results (Figure 5a) are usually worse than the calibration ones (Figure 5b), reducing the range in the validation boxplots. However, PUE and LEM show better classification in the validation period, as there are more hydrological models whose NSE reach higher values than in calibration period. This may occur in cases where the measured data are bi-modal with high and low distributions in the same study area, such as the measured flows [155]. PRI is the only watershed that shows similar behavior to nearly all the models studied. Changes in land use in the last two decades (over 10% noticed in LEM and RVA) are not highlighted when using these metrics, especially in LEM, where the land use has changed almost 20% due to reforestation. Humid and humid sub-humid watersheds show good results in many of the hydrological models, with an average of 5–6 in the grading classification. This value decreases in dry sub-humid catchments until 4–2, and the semi-arid RVA only obtains satisfactory classification with GR2M in the calibration period and Guo5p in the validation period.
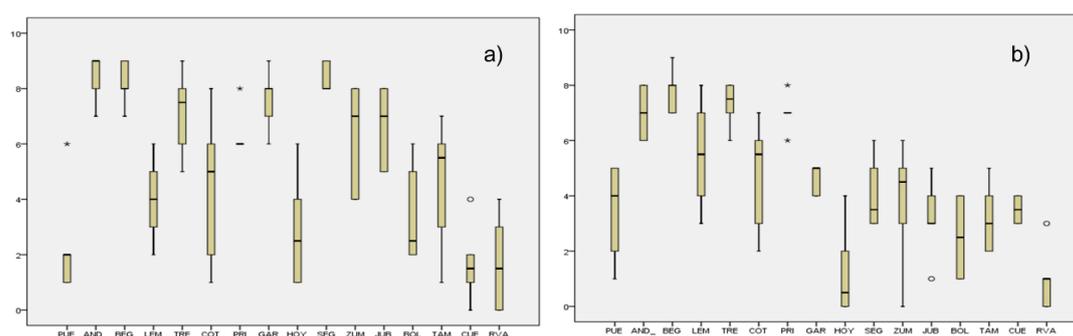
Figure 5. Boxplot for the calibration (**a**) and validation (**b**) of grading classification in the studied watersheds.

Table 6 summarizes the classification sum according to the grading criteria proposed in Table 2 for all studied catchments and models in both the calibration (1980–1995) and validation (1995–2010) periods. We selected the model that best fits each watershed according to the highest and similar values in the calibration and validation periods (Table 6). The GR2M gave the best fit of the catchments, with around 80% in the calibration period and 50% in the validation period with values over 7, meaning they all had a very good fit. The disparities between calibration and validation may be due

to the climatic differences between the calibration and validation periods, as analyzed in previous sections. Semi-arid catchment, RVA, was unsatisfactory for the GR2M and for the rest of the models studied. These watersheds have a very irregular rainfall distribution, producing frequent run-off peaks, and although they have deep soil, low infiltration is produced; therefore, models do not achieve good results.

Table 6. Classification values (calibration: 1980–1995/validation: 1995–2010). Best global fit in bold.

Catchment	Témez	ABCD	GR2M	AWBM	GUO-5P	THOR-MATH	Best Model	Classification
PUE	1/1	2/5	**6/5**	1/2	2/4	2/4	GR2M	**Good**
AND	8/6	9/7	**9/8**	7/6	**9/8**	9/7	GR2M	**Very Good**
BEG	7/8	9/8	**9/9**	8/7	8/8	8/7	GR2M	**Very Good**
LEM	3/8	4/5	6/3	2/7	4/4	**5/6**	TH-MT	**Good**
TRE	7/8	8/8	**9/8**	5/6	8/7	6/7	GR2M	**Very Good**
COT	2/3	4/6	8/5	1/2	6/6	**6/7**	TH-MT	**Good**
PRI	6/7	6/7	**8/8**	6/7	6/7	6/6	GR2M	**Very Good**
GAR	8/5	6/4	**9/5**	8/5	7/4	8/5	GR2M	**Good**
HOY	1/0	**4/2**	6/0	1/0	4/1	1/4	ABCD	**Unsatisfactory**
SEG	8/3	8/3	9/3	8/5	8/4	**9/6**	TH-MT	**Good**
ZUM	4/0	6/3	**8/6**	8/5	8/5	4/4	GR2M	**Good**
JUB	5/1	5/3	8/4	8/3	6/3	**8/5**	TH-MT	**Good**
BOL	2/3	3/4	6/1	2/1	**5/4**	2/2	GUO-5P	**Satisfactory**
TAM	7/2	5/4	1/5	6/2	**6/4**	3/2	GUO-5P	**Satisfactory**
CUE	1/4	2/3	2/3	0/4	**4/3**	1/4	GUO-5P	**Satisfactory**
RVA	3/0	0/1	**4/1**	2/1	1/3	0/0	GR2M	**Unsatisfactory**
Best Fit (Number of Times)	**0**	**1**	**8**	**0**	**4**	**4**		

The Guo5p and Thornthwaite-Mather models were the best in the four catchments. The former has the best fit in some of the humid and humid sub-humid catchments but gives similar results to GR2M. However, Guo5p provides satisfactory results in dry sub-humid regions, whereas the rest of the models show important differences between calibration and validation periods. The ABCD model was the best in only one catchment but showed a value below 3 in the validation period, which means unsatisfactory. Nevertheless, humid watersheds show good behavior with nearly all the models, corroborating the better performance of nearly all models in wetter conditions. In general, the models that showed the best results on average, in all catchments, were first the GR2M, and then the Guo5p and Thornthwaite-Mather, with values around 7/5 (calibration/validation) for the GR2M model and around 5/5 (calibration/validation) for the Guo5p and Thornthwaite-Mather models. The AWBM and Témez showed the worst results, with 4/4 on average, which are almost unsatisfactory.

As with previous comparison methods, the best results were obtained in more humid basins, and the drier regions showed more unsatisfactory results. All of the analyses taken into account did not show a correlation between the catchment area and goodness-of-fit results, so higher model resolution does not seem to be an improvement for humid and sub-humid watersheds [74,156,157].

4. Discussion

4.1. Water Volume Assessment (REV)

The main aim of hydrological balance models is to assess inflows in a water resource system, and it is essential for appropriate analysis of its availability. Therefore, in addition to the assessment performed with the proposed grading classification (Table 2), we should analyze the differences between the total observed and simulated run-off volumes to validate or discard a model. Unlike the previous tests taken into account, REV results do not highlight one or two specific hydrological models, but the best fitted model for each catchment differs quite substantially. Table 7 shows the REV results of these comparisons for the catchments we studied and the models considered. According to this criterion, the Guo-5p and Thornwaite-Mather show the best results, on average, giving the best fitted models in three of the catchments studied. Témez is the worst model for nearly all catchments, reaching disparities that range between 10% and 80% compared to the best performances of other models.

As with the BIC and AIC criteria, when the catchment is more arid, the differences between REV results are higher; therefore, there are several models that achieve REV lower than 10% for humid regions, provided that AI_U is higher than 1.1. However, results for GR2M for LEM are higher, probably caused by the land cover change in the studied period. The worst results are shown in humid areas, with AI_U around 1.0 and dry sub-humid and semi-arid catchments with REV around 20–40%, on average. REV is particularly unsatisfactory in semi-arid regions (RVA), where no model achieves values below 30%. Also noteworthy is the poor results shown by GR2M in dry sub-humid and semi-arid regions, which is in line with the classification of models in Table 6.

Table 7. REV (%) values in period 1980–2010. Best fit (lowest absolute value for each watershed) in bold.

Catchment	Témez	ABCD	GR2M	AWBM	GUO-5P	THOR-MATH
PUE	−36.72	−12.76	11.92	−39.73	−9.86	**−9.56**
AND	−12.40	−6.44	3.14	−11.45	**−0.15**	−9.85
BEG	−9.53	−5.56	**1.72**	−11.25	−2.83	−10.45
LEM	−9.10	**1.04**	14.85	−17.41	1.85	7.11
TRE	−7.41	**−0.67**	7.38	−14.28	−2.23	−9.94
COT	−42.95	−9.74	10.60	−46.93	−5.43	**−5.36**
PRI	−7.94	1.63	2.99	−5.53	**−0.61**	−3.10
GAR	41.31	50.72	13.54	32.88	**−11.47**	40.90
HOY	−51.91	8.98	26.48	−41.58	13.96	**3.08**
SEG	−2.45	11.07	7.47	**1.10**	4.65	3.34
ZUM	15.43	11.43	**0.41**	−2.51	−2.08	−1.63
JUB	−15.12	−11.59	**−8.72**	−10.54	−13.31	−12.31
BOL	−3.30	10.26	49.61	−4.03	17.32	**−1.75**
TAM	89.01	55.40	58.10	71.97	**13.16**	111.98
CUE	5.15	20.24	24.40	**−3.28**	−46.18	12.45
RVA	−63.14	**−29.66**	−47.55	−60.21	−33.07	−68.22
Average (Absolute Value)	25.80	15.45	19.12	23.42	11.14	19.44
Best fit (Number of Times)	**0**	**3**	**3**	**2**	**4**	**4**

4.2. Model Selection

Table 8 shows the proposed model for each catchment, taking into account the proposed set of criteria: AIC, BIC, the grading method, and REV. Values of AIC and BIC in Table 8 denote the ranking position of the model according to these methods, varying between 1 and 6; 1 for the best result (lower value of the index) and 6 for the worst one. As seen in previous sections, the best results are achieved in humid regions for all the criteria taken into account. Furthermore, the AIC and BIC values match in most cases with the other used methods. We confirmed GR2M, as it was the best model in 50% of the catchments analyzed and had the best results in all of the goodness-of-fit tests applied. Moreover, it reached grades from good to very good according to Table 2, not only in humid catchments but also in humid sub-humid ones, with AIU being higher than 0.6 for ZUM and JUB. The good performance of GR2M may be due to its inclusion of a water exchange function alongside two independent parallel routing paths, which is of great importance in the transference between contrasting wet and dry periods [82]. Despite a value of AIU of 1.96 in LEM, only Témez is able to provide satisfactory results for all the criteria, with a REV lower than 10%. However, consideration should be made, as the land cover in LEM has changed more than 20% in the studied period.

