



Parallel Hydrological Model Parameter Uncertainty Analysis Based on Message-Passing Interface

Zhaokai Yin ¹, Weihong Liao ^{2,*}, Xiaohui Lei ^{2,3} and Hao Wang ^{2,3}

- State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University, Tianjin 300350, China; yinzhaokai@tju.edu.cn
- ² State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing 100038, China; lxh@iwhr.com (X.L.); wanghao@iwhr.com (H.W.)
- ³ Science and Technology Innovation Center for Smart Water, Northeastern University, Shenyang 110819, China
- * Correspondence: liaowh@iwhr.com; Tel.: +86-10-6878-5503

Received: 17 August 2020; Accepted: 15 September 2020; Published: 23 September 2020



Abstract: Parameter uncertainty analysis is one of the hot issues in hydrology studies, and the Generalized Likelihood Uncertainty Estimation (GLUE) is one of the most widely used methods. However, the scale of the existing research is relatively small, which results from computational complexity and limited computing resources. In this study, a parallel GLUE method based on a Message-Passing Interface (MPI) was proposed and implemented on a supercomputer system. The research focused on the computational efficiency of the parallel algorithm and the parameter uncertainty of the Xinanjiang model affected by different threshold likelihood function values and sampling sizes. The results demonstrated that the parallel GLUE method showed high computational efficiency and scalability. Through the large-scale parameter uncertainty analysis, it was found that within an interval of less than 0.1%, the proportion of behavioral parameter sets and the threshold value had an exponential relationship. A large sampling scale is more likely than a small sampling scale to obtain behavioral parameter sets at high threshold values. High threshold values may derive more concentrated posterior distributions of the sensitivity parameters than low threshold values.

Keywords: hydrological model; parameter uncertainty; GLUE; parallel computing; MPI

1. Introduction

Accurate hydrological forecasting plays an important role in the scientific management of water resources, flood prevention and the efficient operation of hydraulic projects [1–5]. However, hydrological forecasts are made in the face of pervasive uncertainty, both objectively and subjectively. From the objective point of view, due to the inherent uncertainty of natural water cycle and of human observation and modeling processes, each link in hydrological forecasting, such as the model input, the model structure and the model parameters, has its own uncertainty [6,7]. From the subjective point of view, in the process of hydrological forecasting there are several decisions that have to be made by modelers, such as model scope, equations and parameter values [8]. All these decisions are inevitably influenced by subjective factors, such as values, ethics and politics [9], and will inevitably bring uncertainty to forecasting results [10]. At present, the uncertainty in hydrological forecasting has become a research hotspot.

Among the studies of hydrological uncertainty, the uncertainty of the model parameter is an important aspect [11–14]. In traditional hydrological forecasting studies, the model should be localized by parameter calibration [15–18]. However, many studies have found that it is difficult to find a certain set of parameters that can yield model performances much better than other parameter



sets. In contrast, oftentimes there may be multiple parameter sets that will yield similar model performances. This phenomenon is termed "equifinality" [19]. To assess the parameter uncertainty, different methods have been used, for example, the Generalized Likelihood Uncertainty Estimation (GLUE) [19], Markov Chain Monte Carlo (MCMC) [20], Parameter Solution (ParaSol) [21], Sequential Uncertainty Fitting (SUFI-2) [22], etc. Among them, the GLUE attracts many users due to its simpler concept and convenience in implementation [13,23,24]. As a typical Monte Carlo method, the GLUE method requires sampling from the model parameter space and performing a run of the model with each set of parameters. Obviously, the larger the scale of parameter sampling, i.e., more parameter sets sampled from the parameter space, the more comprehensive the rule of parameter uncertainty that can be described. However, due to the limitation of computational power, the current parameter sampling scale using the GLUE method is mostly limited to the 10⁵ level.

Although multiple criteria were used as likelihood functions in some of the frontier researches of GLUE [13,24,25], most of the GLUE researches still use a single criterion as likelihood function, as in this study. In this case, all the parameter sets that could lead to a model performance above the threshold value of the likelihood function are considered acceptable and are called "behavioral" parameter sets. Meanwhile, the likelihood function and its threshold value are determined subjectively rather than based on statistically consistent error models under the consideration that the structure of errors could be complex and nonstationary. Several studies have reported that the determination of the likelihood function and its threshold values have great impacts on the results of the GLUE method [26–29].

