

Article



# Applying a Multi-Model Ensemble Method for Long-Term Runoff Prediction under Climate Change Scenarios for the Yellow River Basin, China

Linus Zhang<sup>1</sup> and Xiaoliu Yang<sup>2,\*</sup>

- <sup>1</sup> Department of Water Resources Engineering, Lund University, Box-118, 221 00 Lund, Sweden; Linus.zhang@tvrl.lth.se
- <sup>2</sup> College of Urban and Environmental Sciences, Peking University, Beijing 100871, China
- \* Correspondence: xlyang@urban.pku.edu.cn; Tel.: +86-10-6275-9327

Received: 3 January 2018; Accepted: 8 March 2018; Published: 10 March 2018

Abstract: Given the substantial impacts that are expected due to climate change, it is crucial that accurate rainfall–runoff results are provided for various decision-making purposes. However, these modeling results often generate uncertainty or bias due to the imperfect character of individual models. In this paper, a genetic algorithm together with a Bayesian model averaging method are employed to provide a multi-model ensemble (MME) and combined runoff prediction under climate change scenarios produced from eight rainfall–runoff models for the Yellow River Basin. The results show that the multi-model ensemble method, especially the genetic algorithm method, can produce more reliable predictions than the other considered rainfall–runoff models. These results show that it is possible to reduce the uncertainty and thus improve the accuracy for future projections using different models because an MME approach evens out the bias involved in the individual model. For the study area, the final combined predictions reveal that less runoff is expected under most climatic scenarios, which will threaten water security of the basin.

**Keywords:** water security; climate change; rainfall–runoff models; multi-model ensemble method; simulation; Yellow River Basin; genetic algorithms

# 1. Introduction

It is now widely acknowledged that climate change will produce significant effects on the hydrological cycle [1–4]. In recent decades, with outputs from climate models made available, hydrological impacts and responses amid a changing climate with respect to a number of river basins in the world have been studied [5–8].

Recently, climate-change impact studies have started to systematically consider associated uncertainties [7], and rainfall–runoff modeling is recognized as one of the most important sources of uncertainty [9]. It has been reported that the multi-model ensemble (MME) strategy is an efficient method to tackle the uncertainty of rainfall–runoff models [10,11].

All models are imperfect representations of real world processes. Different models have strengths in capturing different aspects of such processes. It is therefore highly desirable, in order to reduce the above-mentioned uncertainties and improve overall accuracy, that the best performing parts of various individual models are combined so that a prediction consensus can be reached. With a multi-model ensemble approach, more reliable runoff predictions can be made from multiple competing predictions made from a number of rainfall–runoff models [11,12]. This method was discussed and used in the pioneering works of [13–16] and others. Shamseldin et al. [12] promoted the concept of combining outputs from various rainfall–runoff models to produce an overall combined output to be used as an alternative to the output of a single individual rainfall–runoff model. A recent application of

this concept to future climate projections can be found in [17,18], where surface air temperature and precipitation are predicted.

The aim of this paper was to tackle the discrepancy of rainfall–runoff modeling in changing climate scenarios. This was done by combining three climate scenarios with eight carefully selected rainfall–runoff models. The final MME approaches were then applied to the Yellow River Basin (YRB) for two selected future target periods. The paper is organized as follows: Section 2 describes the study area and the available data, Section 3 presents the multi-model ensemble methods, Section 4 briefs the selected rainfall–runoff models, Section 5 provides the criteria to evaluate model performance, Section 6 discusses the results, and Section 7 presents the conclusions and the remarks.

#### 2. Study Area and Data

## 2.1. Yellow River Basin

The Yellow River is the second longest river in Asia and the sixth longest in the world, with an estimated length of 5464 km and a drainage area of 752,443 km<sup>2</sup>, and nourishes more than 110 million people with drinking water and irrigation. The basin has an east–west expanse of about 1900 km and a north–south expanse of about 1100 km as shown in Figure 1. The annual runoff of the Yellow River is about 58 billion m<sup>3</sup>, and water resource per capita is only 500 m<sup>3</sup>/person, which implies that the basin is under severe threats of water insecurity. Furthermore, the industrial, drinking, and irrigation water demands are continuously increasing in line with expanding urbanization and intensified economic development. The basin is therefore very sensitive to climate change, which could have an unfavorable impact on its hydrological cycle and on water security. It is therefore vital that hydrological responses to future climate change scenarios are understood so that workable decision-making data for the entire catchment concerning the overall economy and environment can be provided.



Figure 1. Topographical overview map of the Yellow River Basin (YRB).

## 2.2. Data Availability

The Yellow River Basin (YRB) has a relatively long history of measurement and monitoring starting from the 1950s. The China Meteorological Administration (CMA) has the responsibility of operating and maintaining the national network and the quality control of the datasets. Courtesy of

the CMA, the data used in this study were downloaded from the CMA (http://cdc.cma.gov.cn/). In total, 79 meteorological stations in the YRB, with daily and monthly precipitation, air temperature, and wind speed information, and six hydrological stations along the mainstream of the Yellow River, with runoff records for the corresponding period, were selected to provide data for the study period (1961–2000) (see Figure 2). The publication and availability of the drainage network as well as the field data are under the authority of the CMA. Although datasets based on data collected after the year 2000 are not available, the data used in this study based on a period of 40 years are considered highly reliable since they were quality-controlled by the CMA before publishing.



Figure 2. Location and national network coding of the meteorological and hydrological stations in the YRB.

General circulation models (GCMs) are widely used to produce future global climate change scenarios. There are more than 30 GCMs with different assumptions and methodologies available, and all of these models have been used by different academic communities and applied to different countries and regions. In this study, the three following GCMs were considered to best represent the study area and were therefore selected to produce monthly precipitation predictions: CSIRO:MK30, (from the Commonwealth Scientific and Industrial Research Organisation Atmospheric Research, Australia), INM:CM30 (from the Institute for Numerical Mathematics, Russia), and MRI-CGCM2.3.2 (from the Meteorological Research Institute, Japan). Hereafter, they will be referred to as CSIRO, INM, and MRI, respectively. Two different target periods of 20 years—2046–2065 and 2081–2100—were set up under selected emission scenarios of A1B, A2, and B1 based on the IPCC (Intergovernmental Panel on Climate Change) from the IPCC-SRES (IPCC- Special Report on Emissions Scenarios) [19].