Thornthwaite-Mather shows good performance in COT and SEG, although the AIC and BIC criteria do not classify this model as the best one, despite the low number of parameters. ABCD is considered unsatisfactory for HOY, but none of the remaining models obtained better results in this catchment. This may be caused by its altitude, as it is situated at around 1600 m, resulting in the highest altitude in our study, and snow precipitation events are relatively important and could affect hydrological modeling. Nonetheless, REV was below 10%, and the total volume assessment in the studied period can be considered good, although better results of REV (3.08) were achieved with Thornthwaite-Mather in HOY. Furthermore, the ABCD model enables surpluses, even when soil is

not saturated, and the division between run-off and aquifer recharge is modeled by a constant time coefficient, which is rather unrealistic. Dry sub-humid and semi-arid catchments with AIU lower than 0.6 obtained the best performance with Guo5p, but results are unsatisfactory when AIU falls below 0.54, highlighting the relevance of aridity in lumped model performance. Regarding REV results, nearly all models show an error below 10%, although REV values increase to about 20% when the catchment is more arid. We did not find a trend for the catchment area or location and for the models used.

Table 8. Summary of performance of the selected water models. AIC and BIC values are referred to first–third place in increasing order value in Tables 4 and 5, respectively.

Catchment	Area (km^2)	Altitude (MASL)	Model	AIC	BIC	Grading Method	REV (%)
PUE	263.85	400.05	GR2M	1	1	Good	+11.92
AND	778.49	486.01	GR2M	1	1	Very Good	+3.14
BEG	836.89	504.01	GR2M	1	1	Very Good	+1.72
LEM	252.58	342.18	Témez	1	1	Good	−9.10
TRE	413.54	526.64	GR2M	1	1	Very Good	+7.38
COT	488.22	559.51	Th-Mt	2	1	Good	−5.36
PRI	328.16	1255.05	GR2M	1	2	Very Good	+2.99
GAR	69.92	689.98	GR2M	1	1	Good	+13.54
HOY	66.15	1632.12	ABCD	1	1	Unsatisfactory	+8.98
SEG	232.89	1416.46	Th-Mt	2	3	Good	+3.34
ZUM	266.03	1549.95	GR2M	1	1	Good	+0.41
JUB	207.66	1150.05	GR2M	1	1	Good	−8.72
BOL	29.23	600.31	Guo-5p	1	1	Satisfactory	+17.32
TAM	458.12	447.46	Guo-5p	1	2	Satisfactory	+13.16
CUE	139.86	610.63	Témez	1	1	Unsatisfactory	+5.15
RVA	85.68	607.70	Guo-5p	1	2	Unsatisfactory	+11.14

4.3. Models Uncertainty Analysis

After selecting the best model for each catchment in the studied period, defining the thresholds of acceptance of the model depends on its application of itself [158]. We used FITEVAL software [126] to provide an uncertainty framework of the results achieved. This application performs the goodness-of-fit of observed and simulated streamflows by the line balancing graphic versus line 1:1, the computation of NSE and RMSE and their corresponding values for the 95% confidence interval, the qualitative measure of adjustment, and the presence of bias and outliers. Figure 6 collects, as examples of FITEVAL, the outputs obtained for PUE, BEG, CUE, and RVA, and Table 9 shows the summary of results for each catchment and model selected in the previous sections. The values and graphics shown for GR2M models are by far the best compared to the rest of the models selected; therefore, more than 90% of the confidence intervals are above 0.75 NSE, considering that the model was very good in humid watersheds. When AI_U was below 1.0, the confidence interval became greater for all the models, although GR2M also remained lower than the rest with an average range of 0.6–0.8 NSE, which means acceptable and very good models, respectively. On the contrary, Témez and Thornthwaite-Mather did not reach 0.75 NSE, even in humid catchments, and they fell by up to 0.1–0.3 NSE when AIU was below 1.0, such as SEG and CUE. Guo5p had similar performance in the most arid catchments and the confidence interval range comprises up to 70% of the uncertainty in the case of the semi-arid RVA. In general, outliers in catchments with AI_U lower than 1.0 tend to underestimate the streamflow, whereas in the humid ones, the unusual presence of outliers tends to overestimate predictions around 5–10%.

Table 9. Summary of uncertainty analysis for each catchment using FITEVAL in 1980–2010 period.

Catchment	Model	Probability NSE = 0.75–1.00	Probability NSE = 0.65–0.75	Probability NSE = 0.50–0.65	Probability NSE < 0.50	Classification
PUE	GR2M	44.5 %	55.1%	0.4%	0.0%	Good-Very Good
AND	GR2M	100%	0.0%	0.0%	0.0%	Very Good
BEG	GR2M	100%	0.0%	0.0%	0.0%	Very Good
LEM	Témez	3.6%	65.9%	30.5%	0.0%	Acceptable-Good
TRE	GR2M	95.2%	4.8%	0.0%	0.0%	Good-Very Good
COT	Th-Mt	0.1%	87.0%	12.9%	0.0%	Acceptable-Good
PRI	GR2M	91.8%	8.2%	0.0%	0.0%	Good-Very Good
GAR	GR2M	94.6%	5.2%	0.2%	0.0%	Good-Very Good
HOY	ABCD	0.0%	0.4%	49.8%	49.8%	Unsatisfactory-Acceptable
SEG	Th-Mt	25.3%	25.4%	31.0%	18.3%	Unsatisfactory-Very Good
ZUM	GR2M	14.6%	43.7%	41.7%	0.0%	Acceptable-Very Good
JUB	GR2M	17.9%	60.9%	21.0%	0.2%	Acceptable-Very Good
BOL	Guo-5p	1.9%	42.8%	50.1%	5.2%	Unsatisfactory-Good
TAM	Guo-5p	34.3%	50.0%	15.2%	0.5%	Acceptable-Very Good
CUE	Témez	0.3%	10.1%	60.2%	29.4%	Unsatisfactory-Good
RVA	Guo-5p	0.2%	5.0%	29.5%	65.3%	Unsatisfactory-Good
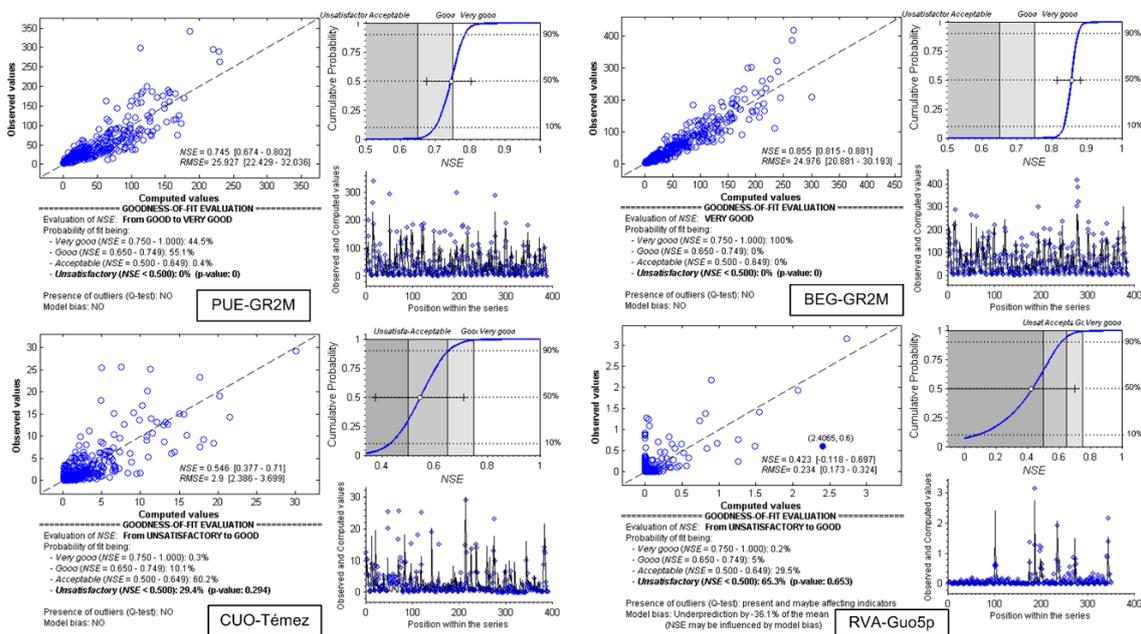
Figure 6. Uncertainty analysis of the models selected in PUE, BEG, CUE, and RVA using FITEVAL application.

4.4. Flow Duration Curves

As our ultimate goal in this study is to use model predictions for the assessment of water availability, we developed FDCs for the observed and simulated streamflows for the whole studied period (1980–2010) (Figure 7) in order to evaluate the quality metrics related to different segments of the streamflow series. As previously determined, when the watershed is more humid, the best performance is obtained. All the watersheds had a similar pattern, in general terms, except the BOL, TAM, CUE, and RVA, which were the driest of those studied and were performed by the Guo5p and Témez models. However, Témez for LEM had good results, as REV suggested, although differences reached 59.4% in the low volumes. The graphics confirm the previous values of criteria and grading classification proposed, showing, in general, good performance in the highest and medium volumes. Lowest volumes had poor performance, especially when AI_U was lower than 1. Concerning the humid and sub-humid regions, GR2M had good performance in probabilities of exceedance lower than 80%, (i.e., in high and medium volumes), but both curves were separated when the streamflow was lower. Thornthwaite-Mater's model structure tends to underestimate low streamflow and to

overestimate high streamflow, thus, even in COT with an AI_U of 1.65, the observed and simulated FDCs are separated for high and low volumes. Table 10 quantifies the variabilities identified by FDCs using the signatures for high, low, and mid-segment volumes, as we described earlier. Deviations in humid and sub-humid watersheds ranged from 0–30% in both high volumes and mid-segment slopes, provided that GR2M was the model used. When AI_U is below 1, the differences are significantly greater, regardless of the model, especially in the mid-segment and low-segment volume. Therefore, values in low volumes for humid watersheds showed an average of 30% but reached values higher than 100% when AI_U was below 1, particularly marked in SEG, BOL, and RVA, which we performed using Guo5p and Thornthwaite-Mater. The best performance was obtained in the AND, BEG, and TRE watersheds (GR2M model), with values below 25% for all the signatures. Nonetheless, low volumes in most humid watersheds accounted for less than 5% of the total streamflow in the watersheds in the studied period, so their importance is relative for water resources assessment related to urban and agriculture water demand. Notwithstanding, as suggested by Operacz et al. [159] and Pusłowska-Tyszewska and Tyszewski [160], low flows have to be taken into account from an environmental point of view to maintain minimum ecological discharge. The differences in low volumes increased when the watershed was less humid. While high volumes' differences stood at around 10–40%, signatures in low volumes reached negative 200–400%, overestimating streamflows in spring and summer seasons. As with humid regions, volumes with a probability of exceedance of 80% accounted for less than 8% of the total volume in the studied period. On the contrary, semi-arid regions showed good performance in high volumes, but both mid-segment slopes and low volumes had differences above 100% in most cases, and high volumes in these watersheds accounted for only 25% of the total streamflow in the studied period.