All input data were downloaded from the Data Distribution Centre (DDC) of the IPCC (http://ipcc-ddc.cru.uea.ac.uk) for the above three GCMs with different spatial resolutions (ranging from 1.9 to 5°). To facilitate hydrological modeling in the following steps, data were downscaled using an inverse distance weighted interpolation procedure, and the resulting values were populated evenly over a  $0.5 \times 0.5^{\circ}$  grid.

## 3. Multi-Model Ensemble Methods

The concept of combining the forecasts obtained from different models or methods was discussed and used in the pioneering works of [13–16] and others. The essence of the concept of these methods is that each model output captures certain important aspects of the information available about the process being modeled, thereby providing a source of information that may be different from that of other models [12]. Combining these various sources of information may enable the user to gain a merged, all-inclusive picture for a given study area. Furthermore, the judicious combination of outputs of different models may assist in understanding the underlying physical processes involved and thus in developing improved individual models. As such, it might be possible to develop a new individual model that can effectively utilize the different types of information reflected in the estimated outputs of each of the models included in the combination [20].

Several methods of combining model outputs have been reported: the simple average method, the weighted average method [12], the neural network method [21] (which is based on a learning procedure through a black box), the first order Takagi–Sugeno method [22] (which is based on a special class of fuzzy system), and the Bayesian model averaging method (BMA) [11]. Compared with more traditional averaging methods, BMA is becoming popular due to its ability to optimize weights based on performance and thus providing a superior choice in modeling. In this paper, we are interested in comparing the results of a genetic-algorithm-based weighting average method with the results of Bayesian model averaging and simple average methods.

#### 3.1. Genetic Algorithms

Genetic algorithms (GAs) are known as global search heuristics for finding exact or approximate solutions to optimization and search problems based on the evolutionary ideas of natural selection. They are often implemented in a computer simulation in which a population of abstract representations of candidate solutions to an optimization problem evolves toward improved solutions. The evolution starts from a population of randomly generated individuals and occurs in generations. In each generation, the fitness function of every individual in the population is evaluated, multiple individuals are stochastically selected from the current population (based on their fitness), and modified to form a new population through genetic operators of crossover (recombination) and mutation. For each new solution shares many of the characteristics of its "parents." New parents are selected for each new child, and the process continues until a new population of solutions of appropriate size is generated. The new population is then used in the next iteration of the algorithm. The algorithm terminates when a termination condition has been reached, commonly a maximum number of generations has been produced [23]. In this paper, a genetic algorithm based on [23] is employed to optimize the set of weights for eight individual rainfall–runoff models.

#### 3.2. Bayesian Model Averaging Scheme

Bayesian model averaging (BMA) is a statistical procedure that infers consensus predictions by weighing individual predictions based on their probabilistic likelihood measures, with the better performing predictions receiving higher weights than the worse performing predictions. Furthermore, BMA provides a description of the total predictive uncertainty that is more reliable than the original ensemble, leading to a sharper and better-calibrated probability density function for the probabilistic predictions [10,24]. A detailed description of BMA scheme implementation is given in [11]. The same procedure from [11] is employed in this study.

#### 4. A Brief Description of the Selected Rainfall–Runoff Models

In climate-change-related hydrological modeling, monthly water balance models are becoming more popular thanks to their flexibility and ease of use [25–27]. In favor of using conceptual water balance models instead of physically based models or black box models, the authors of [26] stated that the detailed realism of a physically based model posed a different set of complications. First, the physically based models require a high resolution, in both space and time, of climatic input data that may not be available; second, it is possible that model parameters may need to change as climate evolves: soil structure may change, for example, as summers become drier, and, more importantly, the distribution and composition of catchment vegetation will probably alter. There are at present too many unknowns for detailed physically based models to be used in climate impact studies [27].

Based on the above considerations, eight numerical models, including five lumped conceptual models, i.e., GR5M, AWBM, SIMHYD, TRPWB, VWBM, one local distributed model (the Yellow River Water Balance model, YRWBM), one physically based model (VIC), and one black-box model (the artificial neural network model, ANN), were employed in this study for comparison (see Table 1 for details). A brief summary of these eight models is given in Table 1. It should be pointed out that, being the only physically based model, the VIC's performance might be affected due to the fact that only averaging parameters were used in the VIC on the sub-basin scale. The outputs of the eight different rainfall–runoff models were then combined using different BMA strategies to find the best fitted ensemble strategy for the YRB.

Model	Туре	Reference
Australian Water Balance	Conceptual	Boughton, 1993 [28]; Boughton and Carroll, 1993 [29]; Boughton,
Model (AWBM)	rainfall-runoff model	2004 [30]; Boughton and Chiew, 2003 [31].
CP5M model	Conceptual	Mouelhi et al., 2006 [32]; Makhlouf and Michel, 1994 [33];
GR5M model	rainfall-runoff model	Perrin et al., 2003 [34].
SIMHVD model	Conceptual	Porter and McMahon, 1976 [35]; Chiew and McMahon, 1990,
Similid inddei	rainfall-runoff model	1991 [36,37]; Chiew et al., 2002 [38].
TRPWB model	Conceptual rainfall-runoff model	Xiong and Guo, 1999 [39]; Hu et al., 2008 [40].
VWBM model	Conceptual rainfall-runoff model	Xu, 1992 [41]; Vandewiele et al., 1992 [42]; Xu and Vandewiele, 1994 [43].
YRWBM model	Conceptual rainfall-runoff model	Guo and Wang, 1994 [44]; Wang et al., 2001 [45]
VIC model	Distributed hydrological model	Liang et al., 1996 [46]; Liang and Xie, 2001 [47].
ANN model	Black-box model	Box, G.E.P. and Jenkins, G.M. [48]; Govindaraju, R.S. and Rao, A.R. [49].