Table 10. Quality metrics of FDC. (MS: Mid-segment slope; HV: High-segment volume; LV: Low-segment volume; Di: Relative difference percentage).

Catchment	Model	Observed			Simulated			Di		
		MS	HV	LV	MS	HV	LV	MS	HV	LV
PUE	GR2M	0.71	1878.65	67.54	0.81	1435.18	86.01	−14.68	23.61	−27.35
AND	GR2M	0.75	2239.05	164.43	0.71	1679.94	122.85	6.0	25.0	25.3
BEG	GR2M	0.76	2283.95	61.11	0.73	1789.04	59.91	4.0	21.7	2.0
LEM	Témez	0.70	943.00	138.90	0.76	569.64	56.39	−8.6	39.6	59.4
TRE	GR2M	0.58	701.70	47.69	0.63	643.24	59.56	−8.6	8.3	−24.9
COT	Th-Mt	0.73	1725.80	110.60	0.44	1189.46	74.69	39.5	31.1	32.5
PRI	GR2M	0.40	321.01	29.96	0.38	216.56	16.61	7.3	32.5	44.6
GAR	GR2M	0.98	107.54	181.15	0.74	94.75	24.97	24.8	11.9	86.2
HOY	ABCD	0.70	147.23	72.38	0.39	109.93	16.41	44.8	25.3	77.3
SEG	Th-Mt	0.36	179.70	24.21	0.57	198.59	129.37	−56.1	−10.5	−434.5
ZUM	GR2M	0.35	118.50	37.60	0.42	80.15	145.56	−23.0	32.4	−287.2
JUB	GR2M	0.42	33.88	27.07	0.31	24.34	32.34	25.0	28.2	−19.5
BOL	Guo-5p	1.10	45.43	115.76	4.02	35.13	696.28	−265.3	22.7	−501.5
TAM	Guo-5p	3.34	283.49	9.26	0.32	299.61	0.10	90.4	−5.7	98.9
CUE	Témez	0.79	168.00	278.79	0.78	145.80	99.88	1.17	13.21	64.17
RVA	Guo-5p	1.09	12.55	37.87	6.05	14.56	93.40	−455.4	−16.1	−146.6

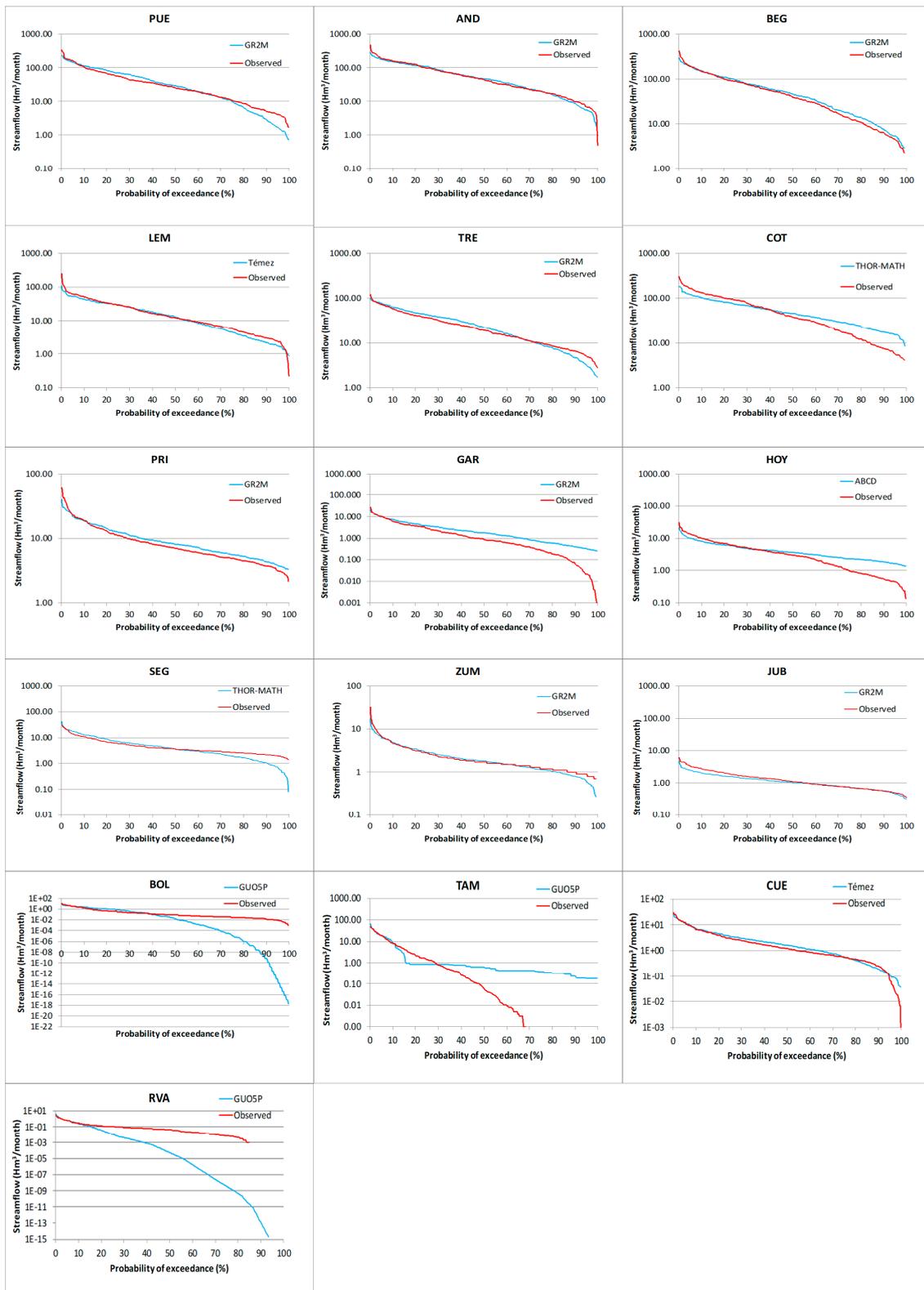**Figure 7.** Flow duration curves for observed and selected models' values in the study period (1980–2010).

5. Conclusions

Spain has wide climate variation due to its complex orography and geographic situation. It has the driest and rainiest regions in continental Europe. Indeed, the 16 basins we selected as case studies cover a range of aridity index classifications from humid to semi-arid. We performed the assessment

of the goodness-of-fit using various model selection criteria, so as to establish a robust methodology that can validate the conclusions obtained and its application in other different regions. Lumped hydrological balance models performed well in humid and sub-humid regions, regardless of the area, so higher model resolution does not seem to be an improvement for humid and sub-humid watersheds. The influence of the aridity index is highlighted in all of the goodness-of-fit tests carried out, which indicates that the climatic characteristics of the watershed are the most important issue in a model's performance. When the watershed is more arid, the distribution of precipitation over the year is more irregular, and although these catchments have deep soil, low infiltration was produced; thus, the models did not achieve good results. Complexity in the models and over-parametrization do not guarantee a better performance, and three-parameters models, such as the Thornwaithe-Mather model, show, in some cases, better monthly simulations than the Guo-5p. However, although the driest regions registered "unsatisfactory" performance for the lumped models used and alternative methods (such as machine-learning modelling and semi-distributed models, such as SWAT [161]) will be assessed in future studies, the estimated run-off volumes with the Guo-5p are very similar to the observed ones, with differences below 12%, which is even lower than in some dry sub-humid regions.

The GR2M model gave the best fit in peninsular Spain. The performance indices proved satisfactory in both calibration and validation periods for most humid and sub-humid watersheds, so this model can be used for future water management in climate change scenarios and similar regions as the one studied, within an aridity index framework of around 1. When the catchment is more humid, any water model fits better. It is the opposite in the driest regions, thus, corroborating previous findings. The Témez model, widely used in Spain, only performed well in humid regions, as many of the other water balance models did, but it had the worst results in the dry sub-humid region. It only gave the best fit in the LEM and CUE, but the former showed a high C.V. in AIC and BIC, probably due to a 20% variation of cover land use in the studied period, and the latter had an unsatisfactory result for most of the tests applied.