# 5. Evaluation of Model Performance

The overall performance of these eight models was evaluated using seven selected performance criteria as shown in Table 2 together with the measured parameter for each criteria. The combined performances of these criteria were then used as the objective functions to examine the general model performances.

Criterion	Abbreviation	Equation	What It Measures
Nash-Sutcliff Measure	NS	$1 - rac{\sum_{t=1}^{T} \left(y_t^{obs} - y_t^{ost} ight)^2}{\sum_{t=1}^{T} \left(y_t^{obs} - \overline{y^{obs}} ight)^2}$	Overall average precision
Root Mean Square Deviation	RMSD	$\sqrt{rac{\sum_{t=1}^{T} \left(y_t^{obs} - y_t^{est} ight)^2}{T}}$	Average error, but a greater focus on high flow
Relative Error	RE	$\frac{\sum_{t=1}^{T} (y_t^{est} - y_t^{obs})}{\sum_{t=1}^{T} u^{obs}}$	Water balance
Mean Absolute Error	ABSERR	$rac{\sum_{t=1}^{T} \left  y_t^{obs} - y_t^{est} \right }{T}$	Average error
Maximum Absolute Error	ABSMAX	$\max_{1 \le t \le T} \left  y_t^{obs} - y_t^{est} \right $	Maximum flood peak at measured flood peak time
Peak Difference	PDIFF	$\max_{1 \leq t \leq T} \left\{ y_t^{obs} \right\} - \max_{1 \leq t \leq T} \left\{ y_t^{est} \right\}$	Maximum flood peak difference (may not be at the same time)
First Lag Autocorrelation	RCOEF	$\frac{\sum_{t=1}^{T} y_{t}^{obs} \cdot y_{t}^{cst} - \frac{\sum_{t=1}^{T} y_{t}^{obs} \cdot \sum_{T=1}^{T} y_{t}^{cst}}{T}}{\sqrt{\left(\sum_{t=1}^{T} y_{t}^{obs} - \frac{(\sum_{t=1}^{T} y_{t}^{obs})^{2}}{T}\right) \left(\sum_{t=1}^{T} y_{t}^{cst} - \frac{(\sum_{t=1}^{T} y_{t}^{cst})^{2}}{T}\right)}$	$\frac{2}{2}$ Correlation

 Table 2. Description of criterions for selected models.

## 6. Results and Discussion

For optimal comparison results, all model simulations were performed with the same calibration, validation, and the simulation periods. The years from 1962 to 1980 was selected as the calibration

period, while the years from 1981 to 2000 were used as the validation period. Two future target simulation periods were chosen: 2046–2065 and 2081–2100. The performance of the seven objective functions were then calculated for each of them and subsequently compared.

#### 6.1. Model Performance during the Calibration Period (1962–1980)

Table 3 lists the performance results of the eight rainfall–runoff models referring to the seven criteria. It can be found that (1) the best benchmark values are achieved by the ANN, (2) all water balance (RE) values are less than 10%, with the best performance by the GR5M, and (3) six of the worst performance values belong to the VIC followed by the TRPW. It can also be noted that, for the Nash–Sutcliff criterion (NS), all models performed well, around or above 0.8, except for the VIC (NS = 0.51).

**Table 3.** Rainfall–runoff model performance during the calibration period (the best value for each criteria is in bold and the worst values are in \*).

Model	NS	RMSD	RE	ABSERF	R ABSMAX	PDIFF	RCOEF
ANN	0.85	12.73	-6.14	8.79	50.61	6.31	0.86
AWBM	0.8	14.65	-9.34	10.86	56.28	-8.6	0.82
GR5M	0.82	13.82	-5.28	9.89	55.35	-9.03	0.83
SIMH	0.77	15.73	-9.91 *	11.4	62.65	1.4	0.79
TRPW	0.8	14.57	-6.6	11.08	45.35	-7.5	0.82
VWBM	0.75	16.6	-8.23	12.02	65.35	8.06	0.76
YRWB	0.84	13.32	-8.87	9.61	61.03	-5.02	0.86
VIC	0.51 *	22.97 *	-5.45	16.93 *	79.48 *	20.14 *	0.58 *

The RMSD and the RE are two typical objective functions in validating rainfall–runoff model results. Table 4 presents the RMSD and RE variations of the eight models for different seasons (spring, summer, autumn, and winter). It can be observed that, for both RMSD and RE, (1) the ANN performs well in summer, and the TRPW performs well in autumn; (2) the ANN and VIC respectively have the best RMSD and RE in spring, and the YRWB and ANN respectively have the best RMSD and RE in winter. Although the results among the models are scattered, it is clear that the ANN showed the overall best performance in terms of seasonal modeling.

**Table 4.** Rainfall–runoff model performance in different seasons during the calibration period (the best value for each criteria is in bold, and the worst ones are in \*).