It is clear that the use of a single comparison method is inaccurate, confusing, and in some cases, useless. When AI_U is higher than 1.5, minor differences in AIC and BIC are found in the model performance; although they are also the basis for the hydrological model selection in the dry and semi-arid regions, these criteria must be used in conjunction with other methods to achieve the best model, especially in humid catchments. We used a grading method based on R^2, NSE, and PBIAS that is complementary to AIC and BIC. Furthermore, the assessment of the margin of error in the total run-off volume using REV is also a key index. However, contrary to that observed in previous tests, the REV results do not highlight specific hydrological models, as well as the criteria above, if they are separately accounted for. These results indicate that it is not adequate to rely on a single hydrological model; also, the choice depends on the purpose of the studies. The uncertainty analysis carried out using the Fiteval software shows that the GR2M model is the best compared to the rest of the models selected; therefore, more than 90% of the confidence intervals are above 0.75 NSE, which shows that the model is very good in humid watersheds. On the contrary, Témez and Thornthwaite-Mather do not reach 0.75 NSE, even in humid catchments

FDCs and their metrics confirm previous results, as the humid and sub-humid region models show good performance in probabilities of exceedance lower than 80% (i.e., in high and medium volumes) but the results worsen the lower the streamflow is. Deviations in humid watersheds range from 0–25% in both high volumes and the mid-segment slope but reach values higher than 80% in low streamflows. Nonetheless, low volumes in most humid watersheds account for less than 5% of the total streamflow in the watersheds, so inaccuracy is admissible to the assessment of interannual water volume purposes. On the contrary, semi-arid regions show good performance in high volumes, which account for only 25% of the total streamflows in the studied period, thus disabling the models selected.

The methodology used can be applied in regions with similar case studies to more accurately assess the resources within a system and provide more sustainable management in a watershed.

Appendix A. Témez Model

This model considers the system to be divided into the upper or non-saturated zone (S), and the lower or saturated zone (G). Some of the precipitation (P) drains directly into the river or through the aquifer, whereas the remainder is converted into ET (real evapotranspiration). Excess (Ex) is divided into run-off (Qs) through river networks at the present time, and infiltration to aquifers, draining one part (Qg) at the present time (month 0), with the rest remaining in the groundwater storage tank (G) for draining at a later date (month i).

Water balance is formed by input flow (P), which is divided into outflow (ET, Qs, and Qg), infiltration in month i (Ii), which is similar to aquifer recharge (Ri), and water storages in soil moisture (Si) and aquifer volume (Vi).

Total excess (Ex) in month i is obtained by Equations (A1)–(A3).

$$Ex_i = 0 \text{ if } Pi \le P0$$

$$Ex_i = \frac{(P_i - P_0)^2}{P_i + \delta - 2*P_0} \text{ if } Pi > P0 \tag{A1}$$

$$\delta = S_{max} - S_{i-1} + ET_i \tag{A2}$$

$$P_0 = C*(S_{max} - S_{i-1}) \tag{A3}$$

where S_{max} is the soil maximum moisture (mm), S_{i-1}, is the soil moisture (mm) at the beginning of the period i, ETP_i is the potential evapotranspiration in the period I, and C is the beginning exceeding coefficient whose range is between 0.2 and 1. Soil moisture at the end of month i is obtained with Equation (A4).

$$S_i = \max(0; S_{i-1} + P_i - Ex_i - ETP_i) \tag{A4}$$

Moreover, ET_i is obtained by Equation (A5) and I_i by Equation (A6).

$$ET_i = \min(S_{i-1} + P_i - Ex_i; \ ETP_i) \tag{A5}$$

$$I_i = I_{max} * \frac{Ex_i}{Ex_i + I_{max}} \tag{A6}$$

Run-off will be calculated as the excess that is not infiltrated, as in Equation (A7).

$$Qs_i = Ex_i - I_i \tag{A7}$$

With respect to the aquifer, two hypotheses are made: (1) It behaves as a lineal water storage (Equation (A8)); and (2) the recharge by infiltration occurs in the middle of the period.

$$Q_i = Q_{i-1} \cdot e^{-\alpha \cdot t} \tag{A8}$$

where Qi is the outflow in time i, α is the aquifer discharge coefficient and t is the time period between i-l and i (Equation (A9)). The hypothesis of lineal water storage implies a relationship between Qi and stored volume in aquifer (G_i) given by Equation (A10).

$$Q_i = (Q_{i-1} \cdot e^{-\alpha \cdot t/2} + \alpha \cdot R_i) e^{-\alpha t/2} \tag{A9}$$

$$Q_i = \alpha \cdot G_i \tag{A10}$$

When applying mass balance equation (Equation (A11)) and combining with previous expressions, Equation (A12) is obtained.

$$I_i - Q_i = \frac{dG_i}{dEx} \tag{A11}$$

$$R_i - \alpha \cdot G_i = \frac{dG_i}{dEx_i} \tag{A12}$$

Aquifer recharge can be expressed as Equation (A13).

$$R_i = \text{Sur} * I_i \tag{A13}$$

where Sur is the basin area.

Finally, Equation (A14) is used to calculate the stored water in the aquifer in month t, Equation (A15) provides aquifer drainage outflow, and total run-off in the monitoring point is obtained by Equation (A16).

$$G_i = G_{i-1} * e^{-\alpha t/2} + \frac{\text{Sur} * I_i}{\alpha} \left(1 - e^{-\alpha t/2}\right) \tag{A14}$$

$$Qg_i = V_{i-1} - V_i + R_i \tag{A15}$$

$$Q_i = Qs_i + Qg_i \tag{A16}$$

Appendix B. ABCD Model

This model (A, B, C, and D parameters) also considers two storage tanks, namely the upper storage (S) or soil tank, and the groundwater storage tank (G). The upper storage tank (S) has two outputs: run-off (Qs) and infiltration. Thus, the model has two inputs, precipitation (P) and ET, and their outputs are soil moisture content at the end of the month (S), monthly available water (W), ET, run-off (Qs), infiltration (I), groundwater run-off (Qb), monthly groundwater storage (G), and total run-off (Q).

The first step in the model consists of assessing the soil moisture at the end of the month (S_i), using Equations (A17)–(A20).

$$W_I = P_i + S_{i-1} \tag{A17}$$

$$Y_i = \frac{W_i + b}{2 * a} - \sqrt{\left(\frac{W_i + b}{2 * a}\right)^2 - \frac{W_i * b}{a}} \tag{A18}$$

$$I = Y_i * e^{\frac{-ETP}{b}} \tag{A19}$$

$$EI_i = Y_i - S_i \tag{A20}$$

The second step is obtaining run-off (Qs) using Equation (A21).

$$Qs = (1 - c) * (W_i - Y_i) \tag{A21}$$

Groundwater run-off is obtained using Equations (A22)–(A24).

$$I = c * (W_i - Y_i) \tag{A22}$$

$$G_i = \frac{(I_i + G_{i-1})}{(1 + d)} \tag{A23}$$

$$I = d * G_i \tag{A24}$$

Finally, total run-off is calculated by adding the surface run-off and groundwater run-off (Equation (A25)).

$$Q_i = Qs_i + Qg_i \tag{A25}$$

Appendix C. GR2M

GR2M transforms precipitation into run-off through the implementation of two equations: production and transfer functions. Initially, P and ET are balanced and precipitation is distributed between the upper storage tank (S), with a limited capacity and groundwater storage tank (G).

Monthly P and ET are fitted using Equations (A26)–(A30).

$$U = \frac{P \cdot ET}{\left(R^{0.5} + E^{0.5}\right)^2} \tag{A26}$$

$$P_n = P - U \tag{A27}$$

$$ET_n = ET - U \tag{A28}$$

$$P_{np} = x_1 \cdot P_n \tag{A29}$$

$$ET_{np} = x_1 \cdot ET_n \tag{A30}$$

where X_1 is one parameter of the model.

Stored moisture in S tank is obtained when using Equation (A31).

$$S1 = \frac{S + A \cdot V}{1 + \frac{S * V}{A}} \tag{A31}$$

where $V = \tan h\left(\frac{P_{np}}{A}\right)$ and A are the maximum storage capacity of tank S.

Likewise, ET causes S1 to become S2 when using Equation (A32).

$$S2 = \frac{S1 * (1 - W)}{1 + W * \left(1 - \frac{S1}{A}\right)} \tag{A32}$$

Rainfall excess is calculated by Equation (A33).

$$Ex = P_{np} + S - S1 \tag{A33}$$

Then, run-off (Qs) is obtained by Equation (A34).

$$Q_s = \alpha * P_e \tag{A34}$$

where α is a parameter of the model.

Stored mosisture in the second tank G (G1 y G2) receives a recharge of $(1 - \alpha) * P_e$ where G (initial stored in the tank) becomes G1, according to Equation (A35).

$$G_1 = G + (1 - \alpha) * P_e \tag{A35}$$

Groundwater run-off (Qg) is obtained by Equation (A36).

$$Q_g = x_2 * G_1 \tag{A36}$$

where X_2 is another parameter of the model.

Finally, stored moisture in the second tank goes from G1 to G2 according to Equation (A37), and monthly total run-off is obtained by Equation (A38).

$$G_2 = G_1 - Q_s \tag{A37}$$

$$Q = Q_s + Q_g \tag{A38}$$

Appendix D. AWBM

This model has three surface water-storage tanks (S1, S2, and S3). The water balance of each is estimated independently, resulting in three surpluses. One part of these surpluses is transformed into run-off (Qs), and the other part percolates to a groundwater storage tank or aquifer (G), which in turn goes to groundwater run-off (Qg). Total flow (Q) is obtained by adding both run-offs.

The maximum capacity of each storage is obtained by Equation (A39).

$$S_1 = \frac{C_{ave} * 0.01}{A_1} S_2 = \frac{C_{ave} * 0.33}{A_2} S_3 = \frac{C_{ave} * 0.66}{A_3} \tag{A39}$$

where Cave is the average soil water capacity and A1, A2, and A3 are the area of each tank.

For each tank P, ET, Ex, I, and total run-off (Q) are calculated according to Equations (A40)–(A43).

$$P_k = A_k * P \tag{A40}$$

$$ET_k = \min(S_k + P_k; A_k * ETP) \tag{A41}$$

$$Ex_k = \max\big(0; S_k + P_k - ET_k - Cap_k\big) \tag{A42}$$

$$S_k = So_k + P_k - ETR_k - Ex_k \tag{A43}$$

where k = 1, 2, and 3, Sok is the initial volume in each tank, and Ak the area of each tank. Run-off (Qs) is obtained using the fraction of total run-off corresponding to the base flow (BFI), using Equations (A44) and (A45).

$$Qs = (1 - BFI) \cdot (Ex_1 + Ex_2 + Ex_3) \tag{A44}$$

$$Ex = (Ex_1 + Ex_2 + Ex_3) \cdot BFI \tag{A45}$$

Then, groundwater run-off is obtained by Equations (A46) and (A47).

$$Qg = Go * K_b \tag{A46}$$

$$Qg_{aquifer} = G_0 + Ex - Qg \tag{A47}$$

where K_b is the recession constant.