Model	RMSD <sub>SPR</sub>	RE <sub>SPR</sub>	RMSD <sub>SUM</sub>	RE <sub>SUM</sub>	RMSD <sub>AUT</sub>	RE <sub>AUT</sub>	RMSD <sub>WIN</sub>	<b>RE</b> <sub>WIN</sub>
ANN	9.49	-10.73	12.81	0.8	18.94	-11.22	5.97	-2.25
AWBM	12.63	-26.36	17.2	-6.14 *	18.96	-13.21	6.57	30.67
GR5M	10.41	-22.52	14.89	1.52	19.81	-10.41	6.41	26.35
SIMH	15.64	-35.31	14.08	1.84	22.65	-16.03 *	5.83	23.89
TRPW	16.16 *	-37.56 *	15.47	-3.17	17.05	-3.38	7.61	32.59 *
VWBM	14.14	-33.05	17.97	1.23	22.89	-13.32	7.46	28.95
YRWB	9.96	-19.55	14.87	-3.58	18.83	-6.98	5.85	-14.73
VIC	15.17	-7.72	24 *	2.46	34.16 *	-10.45	11.73 *	-10.72

## 6.2. Model Performance during the Validation Period (1981–2000)

Similar to the calibration period, the same comparison is done for the validation period. The results of the performance values are summarized in Table 5. For easy comparison, the corresponding values from calibration (Table 3) are displayed in the line below in parentheses. As can be seen in Table 5, most of the model performance is similar to that of the calibration period, except NS values for ANN and VIC, which decreased slightly. The AWBM performed best in terms of the NS, RMSD, and RCOEF, while the ANN's performance dropped remarkably. This fact indicates the risks of using individual models and the need for a multi-model approach.

Model	NS	RMSD	RE	ABSERR	ABSMAX	PDIFF	RCOEF
ANN	0.79 (0.85)	13.92 (12.73)	-0.4 (-6.14)	10.63 (8.79)	54.75 (50.61)	2.12 (6.31)	0.79 (0.86)
AWBM	<b>0.83</b> (0.80)	12.62 (14.65)	-4.53 (-9.34)	9.46 (10.86)	58.47 (56.26)	0.29 (-8.6)	0.83 (0.82)
GR5M	0.81 (0.82)	13.15 (13.82)	3.81 (-5.28)	9.64 (10.86)	66.44 (56.26)	5.02 (-9.03)	0.82 (0.83)
SIMH	0.78 (0.77)	14.14 (15.73)	-4.67(-9.91)	10.46 (11.4)	67.47 (62.65)	19.29* (1.4)	0.79 (0.79)
TRPW	0.79 (0.80)	13.89 (14.57)	-1.47(-6.6)	10.92 (11.08)	52.28 (45.35)	<b>0.18</b> (-7.5)	0.8 (0.82)
VWBM	0.77 (0.75)	14.59 (16.60)	1.1(-8.23)	11.0 (12.02)	63.58 (65.35)	18.85 (8.06)	0.77 (0.76)
YRWB	0.82 (0.84)	12.67 (13.32)	4.16(-8.87)	8.53 (9.61)	64.75 (61.03)	0.73 (-5.02)	<b>0.83</b> (0.86)
VIC	0.44 * (0.51 *)	22.55 * (22.97 *)	-17.46 * (-5.41)	16.39 * (16.93 *)	92.54 * (79.48 *)	-12.62 (20.14 *)	0.52 * (0.58 *)

**Table 5.** Rainfall–runoff model performance during the validation period (the best value for each criteria is in bold, and the worst ones are in \*).

Again, there is a need to compare RMSD and RE variations in the eight models for different seasons of the validation period. As displayed in Table 6, the results of these two parameters show a spread pattern over the four different seasons. The best performers are the GR5M, the TRPW, the AWBM, and the YRWB for the spring, summer, autumn, and winter, respectively, which is clearly different compared to that of the calibration period.

**Table 6.** Rainfall–runoff model performance during the validation period (the best value for each criteria is in bold, and the worst ones are in \*).

Model	RMSD <sub>SPR</sub>	RE <sub>SPR</sub>	RMSD <sub>SUM</sub>	RE <sub>SUM</sub>	RMSD <sub>AUT</sub>	REAUT	RMSD <sub>WIN</sub>	<b>RE</b> <sub>WIN</sub>
ANN	8.8	-11.07	18.28	-5.2	17.34	1.58	7.9	39.58
AWBM	10.35	-23.17	17.44	-7.29	13.24	-0.35	7.09	35.82
GR5M	8.7	-14.14	17.38	4.61	16.16	5.23	7.29	36.44
SIMH	13.42	-33.69	17.18	4.16	17.04	-5.53	5.81	24.72
TRPW	13.47	-34.38 *	16.71	-2.84	15.83	10.94	7.81	33.57
VWBM	11.49	-25.31	18.94	3.29	16.87	3.47	8.7	43.2 *
YRWB	8.94	6.5	16.9	4.44	15.92	3.22	4.78	0.97
VIC	15.83 *	-14.25	31.38 *	-23.34 *	26.05 *	-14.84 *	10.97 *	-7.34

## 6.3. The Performance of Multi-Model Ensemble

## 6.3.1. Model Performance

All models are imperfect. This is generally true and has been shown in particular by the above examples. It is therefore interesting to see whether the ensemble approach and/or any other averaging methods can improve the objective function. In the following section, we compare the best individual model with three different methods:—the simple average method (SAM), a genetic algorithm (GA), and Bayesian model averaging (BMA)—to investigate the differences between them.

Table 7 illustrates the results of the four approaches for the same calibration period, with the RMSD and RE broken down into four different seasons. From the table, it is apparent that the GA and BMA approaches outperformed the best individual model for NS. Furthermore, for the seasonal cases, GA and BMA results showed values that are clearly superior to those of the best individual model for the RMSD for all seasons. Water balance (RE) is the only value where the best individual model had some slightly improved results.

For the case of the validation period, the results are slightly different, as shown in Table 8. In Table 8, the best individual model value for NS (0.826) is from the AWBM, which is still the lowest compared to others. For the seasonal changes of the RMSD and RE, the GA showed the best results for spring and summer seasons. The best individual model showed slightly improved results compared with the others, indicating improved water balance estimation, especially for the winter.

It is worth pointing out that objective functions from the best individual models are weighted through genetic and Bayesian treatment in order to produce a complementary objective function with improved accuracy. The final weights are not lineally corresponding to the best single values based on individual models. In other words, a higher NS function of an individual model may not necessarily result in higher weights for the GA and BMA approaches.

Averaging Method	NS	RMSD <sub>SP</sub>	R RE <sub>SPR</sub>	RMSD <sub>SUM</sub>	RE <sub>SUM</sub>	RMSD <sub>AUT</sub>	REAUT	RMSD <sub>WI</sub>	N RE <sub>WIN</sub>
SAM	0.849	10.98	-24.10	12.73	-0.63	18.63	-10.62	4.98	14.34
BMA	0.856	10.66	-23.18	12.55	-0.65	18.21	-10.54	4.86	12.90
GA	0.891	7.69	-12.09	12.36	4.97	15.41	-4.14	5.07	9.36
Best individual	0.851	9.49	-7.72	12.81	0.8	17.05	-3.38	5.83	-2.25
model *	ANN	ANN	VIC	ANN	ANN	TRPW	TRPW	SIMH	ANN
* Best value from Table 3.									

**Table 7.** Multi-model ensemble method performance compared with the best individual model duringthe calibration period (the best values are in bold).

**Table 8.** Multi-model ensemble method performance compared with the best individual model during the validation period (the best values are in bold).

Averaging Method	NS	RMSD <sub>SP</sub>	R RE <sub>SPR</sub>	RMSD <sub>SUM</sub>	RE <sub>SUM</sub>	RMSD <sub>AUT</sub>	RE <sub>AUT</sub>	RMSD <sub>WI</sub>	N RE <sub>WIN</sub>
SAM	0.834	9.28	-18.68	16.54	-2.77	14.56	0.47	5.89	25.87
BMA	0.838	8.98	-17.6	16.33	-2.68	14.48	0.71	5.90	26.28
GA	0.840	7.16	-6.09	15.96	1.53	15.28	7.64	6.84	34.46
Best individual	0.826	8.7	6.5	16.71	-2.84	13.24	-0.35	4.78	0.97
model *	AWBM	GR5M	YRWB	TRPW	TRPW	AWBM	AWBM	YRWB	YRWB

\* Best value from Table 5.

## 6.3.2. Rainfall-Runoff Simulation and Prediction

Based on the previous sections, it is clear that the genetic algorithm (GA) approach has the best performance and accuracy within the comparison. The GA method was therefore selected to simulate the runoff for the study area.

In Figures 3 and 4, the simulation results of the GA method are plotted against the observed monthly runoff in calibration (1962–1980) and validation (1981–2000) periods, respectively, for the Huayuankou station (the lower reach of the entire YRB). As can be deduced from the figures, the simulation accuracy is satisfactory for most of the years in both periods (with NS = 891 and NS = 840, respectively).



Figure 3. Observed and simulated monthly flows (10<sup>9</sup> m<sup>3</sup>) at Huayuankou station for the calibration period.



Figure 4. Observed and simulated monthly flows (10<sup>9</sup> m<sup>3</sup>) at Huayuankou station for the validation period.

# 6.4. Runoff Prediction under Varying Climate Scenarios

As described in Section 2.2, future climate scenarios of A1B, A2, and B1 from the IPCC-SRES are used by the selected GCMs with downscaled outputs to produce the input for our eight rainfall–runoff models. The models are now furnished with the GA method for the best combination of objective functions.

Figures 5 and 6 demonstrate the simulation results for the target periods of 2046–2065 and 2081–2100, respectively.



**Figure 5.** Simulated monthly flows  $(10^9 \text{ m}^3)$  during the period of 2046–2065.



**Figure 6.** Simulated monthly flows  $(10^9 \text{ m}^3)$  during the period of 2081–2100.

Mean values of the annual runoff in the periods of 2046–2065 and 2081–2100 are shown in Table 9. With a baseline mean annual runoff of 56.7 billion m<sup>3</sup> (1961–2000), it is concluded that the greatest mean annual runoff scenario under climate change scenarios downscaled from MRI and A2 would be 53.1 billion m<sup>3</sup>, while the smallest mean annual runoff driven by the CSIRO and B1 scenarios would be 47.21 billion m<sup>3</sup>. In the period of 2081–2100, the greatest mean annual runoff scenario under climate change scenarios downscaled from MRI and A2 would be 55.3 billion m<sup>3</sup>, while the smallest mean annual runoff driven by the CSIRO and B1 scenarios that the greatest mean annual runoff driven by the CSIRO and B1 scenarios would be 47.02 billion m<sup>3</sup>. It is interesting to note that the greatest and smallest runoff scenarios would be under the combined scenarios of MRI-A2 and CSIRO-B1 in both periods of 2046–2065 and 2081–2100.

SR	ES		A2		B1			
GCMs				MDI	CCUDO		101	
Periods	Quantiles	- CSIRO	IINM	MKI	CSIKO	INM	MKI	
	Mean	52.59	51.86	53.10	47.21	48.76	51.66	
	R25	46.35	48.02	48.14	44.09	46.61	49.28	
2046-2065	R50	52.78	52.74	52.74	46.36	48.29	50.47	
	R75	57.94	55.08	56.25	50.44	50.22	53.80	
	Mean	54.14	48.82	55.30	47.02	51.19	53.46	
2081-2100	R25	48.38	45.42	50.80	42.63	47.72	52.13	
	R50	55.01	48.04	55.26	45.65	50.23	53.48	
	R75	59.57	51.65	59.32	50.67	54.26	55.64	

Table 9. Quantile and mean values for annual runoff in the periods of 2046-2065 and 2081–2100 (Unit: billion  $m^3$ /annual).

The change rate of quantile and mean values for annual runoff in the periods of 2046–2065 and 2081–2100 compared with that in the baseline period of 1961–2000 is shown in Table 10. Overall, it displays slight decreasing trends for runoff in the YRB in the simulated period. Average values of mean annual runoff in the periods of 2046–2065 and 2081–2100 are 50.86 and 51.65 billion m<sup>3</sup>, which

are remarkably smaller than runoff in the baseline period by 10.3 and 8.9%, respectively. Generally, runoff under the A2 scenario is greater than that under the B1 scenario.