Finally, total flow is obtained by adding both run-off from surface (E_{sup}) and groundwater ($E_{aquifer}$) tanks (Equation (A48)).

$$Q_{total} = Q_S + Qg_{aquifer} \tag{A48}$$

Appendix E. Guo-5P

This model has a similar performance to the two-parameter Guo model, and its use is particularly recommended in humid and semi-humid regions. Precipitation and evapotranspiration (P and ET) are the input data, on the basis of which the remaining parameters are estimated: ET, soil water storage (S), water surpluses, surface run-off (Qs), subsurface run-off (Qb), aquifer recharge, groundwater storage (G), groundwater run-off (Qg), and total flow (Q).

Real ET is calculated by Equation (A49).

$$ETR_i = ET_i * K_0 \tag{A49}$$

where K_0 is a parameter of the model.

Soil water storage (Si) is limited by soil maximum moisture (SMAX), which is obtained when using Equations (A50) and (A51)

$$\text{If } Pi > ETRi \text{ then } S_i = S_{i-1} + P_i - ETR_i \tag{A50}$$

$$\text{If } Pi < ETRi \text{ then } S_i = S_{i-1} * e^{\frac{-(ETR-Pi)}{SMAX}} \tag{A51}$$

When P is higher than ET and soil moisture reaches its maximum (SMAX), water excess (Exi) is obtained by Equation (A52).

$$Ex_i = S_i - SMAX \tag{A52}$$

Part of the water excess becomes run-off (QSi), whereas the rest of the water excess (WSi) is divided between subsurface run-off (Qsi) and infiltration (Ii) according to Equations (A53)–(A56).

$$Q_{si} = C \cdot (S_i - SMAX) = C \cdot Ex \tag{A53}$$

$$WS_i = (1 - C) \cdot (S_i - SMAX) = (1 - C) \cdot Ex \tag{A54}$$

$$QI_i = K_1 \cdot WS_i \tag{A55}$$

$$I_i = (1 - K_1) \cdot WS_i \tag{A56}$$

where C and K_1 are parameters of the model and QIi is the sub-superficial run-off.

The model considers that aquifer balance is obtained when using Equation (A57).

$$G_i = G_{i-1} + I_i - K_2 * G_{i-1} \tag{A57}$$

where K_2 is a parameter of the model.

The underground run-off (Qg) is calculated by Equation (A58).

$$Qg_i = K_2 * G_i \tag{A58}$$

Finally, total run-off (Q) is assessed adding run-off, sub-superficial run-off, and underground run-off (Equation (A59)).

$$Q = Q_{si} + QI_i + Qg_{i-1} \tag{A59}$$

Appendix F. Thornthwaite-Mather

The model distinguishes two water storage tanks—surface (S) and groundwater (G)—which lead to the output flow (Q) through different calculations.

A part of P is converted into run-off (Qd) directly (Equation (A60)).

$$Qd = P \propto \tag{A60}$$

The remaining precipitation that comes into the model (Pi) is compared to ET, obtaining soil moisture capacity (S). If Pi is higher than ET, then there is water excess and real evapotranspiration (ETR), and final storage (S_i) is obtained by Equation (A61), and is otherwise obtained using Equation (A62).

$$S_i = \min(S_{i-1} + P_i - ETP); \varnothing) \tag{A61}$$

$$ET_i = S_{i-1} - S + P_i \tag{A62}$$

Thus, if surface storage (S) is full, flow to the tank is calculated by Equation (A63), and is otherwise obtained with $\Delta Q = 0$.

$$\Delta Q = P_i - ETP_i + S_i - 1 - \varnothing \tag{A63}$$

Finally, only one part of the available water in an underground water tank (G) will be added to total run-off (Q), which will be obtained by Equation (A64).

$$G_i = Q(1 - \lambda) \tag{A64}$$

where λ is a parameter of the model.

Therefore, Q is obtained by adding direct and underground run-off (Equation (A65)).

$$Q_t = Qd + Qg \tag{A65}$$

References

1. Wurbs, R.A. Texas water availability modeling system. *J. Water Resour. Plan. Manag.* **2005**, *131*, 270–279. [CrossRef]
2. Senent-Aparicio, J.; Liu, S.; Pérez-Sánchez, J.; López-Ballesteros, A.; Jimeno-Sáez, P. Assessing Impacts of Climate Variability and Reforestation Activities on Water Resources in the Headwaters of the Segura River Basin (SE Spain). *Sustainability* **2018**, *10*, 3277. [CrossRef]
3. Puricelli, M.M. Estimación y Distribución de Parámetros del Suelo Para la Modelación Hidrológica. Ph.D. Thesis, Departamento de Ingeniería Hidráulica y Medio Ambiente, Universidad Politécnica de Valencia, Valencia, Spain, 2003.
4. Essam, M. Water Flow and Chemical Transport in a Subsurface Drained Watershed. Ph.D. Thesis, University of Illinois, Champaign County, IL, USA, 2007.
5. Shimon, C. *Water Resources*; Island Press: Washington, DC, USA, 2010.
6. Unesco. *Methods for Water Balance Computation*; Instituto de Hidrología de España: Madrid, Spain, 1981.
7. Thornthwaite, C.W. A new and improved classification of climates. *Geogr. Rev.* **1948**, *38*, 55–94. [CrossRef]
8. Thornthwaite, C.W.; Mather, J.R. *Instructions and Tables for Computing Potential Evapotranspiration and the Water Balance*; Publications in Climatology; Drexel Institute of Technology, Laboratory of Climatolog: Centerton, NJ, USA, 1957; Volume 10.
9. Alley, W.M. On the treatment of evapotranspiration, soil moisture accounting, and aquifer recharge in monthly water balance models. *Water Resour. Res.* **1984**, *20*, 1137–1149. [CrossRef]
10. Xiong, L.; Guo, S. A two-parameter monthly water balance model and its application. *J. Hydrol.* **1999**, *216*, 111–123. [CrossRef]
11. Makhlouf, Z.; Michel, C. A two-parameter monthly water balance model for French watersheds. *J. Hydrol.* **1994**, *162*, 299–318. [CrossRef]
12. Giakoumakis, S.; Tsakiris, G.; Efremides, D. *On the Rainfall-Run-Off Modelling in a Mediterranean Island Environment*; Advances in Water Resources Technology; Tsakiris, G., Ed.; Balkema: Rotterdam, The Netherlands, 1991.
13. Mimikou, M.A.; Hadjisawa, P.S.; Kouvopoulos, Y.S.; Afrateos, H. Regional climate change impacts: II. Impacts on water management works. *Hydrol. Sci. J.* **1991**, *36*, 259–270. [CrossRef]
14. Alley, W.M. Water balance models in one-month-ahead streamflow forecasting. *Water Resour. Res.* **1985**, *21*, 597–606. [CrossRef]
15. Abulohom, M.S.; Shah, S.M.S.; Ghumman, A.R. Development of a rainfall-run-off model, its calibration and validation. *Water Resour. Manag.* **2001**, *15*, 149–163. [CrossRef]
16. Yates, D.; Strzepek, K. Modeling the Nile basin under climate change. *J. Hydrol. Eng.* **1998**, *3*, 98–108. [CrossRef]
17. Vandewiele, G.; Xu, C.-Y. Methodology and comparative study of monthly water balance models in Belgium, China and Burma. *J. Hydrol.* **1992**, *134*, 315–347. [CrossRef]
18. Andréassian, V.; Hall, A.; Chahinian, N.; Schaake, J. Introduction and Synthesis: Why Should Hydrologists Work on A Large Number of Basin Data Sets? In *Large Sample Basin Experiments for hydrological Model Parameterization, Results of the Model Parameter Experiment—MOPEX*; IAHS Publication No. 307; IAHS Press: Paris, France, 2006.

19. Perrin, C.; Michel, C.; Andreassian, V. Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. *J. Hydrol.* **2001**, *242*, 275–301. [CrossRef]

20. Chiew, F.H.S.; McMahon, T.A. *Australian Data for Rainfall-Run-Off Modelling and the Calibration of Models Against Streamflow Data Recorded Over Different Time Periods*; Civil Engineering Transactions, The Institution of Engineers: Sydney, Australia, 1993; Volume 35, pp. 261–274.

21. World Meteorological Organization. *Intercomparison of Conceptual Models Used in Operational Forecasting*; Operational Hydrology Report No. 7, WMO-No. 429; WMO: Geneva, Switzerland, 1975.

22. Xu, C.Y.; Singh, V.P. Dependence of evaporation on meteorological variables at different time-scales and intercomparison of estimation methods. *Hydrol. Process.* **1998**, *12*, 429–442. [CrossRef]

23. Todini, E. Rainfall-runoff modeling—past present and future. *J. Hydrol.* **1998**, *100*, 341–352. [CrossRef]

24. Chiew, F.H.S.; Kirono, D.G.C.; Kent, D.M.; Frost, A.J.; Charles, S.P.; Timbal, B.; Nguyen, K.C.; Fu, G. Comparison of run-off modelled using rainfall from different downscaling methods for historical and future climates. *J. Hydrol.* **2010**, *387*, 10–23. [CrossRef]

25. Zhang, Q.; Liu, J.; Singh, V.P.; Gu, X.; Chen, X. Evaluation of impacts of climate change and human activities on streamflow in the Poyang Lake basin, China. *Hydrol. Process.* **2016**, *30*, 2562–2576. [CrossRef]

26. Wang, D.; Tang, Y. A one-parameter Budyko model for water balance captures emergent behavior in Darwinian hydrologic models. *Geophys. Res. Lett.* **2014**, *41*, 4569–4577. [CrossRef]

27. Sankarasubramanian, A.; Vogel, R.M. Annual hydroclimatology of the United States. *Water Resour. Res.* **2002**, *38*, 19-1–19-2. [CrossRef]

28. Lacombe, G.; Ribolzi, O.; De Rouw, A.; Pierret, A.; Latsachak, K.; Silvera, N.; Dinh, R.P.; Orange, D.; Janeau, J.L.; Soulileuth, B.; et al. Contradictory hydrological impacts of afforestation in the humid tropics evidenced by long-term field monitoring and simulation modelling. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 2691–2704. [CrossRef]

29. Mouelhi, S.; Michel, C.; Perrin, C.; Andreassian, V. Stepwise development of a two parameter monthly water balance model. *J. Hydrol.* **2006**, *318*, 200–214. [CrossRef]

30. Burnash, R.J.C.; Ferral, R.L.; Maguire, R.A. *A Generalized Streamflow Simulation System: Conceptual Models for Digital Computers*; Joint Federal State River Forecast Center: Sacramento, CA, USA, 1973.