SR	ES		A2		B1			
GC	GCMs				CCIDO		MDI	
Periods	Quantiles	- CSIRO	INM	MKI	CSIRO	INM	MKI	
2046–2065	Mean R25 R50 R75	-6.48 0.32 0.91 -14.02	-7.78 3.94 0.85 -18.26	-5.57 4.20 0.84 -16.53	-16.05 -4.57 -11.36 -25.14	-13.29 0.89 -7.67 -25.47	-8.13 6.67 -3.50 -20.16	
2081–2100	Mean R25 R50 R75	-3.72 4.73 5.19 -11.59	-13.18 -1.68 -8.15 -23.35	-1.66 9.96 5.65 -11.97	-16.38 -7.72 -12.71 -24.81	-8.96 3.29 -3.95 -19.47	-4.93 12.85 2.25 -17.43	

**Table 10.** Change rate of quantile and mean values for annual runoff in the periods of 2046–2065 and 2081–2100 compared with that in the period of 1961–2000 (Unit: %).

The distributions of the simulated monthly runoff over the year for the two targeted periods are plotted together with the baseline runoff in Figures 7 and 8, respectively. For both periods, it is clear that the largest variations between climate scenarios are during May–June and August–October.

Compared to the baseline runoff, it is quantified that average spring, summer, autumn, and winter runoff of the Yellow River Basin would change -16.7-23.3%, -24.5--13.9%, -28.7--3.3%, and 28.4-49.0% for the 2046–2065 time span and change -19.1-43.2%, -25.4--12.4%, -36.1--5.8%, and 28.4-61.7% for the 2081–2100 time span.



**Figure 7.** The average simulated monthly flows  $(10^9 \text{ m}^3)$  during the period of 2046–2065.



Figure 8. The average simulated monthly flows  $(10^9 \text{ m}^3)$  during the period of 2081–2100.

#### 7. Discussion and Conclusions

Three GCMs were employed to simulate future hydrological scenarios for the Yellow River Basin in climate scenarios A1B, A2, and B1 from the IPCC-SRES. The GCM outputs were then downscaled to a grid size of  $0.5 \times 0.5^{\circ}$  for use with the eight rainfall–runoff models selected for this particular study. These runoff outputs were then combined using a multi-model ensemble (MME), which, compared with any individual model, can be expected to be more accurate and reliable. It should be noted that the choice of these three GCMs, among many others, might not be representative. It would be interesting to include other GCMs in a similar study in the future.

For regional scale modeling, GCMs involve high uncertainty due to the GCMs' inter-model uncertainty as well as the coarse spatial resolution. One way to reduce this uncertainty might be to quantify the discrepancies by showing error bars for each case so that the inter-model uncertainty can be addressed more easily. Similar to the current study for hydrological outputs, another way to increase the reliability might be to employ an MME of GCM models, as described in [50]. The performance of the VIC, the only distributed model, compared with other conceptual or black-box models, was worse in many cases. This is partly due to the lack of detailed distributed parameter inputs and partly due to the coarse averaging effects in the space and time domains. In a recent study, a long-term model with the VIC was successfully applied [51]. One future step could be to incorporate the results of [51] to an MME procedure so that the distributed model representation can be enhanced and inter-model uncertainty can thus be reduced.

For the calibration and validation periods of the YRB, it was confirmed that the MBA and GA approaches had an overall performance that was a substantial improvement in terms of rainfall–runoff outputs compared with the individual models and the simple average method (SAM). This indicates application potential for MME approaches in other areas. For example, more recent CMIP5 datasets could be used in the future to generate more reliable scenarios.

For cases of seasonal modeling, various individual models outperformed the MME approaches (Tables 7 and 8). This might be because individual models are more sensitive to temporal variations than those of an MME, where temporal variations may be evened out.

Based on the combined predictions for the two target periods (Figures 7 and 8), compared to the baseline, average annual runoff of the YRB will, for the 2046–2065 time span, decrease 5.6~7.8% in Scenario A2 and 8.1–16.1% in Scenario B1 and, for the 2081–2100 time span, decrease 1.7~13.2% in Scenario A2 and 4.9–16.4% in Scenario B1. This will threaten the water security in the YRB.

Relative to the baseline runoff, average spring, summer, autumn, and winter runoff of the YRB will change -16.7-23.3%, -24.5--13.9%, -28.7--3.3%, and 28.4-49.0% over the 2046–2065 period and, for the 2081–2100 period, change -19.1-43.2%, -25.4--12.4%, -36.1--5.8%, and 28.4-61.7% (Figures 7 and 8). This means that, from January to June, the runoff in most of the simulated scenarios (4 of 6 cases) will be higher than the corresponding baseline period, while the opposite situation will appear for the second half of the year, implying increased severity in the water supply since the first half of the year is generally more vulnerable.

In summary, a multi-model ensemble (MME) using a genetic algorithm (GA) and a Bayesian model averaging method (BMA) is concluded to enhance rainfall–runoff predictions by providing a combined, better weighted parameter set compared to individual models. These results show that it is possible to reduce the uncertainty and thus improve the accuracy for future projections using MME-weighted models, because the MME approach will even out the bias and exploit the best performances of the individual models. Three future tasks are expected to be undertaken: one is the use of the same methodology in other river basins (e.g., the Yangtze River Basin) to confirm the usefulness; another is the optimization of choice of GCMs and of more recent climate change scenarios; a third is the optimization of the setup and composition of the rainfall–runoff models such that the balance between conceptual, black-box, and distributed models, as well as other models such as those reported by [52–54], is improved.

Author Contributions: Xiaoliu Yang conceived and designed the study; Xiaoliu Yang analyzed the data and made model work with contribution of Linus Zhang. Both authors wrote the paper together.

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

#### References

- 1. Ragab, R.; Prudhomme, C. Climate change and water resources management in arid and semi-arid regions: prospective and challenges for the 21st century. *Biosyst. Eng.* **2002**, *81*, 3–34. [CrossRef]
- 2. Arnell, N.W. Climate change and global water resources: SRES emissions and socio-economic scenarios. *Glob. Environ. Chang.* **2004**, *14*, 31–52. [CrossRef]
- 3. Jonathan, I.M.; Graciana, P.; Kenneth, M.M. Managing water under climate change for peace and prosperity in Swaziland. *Phys. Chem. Earth* **2005**, *30*, 943–949.
- Kundzewicz, Z.W.; Mata, L.J.; Arnell, N.W.; Döll, P.; Jimenez, B.; Miller, K.; Oki, T.; Sen, Z.; Shiklomanov, I. The implications of projected climate change for freshwater resources and their management. *Hydrol. Sci.* 2008, 53, 3–10. [CrossRef]
- Christensen, N.S.; Wood, A.W.; Voisin, N.; Lettenmaier, D.P.; Palmer, R.N. The effects of climate change on the hydrology and water resources of the Colorado River basin. *Clim. Chang.* 2004, *62*, 337–363. [CrossRef]
- 6. Elshamy, M.E.; Seiertad, I.A.; Sorteberg, A. Impact of climate change on Blue Nile flows using bias-corrected GCM scenarios. *Hydrol. Earth Syst. Sci.* **2009**, *13*, 551–565. [CrossRef]
- 7. Poulin, A.; Brissette, F.; Leconte, R.; Arsenault, R.; Malo, J.S. Uncertainty of hydrological modelling in climate change impact studies in a Canadian, snow-dominated river basin. *J. Hydrol.* **2011**, *409*, 626–636. [CrossRef]
- 8. Montenegro, S.; Ragab, R. Impact of possible climate and land use changes in the semi arid regions: A case study from North Eastern Brazil. *J. Hydrol.* **2012**, *434–435*, 55–68. [CrossRef]
- Wilby, R.L. Water, Hydropower and Climate Change. In Proceedings of the Water Management, Climate Change Impacts on Hydroelectric Water Resource Management, CEATI, Montreal, QC, Canada, 8–9 October 2008.

- Neuman, S.P. Maximum likelihood Bayesian averaging of uncertain model predictions. *Stoch. Environ. Res. Risk Assess.* 2003, 17, 291–305. [CrossRef]
- 11. Duan, Q.; Ajami, N.K.; Gao, X.; Sorooshian, S. Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Adv. Water Resour.* **2007**, *30*, 1371–1386. [CrossRef]
- 12. Shamseldin, A.Y.; O'Connor, K.M.; Liang, G.C. Methods for combining the outputs of different rainfall–runoff models. *J. Hydrol.* **1997**, 197, 203–229. [CrossRef]
- 13. Bates, J.M.; Granger, C.W.J. The combination of forecasts. J. Oper. Res. Soc. 1969, 20, 451–468. [CrossRef]
- 14. Dickinson, J.P. Some statistical results in the combination of forecasts. *J. Oper. Res. Soc.* **1973**, 24, 253–260. [CrossRef]
- 15. Dickinson, J.P. Some comments on the combination of forecasts. J. Oper. Res. Soc. 1975, 26, 205–210. [CrossRef]
- 16. Newbold, P.; Granger, C.W.J. Experience with forecasting univariate time series and the combination of forecasts. *J. R. Stat. Soc. Ser. A.* **1974**, *137 Pt 2*, 131–146. [CrossRef]
- 17. Chandler, R.E. Exploiting strength, discounting weakness: combining information from multiple climate simulators. *Philos. Trans. R. Soc. A* 2013, *371*, 20120388. [CrossRef] [PubMed]
- Ahn, J.-B.; Lee, J. A new multimodel ensemble method using nonlinear genetic algorithm: An application to boreal winter surface air temperature and precipitation prediction. *J. Geophys. Res. Atmos.* 2016, 121, 9263–9277. [CrossRef]
- 19. IPCC. *IPCC Special Report Emissions Scenarios;* Cambridge University Press: Cambridge, UK, 2000; ISBN 92-9169-113-5.
- 20. Clemen, R.T. Combining forecasts: A review and annotated bibliography. *Int. J. Forecast.* **1989**, *5*, 559–583. [CrossRef]
- 21. Araghinejad, S.; Azmi, M.; Kholghi, M. Application of artificial neural network ensembles in probabilistic hydrological forecasting. *J. Hydrol.* **2011**, 407, 94–104. [CrossRef]
- 22. Xiong, L.; Shamseldin, A.Y.; O'Connor, K.M. A non-linear forecasts combination of the forecasts of rainfall–runoff models by Takagi–Sugeno fuzzy systems. *J. Hydrol.* **2001**, 245, 196–217. [CrossRef]
- 23. Goldberg, D.E. *Genetic Algorithms in Search, Optimization and Machine Learning;* Addison-Wesley Longman Pub. Co., Inc.: Boston, MA, USA, 1989; ISBN 0201157675.
- 24. Hoeting, H.A.; Madigan, D.; Raftery, A.E.; Volinsky, C.T. Bayesian model averaging: A tutorial. *Stat. Sci.* **1999**, *14*, 382–417, with discussion.
- 25. Gleik, P.H. Methods for evaluating the regional hydrologic impacts of global climatic changes. *J. Hydrol.* **1986**, *88*, 97–116. [CrossRef]
- 26. Arnall, N.W. Factors controlling the effects of climate change on river flow regimes in a humid temperate environment. *J. Hydrol.* **1992**, *132*, 321–342. [CrossRef]
- 27. Xu, C.Y.; Singh, V.P. A Review on Monthly Water Balance Models for Water Resources Investigations. *Water Resour. Manag.* **1998**, *12*, 31–50. [CrossRef]
- 28. Boughton, W.C. A hydrograph-based model for estimating the water yield of ungauged catchments. In Proceedings of the 1993 Hydrology and Water Resources Conference, Institution of Engineers, Australia, National Conference Publication No. 93/14, Newcastle, Australia, 30 June–2 July 1993; pp. 317–324.
- 29. Boughton, W.C.; Carroll, D.G. A simple combined water balance/ flood hydrograph model. In Proceedings of the 1993 Hydrology and Water Resources Conference, Institution of Engineers, Australia, National Conference Publication No. 93/14, Newcastle, Australia, 30 June–2 July 1993; pp. 299–304.
- 30. Boughton, W.C. The Australian water balance model. Environ. Model. Softw. 2004, 19, 943–956. [CrossRef]
- 31. Boughton, W.C.; Chiew, F. *Calibrations of the AWBM for Use on Ungauged Catchments*; Technical Report 03/15, CRC for Catchment Hydrology; Monash University: Monash, Australia, 2003; 37p.
- 32. Mouelhi, S.; Michel, C.; Perrin, C.; Andréassian, V. Stepwise development of a two-parameter monthly water balance model. *J. Hydrol.* **2006**, *318*, 200–214. [CrossRef]
- Makhlouf, Z.; Michel, C. A two-parameter monthly water balance model for French watersheds. *J. Hydrol.* 1994, 162, 299–318. [CrossRef]
- Perrin, C.; Michel, C.; Andréassian, V. Improvement of a parsimonious model for runoff simulation. *J. Hydrol.* 2003, 279, 275–289. [CrossRef]
- 35. Porter, J.W.; McMahon, T.A. *The Monash Model: User Manual for Daily Program HYDROLOG*; Research Report No. 2/76; Department of Civil Engineering, Monash University: Monash, Australia, 1976; 41p.