31. Guo, S. *Impact of Climate Change on Hydrological Balance and Water Resource Systems in the Dongjiang Basin, China*; Modeling and Management of Sustainable Basin-Scale Water Resource Systems (Proceedings of a Boulder Symposium); LAHS Publ.: Los Alamos, NM, USA, 1995; Volume 231.

32. Singh, K.; Kumar, A. Evaluation of relief aspects morphometric parameters derived from different sources of DEMs and its effects over time of concentration of run-off (TC). *Earth Sci. Inform.* **2016**. [CrossRef]

33. Singh, S.K. Transmuting synthetic unit hydrograph into gamma distribution. *J. Hydrol. Eng.* **2000**, *5*, 380–385. [CrossRef]

34. Ferrer, J. *Recomendaciones para el Cálculo Hidrometeorológico de Avenidas*; CEDEX M-37, CEDEX: Madrid, Spain, 1993.

35. Témez, J.R. Extended and Improved Rational Method: Version of the Highways Administration of Spain. In Proceedings of the XXIV Congress of IAHR, Madrid, Spain, 9–13 September 1991; Volume A, pp. 33–40.

36. Témez, J.R. *Cálculo Hidrometeorológico de Caudales Máximos en Pequeñas Cuencas Naturales*; Dirección General de Carreteras, MOPU: Madrid, Spain, 1987.

37. Lyon, S.W.; Walter, M.T.; Gerard-Marchant, P.; Steenhuis, T.S. Using a topographic index to distribute variable source area run-off predicted with the SCS curve-number equation. *Hydrol. Proc.* **2004**, *18*, 2757–2771. [CrossRef]

38. Frankenberger, J.R.; Brooks, E.S.; Walter, M.T.; Walter, M.F.; Steenhuis, T.S. A GIS-Based Variable Source Area Hydrology Model. *Hydrol. Process.* **1999**, *13*, 804–822. [CrossRef]

39. Calvo, J.C. An evaluation of Thornthwaite's water balance technique in predicting stream run-off in Costa Rica. *Hydrol. Sci. J.* **1986**, *31*, 51–60. [CrossRef]

40. Croke, B.F.W.; Andrews, F.; Jakeman, A.J.; Cuddy, S.M.; Luddy, A. IHACRES classic plus: A redesign of the IHACRES rainfall-run-off model. *Environ. Model. Softw.* **2006**, *21*, 426–427. [CrossRef]

41. Chiew, F.H.S.; McMahon, T.A. Modelling the impacts of climate change on Australian streamflow. *Hydrol. Process.* **2002**, *16*, 1235–1245. [CrossRef]

42. Perrin, C.; Michel, C.; Andréassian, V. Improvement of a parsimonious model for streamflow simulation. *J. Hydrol.* **2003**, *279*, 275–289. [CrossRef]

43. Boughton, W. New approach to calibration of the AWBM for use on ungauged catchments. *J. Hydrol. Eng.* **2009**, *14*, 562–568. [CrossRef]

44. Boughton, W. Effect of data length on rainfall–run-off modelling. *Environ. Model. Softw.* **2007**, *22*, 406–413. [CrossRef]

45. Boughton, W. Calibrations of a daily rainfall–run-off model with poor quality data. *Environ. Model. Softw.* **2006**, *21*, 1114–1128. [CrossRef]

46. Boughton, W. The Australian Water Balance Model. *Environ. Model. Softw.* **2004**, *19*, 943–956. [CrossRef]

47. Boughton, W.; Chiew, F. Estimating run-off in ungauged catchments from rainfall, PET and the AWBM model. *Environ. Model. Softw.* **2007**, *22*, 476–487. [CrossRef]

48. O'Connell, P.E.; Nash, J.E.; Farrell, J.P. River flow forecasting through conceptual models, Part 2, The Brosna catchment at Ferbane. *J. Hydrol.* **1970**, *10*, 317–329. [CrossRef]

49. Singh, V.P. Watershed Modeling. In *Computer Models of Watershed Hydrology*, 1st ed.; Water Resources Publications: Highlands Ranch, CO, USA, 1995.

50. Singh, V.P.; Frevert, D.K. *Mathematical Models of Small Watershed Hydrology and Applications*; Water Resources Publications: Littleton, CO, USA, 2002.

51. Eder, G.; Fuchs, M.; Nachtnebel, H.; Loibl, W. Semi-distributed modelling of the monthly water balance in an alpine catchment. *Hydrol. Process.* **2005**, *19*, 2339–2360. [CrossRef]

52. Arnold, J.G.; Srinivasan, R.; Muttiah, R.S.; Williams, J.R. Large-area hydrologic modeling and assessment: Part I. Model development. *J. Am. Water Resour. Assoc.* **1998**, *34*, 73–89. [CrossRef]

53. Beven, K.J. *Rainfall-Run-off Modelling: The Primer*; Wiley: Chichester, UK, 2001.

54. Ewen, J.; Parkin, G. Validation of catchment models for prediction land use and climate change impact. 1. Method. *J. Hydrol.* **1996**, *175*, 583–594. [CrossRef]

55. Carpenter, T.M. Discretization scale dependencies of the ensemble flow range versus catchment area relationship in distributed hydrologic modeling. *J. Hydrol.* **2006**, *328*, 242–257. [CrossRef]

56. Viney, N.R.; Croke, B.F.W.; Breuer, L.; Bormann, H.; Bronstert, A.; Frede, H.; Gräff, T.; Hubrechts, L.; Huisman, J.A.; Jakeman, A.J.; et al. Ensemble modelling of the hydrological impacts of land use change. In Proceedings of the MODSIM05—International Congress on Modelling and Simulation: Advances and Applications for Management and Decision Making, Melbourne, Australia, 12–15 December 2005; Zerger, A., Argent, R.M., Eds.; ANU: Canberra, Australia, 2005; pp. 2967–2973.

57. Croke, B.F.W.; Merritt, W.S.; Jakeman, A.J. A dynamic model for predicting hydrologic response to land cover changes in gauged and ungauged catchments. *J. Hydrol.* **2004**, *291*, 115–131. [CrossRef]

58. Vansteenkiste, T.; Tavakoli, M.; Ntegeka, V.; De Smedt, F.; Batelaan, O.; Pereira, F.; Willems, P. Intercomparison of hydrological model structures and calibration approaches in climate scenario impact projections. *J. Hydrol.* **2014**, *519*, 743–755. [CrossRef]

59. Samaniego, L.; Kumar, R.; Attinger, S. Multiscale parameter regionalization of a grid–based hydrologic model at the mesoscale. *Water Resour. Res.* **2010**, *46*, W05523. [CrossRef]

60. Beck, H.E.; van Dijk, A.I.J.M.; de Roo, A.; Miralles, D.G.; McVicar, T.R.; Schellekens, J.; Bruijnzeel, L.A. Global-scale regionalization of hydrologic model parameters. *Water Resour. Res.* **2016**, *52*, 3599–3622. [CrossRef]

61. Yang, X.; Michel, C. Flood forecasting with a watershed model: A new method of parameter updating. *Hydrol. Sci. J.* **2000**, *45*, 537–546. [CrossRef]

62. Cameron, D.S.; Beven, K.J.; Tawn, J.; Blazkova, S.; Naden, P. Flood frequency estimation by continuous simulation for a gauged upland catchment (with uncertainty). *J. Hydrol.* **1999**, *219*, 169–187. [CrossRef]

63. Uhlenbrock, S.; Seibert, J.; Leibundgut, C.; Rohde, A. Prediction uncertainty of conceptual rainfall-run-off models caused by problems in identifying model parameters and structure. *Hydrol. Sci. J.* **1999**, *44*, 779–797. [CrossRef]

64. Yang, X.; Parent, E.; Michel, C.; Roche, P.A. Comparison of real-time reservoir operation techniques. *J. Water Resour. Plan. Manag.* **1995**, *121*, 345–351. [CrossRef]

65. Lobligeois, F.; Andréassian, V.; Perrin, C.; Tabary, P.; Loumagne, C. When does higher spatial resolution rainfall information improve streamflow simulation? An evaluation on 3620 flood events. *Hydrol. Earth Syst. Sci.* **2013**, *10*, 12485–12536. [CrossRef]

66. Smith, M.B.; Koren, V.; Zhang, Z.; Zhang, Y.; Reed, S.M.; Cui, Z.; Moreda, F.; Cosgrove, B.A.; Mizukami, N.; Anderson, E.A. Results of the DMIP 2 Oklahoma experiments. *J. Hydrol.* **2012**, *418–419*, 17–48. [CrossRef]

67. Apip; Sayama, T.; Tachikawa, Y.; Takara, K. Spatial lumping of a distributed rainfall sediment-run-off model and its effective lumping scale. *Hydrol. Process.* **2012**, *26*, 855–871. [CrossRef]

68. Breuer, L.; Huisman, J.A.; Willems, P.; Bormann, H.; Bronstert, A.; Croke, B.F.W.; Frede, H.-G.; Gräff, T.; Hubrechts, L.; Jakeman, A.J.; et al. Assessing the impact of land use change on hydrology by ensemble modeling (LUCHEM) I: Model intercomparison of current land use. *Adv. Water Resour.* **2009**, *32*, 129–146. [CrossRef]