- 36. Chiew, F.H.; McMahon, T.A. Estimating groundwater recharge using a surface watershed modeling approach. *J. Hydrol.* **1990**, *114*, 285–304. [CrossRef]
- Chiew, F.H.; McMahon, T.A. Improved Modelling of the Groundwater Processes in MODHYDROLOG; International Hydrology and Water Resources Symposium, National Conference Publication No. 91/22; Institution of Engineers: Perth, Australia, 1991; Volume 2, pp. 492–497.
- Chiew, F.H.; Peel, M.C.; Western, A.W. Application and testing of the simple rainfall-runoff model SIMHYD. In *Mathematical Models of Small Watershed Hydrology and Applications*; Singh, V.P., Frevert, S.K., Eds.; Water Resources Publications: Littleton, CO, USA, 2002; pp. 335–367.
- 39. Xiong, L.; Guo, S. A two-parameter monthly water balance model and its application. *J. Hydrol.* **1999**, 216, 111–123. [CrossRef]
- 40. Hu, A.; Guo, S.; Xiong, L.; Guo, J. Three-parameter monthly water balance model and its application. *Water Resour. Power* **2008**, *26*, 8–10, 42. (In Chinese)
- 41. Xu, C.Y. Monthly Water Balance Models in Different Climatic Regions. Ph.D. Thesis, V.U.B. Hydrologie 22, Vrije Universiteit, Brussel, Belgium, 1992.
- 42. Vandewiele, G.L.; Xu, C.Y.; Ni-Lar-Win. Methodology and comparative study of monthly water balance models in Belgium, China and Burma. *J. Hydrol.* **1992**, *134*, 315–347. [CrossRef]
- 43. Xu, C.Y.; Vandewiele, G.L. Sensitivity of monthly rainfall runoff models to input errors and data length. *Hydrol. Sci. J.* **1994**, *39*, 157–176. [CrossRef]
- 44. Guo, S.L.; Wang, G.Q. *Monthly Water Balance Model for the Yellow River Basin*; Renmin Huanghe: Zhengzhou, China, 1994; Volume 122, pp. 13–16. (In Chinese)
- 45. Wang, G.Q.; Wang, Y.Z.; Kang, L.L. Sensitivity of runoff in the middle Yellow River basin to climate change. *J. Appl. Meteorol. Sci.* **2001**, *13*, 117–121 (In Chinese with English Abstract). (In Chinese)
- 46. Liang, X.; Lettrnmaier, D.P.; Wood, E.F. Surface soil moisture parameterization of the VIC-2L model: evaluation and modification. *Glob. Planet. Chang.* **1996**, *13*, 196–206. [CrossRef]
- 47. Liang, X.; Xie, Z.H. A new surface runoff parameterization with subgrid-scale soil heterogeneity for land surface models. *Adv. Water Resour.* **2001**, *24*, 1173–1192. [CrossRef]
- 48. Box, G.E.P.; Jenkins, G.M. Time Series Analysis: Forecasting Control; Holden-Day: Oakland, CA, USA, 1976.
- 49. Govindaraju, R.S.; Rao, A.R. (Eds.) *Artificial Neural Networks in Hydrology*; Springer: Dordrecht, The Netherlands, 2000; ISBN 978-90-481-5421-0.
- 50. Pierce, D.W.; Barnett, T.P.; Santer, B.D.; Gleckler, P.J. Selecting global climate models for regional climate change studies. *Proc. Nat. Acad. Sci. USA* **2009**, *106*, 8441–8446. [CrossRef] [PubMed]
- 51. Zhao, R.; Tao, F. Contributions of climate change and human activities to runoff change in seven typical catchments across China. *Sci. Total Environ.* **2017**, *605–606*, 219–229. [CrossRef] [PubMed]
- 52. Chen, X.Y.; Chaun, K.W.; Busari, A.O. A comparative study of population-based optimization algorithms for downstream river flow forecasting by a hybrid neural network model. *Eng. Appl. Artif. Intell.* **2015**, *46*, 258–268. [CrossRef]
- 53. Wang, W.C.; Xu, D.M.; Chau, K.W.; Lei, G.J. Assessment of river water quality based on theory of variable fuzzy sets and fuzzy binary comparison method. *Water Resour. Manag.* **2014**, *28*, 4183–4200. [CrossRef]
- 54. Chau, K.W. A split-step particle swarm optimization algorithm in river stage forecasting. *J. Hydrol.* **2007**, 346, 131–135. [CrossRef]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).