69. Zhang, Z.; Koren, V.; Smith, M.; Reed, S.; Wang, D. Use of next generation weather radar data and basin disaggregation to improve continuous hydrograph simulations. *J. Hydrol. Eng.* **2004**, *9*, 103–115. [CrossRef]

70. Koren, V.; Reed, S.; Smith, M.; Zhang, Z.; Seo, D.J. Hydrology laboratory research modeling system (HL-RMS) of the US national weather service. *J. Hydrol.* **2004**, *291*, 297–318. [CrossRef]

71. Ajami, N.K.; Gupta, H.; Wagener, T.; Sorooshian, S. Calibration of a semidistributed hydrologic model for streamflow estimation along a river system. *J. Hydrol.* **2004**, *298*, 112–135. [CrossRef]

72. Reed, S.; Koren, V.; Smith, M.B.; Zhang, Z.; Moreda, F.; Seo, D.; Dmipparticipants, A. Overall distributed model intercomparison project results. *J. Hydrol.* **2004**, *298*, 27–60. [CrossRef]

73. Boyle, D.P.; Gupta, H.V.; Sorooshian, S.; Koren, V.; Zhang, Z.; Smith, M. Toward improved streamflow forecasts: Value of semidistributed modeling. *Water Resour. Res.* **2001**, *37*, 2749–2759. [CrossRef]

74. Refsgaard, J.C.; Knudsen, J. Operational validation and intercomparison of different types of hydrological models. *Water Resour. Res.* **1996**, *32*, 2189–2202. [CrossRef]

75. Shah, S.M.S.; O'Connell, P.E.; Hosking, J.R.M. Modelling the effects of spatial variability in rainfall on catchment response. 2. Experiments with distributed and lumped models. *J. Hydrol.* **1996**, *175*, 89–111. [CrossRef]

76. Bormann, H.; Breuer, L.; Giertz, S.; Huisman, J.A.; Viney, N.R. Spatially explicit versus lumped models in catchment hydrology—Experiences from two case studies. In *Uncertainties in Environmental Modelling and Consequences for Policy Making. NATO Science for Peace and Security Series C: Environmental Security*; Baveye, P.C., Laba, M., Mysiak, J., Eds.; Springer: Dordrecht, The Netherlands, 2009.

77. Vansteenkiste, T.; Tavakoli, M.; Van Steenbergen, N.; De Smedt, F.; Batelaan, O.; Pereira, F.; Willems, P. Intercomparison of five lumped and distributed models for catchment run-off and extreme flow simulation. *J. Hydrol.* **2014**, *511*, 335–349. [CrossRef]

78. Martínez-Santos, P.; Andreu, J.M. Lumped and distributed approaches to model natural recharge in semiarid karst aquifers. *J. Hydrol.* **2010**, *388*, 389–398. [CrossRef]

79. Haque, M.M.; Rahman, A.; Hagare, D.; Kibria, G.; Karim, F. Estimation of catchment yield and associated uncertainties due to climate change in a mountainous catchment in Australia. *Hydrol. Process.* **2015**, *29*, 4339–4349. [CrossRef]

80. Jiang, T.; Chen, Y.Q.; Xu, C.Y.; Chen, X.H.; Chen, X.; Singh, V.P. Comparison of hydrological impacts of climate change simulated by six hydrological models in the Dongjiang Basin, South China. *J. Hydrol.* **2007**, *336*, 316–333. [CrossRef]

81. Kumar, A.; Singh, R.; Jena, P.P.; Chatterjee, C.; Mishra, A. Identification of the best multi-model combination for simulating river discharge. *J. Hydrol.* **2015**, *525*, 313–325. [CrossRef]

82. Van Esse, W.R.; Perrin, C.; Booij, M.J.; Augustijn, D.C.M.; Fenicia, F.; Kavetski, D.; Lobligeois, F. The influence of conceptual model structure on model performance: A comparative study for 237 French catchments. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 4227–4239. [CrossRef]

83. Velázquez, J.A.; Anctil, F.; Perrin, C. Performance and reliability of multimodel hydrological ensemble simulations based on seventeen lumped models and a thousand catchments. *Hydrol. Earth Syst. Sci.* **2010**, *14*, 2303–2317. [CrossRef]

84. Broderick, C.; Matthews, T.; Wilby, R.L.; Bastola, S.; Murphy, C. Transferability of hydrological models and ensemble averaging methods between contrasting climatic periods. *Water Resour. Res.* **2016**, *52*, 8343–8373. [CrossRef]

85. Bourgin, F.; Andréassian, V.; Perrin, C.; Oudin, L. Transferring global uncertainty estimates from gauged to ungauged catchments. *Hydrol. Earth Syst. Sci.* **2015**, *19*, 2535–2546. [CrossRef]

86. Ye, W.; Bates, B.; Viney, N.; Sivapalan, M.; Jakeman, A. Performance of conceptual rainfall-run-off, models in low-yielding ephemeral catchments. *Water Resour. Res.* **1997**, *33*, 153–166. [CrossRef]

87. Kirchner, J.W. Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology. *Water Resour. Res.* **2006**, *42*, W03S04. [CrossRef]

88. Escriva-Bou, A.; Pulido-Velazquez, M.; Pulido-Velazquez, D. Economic value of climate change adaptation strategies for water management in Spain's Jucar basin. *J. Water Res. Plan. ASCE* **2017**, *143*, 04017005. [CrossRef]

89. *MIMAM, Libro Blanco del Agua en España*; Centro de Publicaciones del Ministerio de Medio Ambiente: Madrid, Spain, 2000.

90. Cabezas, F.; Estrada, F.; Estrela, T. Algunas contribuciones técnicas del Libro Blanco del Agua en España. *Ingeniería Civil* **1999**, *115*, 79–96.

91. Estrela, T.; Cabezas, F.; Estrada, F. La evaluación de los recursos hídricos en el libro blanco del agua en España. *Ingeniería del Agua* **1999**, *6*, 125–138. [CrossRef]

92. Estrela, T. *Modelos Matemáticos para la Evaluación de Recursos Hídricos*; Centro de Estudios Hidrográficos y Experimentación de Obras Públicas, CEDEX: Madrid, Spain, 1992.

93. Fernandez, W.; Vogel, R.; Sankarasubramanian, A. Regional calibration of a watershed model. *Hydrol. Sci. J.* **2000**, *45*, 689–707. [CrossRef]

94. Martinez, G.F.; Gupta, H.V. Toward improved identification of hydrological models: A diagnostic evaluation of the "abcd" monthly water balance model for the conterminous United States. *Water Resour. Res.* **2010**, *46*, W08507. [CrossRef]

95. Sankarasubramanian, A.; Vogel, R.M. Hydroclimatology of the continental United States. *Geophys. Res. Lett.* **2003**, *30*. [CrossRef]

96. Thomas, H. *Improved Methods for National Water Assessment*; Report WR15249270; US Water Resource Council: Washington, DC, USA, 1981.

97. Pellicer-Martinez, F.; Martínez-Paz, J.M. Contrast and transferability of parameters of lumped water balance models in the Segura River Basin (Spain). *Water Environ. J.* **2015**, *29*, 43–50. [CrossRef]

98. Coron, L.; Thirel, G.; Delaigue, O.; Perrin, C.; Andréassian, V. The Suite of Lumped GR Hydrological Models in an R Package. *Environ. Model. Softw.* **2017**, *94*, 166–171. [CrossRef]

99. Wriedt, G.; Bouraoui, F. *Towards A General Water Balance Assessment of Europe*; Joint Research Centre—Institute for Environment and Sustainability, Office for Official Publications of the European Communities: Luxembourg, 2009.

100. Sharifi, F.A.S.; Safarpour, S.; Ayubzadeh, S.A. Evaluation of AWBM 2002 simulation model in 6 Iranian representative catchments. *Pajouhesh-Va-Sazandegi. Nat. Resour.* **2004**, *17*, 35–42.

101. Sharafi, F.; Adamowski, J.; Barkhordari, J.; Saadat, H. Decision Support Tool for Evaluating Changes in Arid and Tropical Watersheds. *J. Agric. Eng.* **2012**, *49*, 33.

102. Makungo, R.; Odiyo, J.O.; Ndiritu, J.G.; Mwaka, N. Rainfall-run-off modelling approach for ungauged catchments: A case study of Nzhelele River subquaternary catchment. *Phys. Chem. Earth* **2010**, *35*, 596–607. [CrossRef]

103. Peranginangin, N.; Sakthivadievel, R.; Scott, N.R.; Kendy, E.; Steenhuis, T.S. Water accounting for conjunctive groundwater/surface water management: Case of the Singkarak-Ombilin River basin, Indonesia. *J. Hydrol.* **2004**, *292*, 1–22. [CrossRef]

104. Barros, R.; Isidoro, D.; Aragüés, R. Long-term water balances in La Violada irrigation district (Spain): I. Sequential assessment and minimization of closing errors. *Agric. Water Manag.* **2011**, *102*, 35–45. [CrossRef]

105. Sharifi, M.A.; Rodriguez, E. Design and development of a planning support system for policy formulation in water resources rehabilitation: The case of Alcázar De San Juan District in Aquifer 23, La Mancha, Spain. *J. Hydroinform.* **2002**, *4*, 157–175. [CrossRef]

106. Estrela, T.; Sahuquillo, A. Modeling the Response of a Karstic Spring at Arteta Aquifer in Spain. *Groundwater* **1997**, *35*, 18–24. [CrossRef]

107. Donker, N. WTRBLN, a computer program to calculate water balance. *Comput. Geosci.* **1987**, *13*, 95–122. [CrossRef]

108. Pulido-Velázquez, D.; Ahlfeld, D.; Andreu, J.; Sahuquillo, A. Reducing the computational cost of unconfined groundwater flow in conjunctive-use models at basin scale assuming linear behaviour: The case of Adra-Campo de Dalías. *J. Hydrol.* **2008**, *353*, 159–174. [CrossRef]

109. Pulido-Velazquez, D.; Sahuquillo, A.; Andreu, J.; Pulido-Velazquez, M. An efficient conceptual model to simulate surface water body-aquifer interaction in Conjunctive Use Management Models. *Water Resour. Res.* **2007**, *43*, W07407. [CrossRef]

110. Köppen, W. *Das Geographische System der Klimate, Handbuch der Klimatologie, 1*; Köppen, W., Geiger, R., Eds.; Borntraeger: Berlin, Germany, 1936.

111. Köppen, W. *Klassification der KlimatenachTemperatur, Niederschlag and Jahreslauf*; Petermanns Geographische Mitteilungen: Gotha, Germany, 1918; pp. 193–203, 243–248.

112. Köppen, W. Die Wärmezonen der Erde, nach der Dauer der heissen, gemässigten und kaltenZeit und nach der Wirkung der Wärme auf die organische Welt betrachtet. *Meteorol. Z.* **1884**, *1*, 5–226.

113. UNEP (United Nations Environment Programme). *World Atlas of Desertification*, 2nd ed.; UNEP: London, UK, 1997.

114. Álvarez, J.; Sanchez, A.; Quintas, L. SIMPA, a GRASS based Tool for Hydrological Studies. *Int. J. Geo-Inf.* **2005**, *1*, 1.

115. Estrela, T.; Quintas, L. A distributed hydrological model for water resources assessment in large basins. In Proceedings of the Rivertech '96-1st International Conference on New/Emerging Concepts for Rivers, Chicago, IL, USA, 22–26 September 1996; Volumes 1–2, pp. 861–868.

116. Pérez-Martín, M.A.; Estrela, T.; Andreu, J.; Ferrer, J. Modeling water resources and river-aquifer interaction in the Júcar River basin. Spain. *Water Resour. Manag.* **2014**, *28*, 4337–4358. [CrossRef]

117. Gonzalez-Zeas, D. Análisis Hidrológico de los Escenarios de Cambio Climático en España. Ph.D. Thesis, Technical Universtiy of Madrid, Madrid, Spain, 2010.

118. Belmar, O.; Velasco, J.; Martinez-Capel, F. Hydrological classification of natural flow regimes to support environmental flow assessments in intensively regulated Mediterranean rivers, Segura River basin (Spain). *J. Environ. Manag.* **2011**, *47*, 992–1004. [CrossRef]

119. eea.europa.eu [Internet]. European Environmental Agency, 2019. Available online: https://www.eea.europa.eu/publications/COR0-landcover (accessed on 10 February 2019).

120. ign.es [Internet]. Instituto Geográfico Nacional, 2019. Available online: http://www.ign.es/web/ign/portal (accessed on 12 February 2019).

121. Akaike, H. Information theory and an extension of the maximum likelihood principle. In Proceedings of the 2nd International Symposium on Information Theory, USSR, Tsahkadsor, Armenia, 2–8 September 1971; Petrov, B.N., Csáki, F., Eds.; Akadémiai Kiadó: Budapest, Hungary, 1973; pp. 267–281.

122. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [CrossRef]

123. Fabozzi, F.J.; Focardi, S.M.; Svetlozar, T.; Bala, G. *The Basics of Financial Econometrics: Tools, Concepts, and Asset Management Applications*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2014.

124. Moriasi, D.N.; Gitau, M.W.; Pai, N.; Daggupati, P. Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria. *Trans. ASABE* **2015**, *58*, 1763–1785. [CrossRef]

125. Bressiani, D.A.; Srinivasan, R.; Jones, C.A.; Mendiondo, E.M. Effects of spatial and temporal weather data resolutions on streamflow modeling of a semi-arid basin, Northeast Brazil. *Int. J. Agric. Biol. Eng.* **2015**, *8*, 125–139.

126. Ritter, A.; Muñoz-Carpena, R. Predictive ability of hydrological models: Objective assessment of goodness-of-fit with statistical significance. *J. Hydrol.* **2013**, *480*, 33–45. [CrossRef]

127. Témez, J.R. *Modelo Matemático de Trasformación "Precipitación-Escorrentía"*; Asociación de Investigación Industrial Eléctrica (ASINEL): Madrid, Spain, 1977.

128. Edijatno, N.; Michel, C. Un modèle pluie-débit journalier a trois paramètres. *LHBL* **1989**, *2*, 113–121. [CrossRef]

129. Kabouya, M. Modélisation pluie-débit aux pas de temps mensuel et annuel en Algérie septentrionale. Ph.D. Thesis, Université Paris Sud Orsay, Orsay, France, 1990.

130. Paturel, J.E.; Servat, E.; Vassiliadis, A. Sensitivity of conceptual rainfall-run-off algorithms to errors in input data. Case of the GR2M model. *J. Hydrol.* **1995**, *168*, 111–125. [CrossRef]

131. Nief, H.; Paturel, J.E.; Servat, E. Study of parameter stability of a lumped hydrologic model in a context of climatic variability. *J. Hydrol.* **2003**, *278*, 213–230.

132. Thornthwaite, C.W.; Mather, J.R. The water balance. *Publ. Climatol.* **1955**, *8*, 1–104.

133. KlemeŠ, V. Operational testing of hydrological simulation models. *Hydrol. Sci. J.* **1986**, *31*, 13–24. [CrossRef]

134. Fylstra, D.; Lasdon, L.; Watson, J.; Waren, A. Design and use of the Microsoft Excel Solver. *Interfaces* **1998**, *28*, 29–55. [CrossRef]

135. Ríos, S. *Investigación Operativa*; Centro de Estudios Ramón Areces: Madrid, Spain, 1988.

136. Abadie, J. The GRG Method for Nonlinear Programming. In *Design and Implementation of Optimization Software*; Greenberg, H.J., Ed.; Sijthoff and Noordhoof: Alphen aan den Rijn, The Netherlands, 1978.

137. Lasdon, L.S.; Waren, A.D. Generalized Reduced Gradient Software for Linearly and Nonlinearly Constrained Problems. In *Design and Implementation of Optimization Software*; Greenberg, H.J., Ed.; Sijthoff and Noordhoof: Alphen aan den Rijn, The Netherlands, 1978.

138. Lasdon, L.S.; Waren, A.D.; Jain, A.; Ratner, M. Design and Testing of a Generalized Reduced Gradient Code for Nonlinear Constrained Programming. *ACM Trans. Math. Softw.* **1978**, *4*, 34–50. [CrossRef]

139. Nuno, R.C.; Lopes, Z.; Pereira, Z.; Pereira, L. Multiple response optimization: A global criterion-based method. *J. Chemom.* **2010**, *24*, 333–342. [CrossRef]

140. Ravinder, H.V. Determining the Optimal Values of Exponential Smoothing Constants—Does Solver Really Work? *AJBE* **2013**, *6*, 347–360. [CrossRef]

141. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models: Part 1. A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290. [CrossRef]

142. Pearson, K. Notes on regression and inheritance in the case of two parents. *Proc. R. Soc. Lond.* **1985**, *58*, 240–242.

143. 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]

144. Karpouzos, D.K.; Baltas, E.A.; Kavalieratou, S.; Babajimopoulos, C. A hydrological investigation using a lumped water balance model: The Aison River Basin case (Greece). *Water Environ. J.* **2011**, *25*, 297–307. [CrossRef]

145. 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*, W09417. [CrossRef]

146. Shafifii, M.; Tolson, B.A. Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. *Water Resour. Res.* **2015**, *5*, 3796–3814. [CrossRef]

147. Bai, P.; Liu, X.; Liang, K.; Liu, C. Comparison of performance of twelve monthly balance models in different climatic catchments of China. *J. Hydrol.* **2015**, *529*, 1030–1040. [CrossRef]

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

149. Jakeman, A.J.; Hornberger, G.M. How much complexity is warranted in a rainfall-run-off model? *Water Resour. Res.* **1993**, *29*, 2637–2649. [CrossRef]

150. Michaud, J.; Sorooshian, S. Comparison of simple versus complex distributed run-off models on a midsized semiarid watershed. *Water Resour. Res.* **1994**, *30*, 593–605. [CrossRef]

151. Beven, K. Changing ideas in hydrology—The case of physically-based models. *J. Hydrol.* **1989**, *105*, 157–172. [CrossRef]

152. Yew Gan, T.; Dlamini, E.M.; Biftu, G.F. Effects of model complexity and structure, data quality, and objective functions on hydrologic modeling. *J. Hydrol.* **1997**, *192*, 81–103. [CrossRef]

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

154. Legates, D.R.; McCabe, G.J. Evaluating the use of "goodness–of–fit" measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **1999**, *35*, 233–241. [CrossRef]

155. Cho, H.; Olivera, F. Effect of the spatial variability of land use, soil type, and precipitation on streamflows in small watersheds. *JAWRA* **2009**, *45*, 1423–1431. [CrossRef]

156. El-Nasr, A.; Arnold, J.G.; Feyen, J.; Berlamont, J. Modelling the hydrology of a catchment using a distributed and a semi-distributed model. *Hydrol. Process.* **2005**, *19*, 573–587. [CrossRef]

157. Booij, M.J. Impact of climate change on river flooding assessed with different spatial model resolutions. *J. Hydrol.* **2005**, *303*, 176–198. [CrossRef]

158. Beven, K. A Manifesto for the Equifinality Thesis. *J. Hydrol.* **2006**, *320*, 18–36. [CrossRef]

159. Operacz, A.; Wałega, A.; Cupak, A.; Tomaszewska, B. The comparison of environmental flow assessment—The barrier for investment in Poland or river protection? *J. Clean. Prod.* **2018**, *193*, 575–592. [CrossRef]

160. Pusłowska-Tyszewska, D.; Tyszewski, S. Attempt at Implementing the 2015 "Ecological Flow Assessment Method for Poland" In the Wieprza River Catchment. *Acta Sci. Pol. Form. Circumiectus* **2018**, *17*, 181–193. [CrossRef]

161. Neitsch, S.; Arnold, J.; Kiniry, J.E.A.; Srinivasan, R.; Williams, J. Soil and Water Assessment Tool User's Manual: Version 2012. Available online: https://swat.tamu.edu/docs/ (accessed on 12 February 2019).