To deal with the problem of computing efficiency, some researchers choose parallel computing. Compared with traditional computing, which executes tasks sequentially, parallel computing refers to the process of executing multiple tasks on multiple processors or computing units at the same time to solve the complex problems efficiently. In parallel computing, the complex problems are usually divided into multiple subproblems, and then, the unrelated ones are allocated to the processors for execution at the same time. These steps can significantly reduce the overcall running time and improve the computational efficiency. Parallel computing cannot be separated from the support of related hardware and software environments. The hardware environment mainly includes the multicore central processing unit (CPU), graphics processing unit (GPU) and computer cluster, while the software environment mainly includes the Message-Passing Interface (MPI), OpenMP, Compute Unified Device Architecture (CUDA, NVIDIA Corporation, Santa Clara, CA, USA) and MapReduce. Since parallel computing was proposed, it has been applied in many fields, such as transportation [30], geology [31], information science [32], geospatial algorithms [33,34], etc.

In the field of hydrology, parallel computing has also been used to improve the computational efficiency in (1) hydrological model calibration [35–39], (2) grid computing of the distributed hydrological model [40–45] and (3) the efficient solution of the hydrodynamic model [46–51]. However, the research on the uncertainty of hydrological model parameters is still in the initial stage. Kan et al. [52] used OpenMP and CUDA to achieve the parallel GLUE algorithm on a multicore CPU and GPU and promoted the sampling scale of hydrological model parameter uncertainty research to the million level, i.e., 10⁶. In fact, after parameter sampling, the prediction process of each parameter set in the GLUE method is independent of each other. It is very suitable to use parallel computing to improve the calculation efficiency, because the GLUE method does not need substantial inter-process communication when it is calculated in different processes.

In this study, a parallel GLUE method will be established based on the MPI. A famous hydrological model, the Xinanjiang model, will be used in streamflow simulation. Based on it, a highly efficient large-scale parameter uncertainty analysis on supercomputers will be carried out. In the first experiment, a small sampling scale, a total of 10,000 sets, will be used to perform a computational efficiency assessment based on different numbers of processors. In the second experiment, several sampling scales, the maximum one of which reaches 10 million, i.e., 10⁷, and several likelihood threshold values will be used to assess the parameter uncertainty affected by them. The maximum number of processors used in the second experiment is 2400.

2. Methods

2.1. The GLUE Method

The GLUE method was first proposed by Beven and Binley [19] to assess parameter uncertainty based on their research on model calibration. It was concluded that multiple parameter sets may yield similar model performances, i.e., "equifinality". The GLUE method includes several steps: (1) select a hydrological model and determine the model parameter space; (2) select a likelihood function and determine the threshold value and the priori probability distribution; (3) randomly extract parameter sets from the parameter space; (4) run the model with the extracted parameter sets, calculate the likelihood value according to the model results and save the behavioral parameter sets if the likelihood value exceeds the threshold; (5) rescale the threshold values to formulate a cumulative distribution. In this study, the Nash–Sutcliffe efficiency (NSE) [53] was selected as the likelihood function, as in many other studies [23,28,29,52]:

$$NSE = 1 - \frac{\sum_{t=1}^{T} (Q_{obs}^{t} - Q_{sim}^{t})^{2}}{\sum_{t=1}^{T} (Q_{obs}^{t} - \overline{Q_{obs}})^{2}}$$
(1)

where Q_{obs}^t and Q_{sim}^t represent the observed and simulated streamflow at time t, $\overline{Q_{obs}}$ represents the mean value of the observed streamflow, and T is the total time steps of the observed and simulated streamflow. Different threshold values and sampling scales were considered to evaluate their impacts on the results of the GLUE method.

2.2. MPI

In this study, an MPI is used to implement the parallel GLUE algorithm. An MPI is a message passing library interface protocol that mainly aims at the message passing parallel programming model. In this model, data is transferred between the address spaces of different processes through performing a cooperative operation between processes. In addition to the traditional messaging model, an MPI also provides support for cluster operations, remote memory access, dynamic process creation and parallel I/O. An MPI has become an important part of parallel computing because of its high communication performance, good program portability and powerful functions. An MPI is not a language but a definition for functions, which are expressed as functions, subroutines or methods. At present, there are several implementations of MPI invited by multiple organizations, such as MPICH, OPEN MPI and Intel MPI (Intel Corporation, Santa Clara, CA, USA). The MPICH used in this study is an open source MPI developed by Argonne National Laboratory and Mississippi State University [54–56].

2.3. The Parallel GLUE Method

In the GLUE method, the processes of the model run with candidate parameters and the likelihood value calculation and evaluation are independent of each other after the parameter sampling process. In the original GLUE method, this part performs sequentially. In this study, this part of the GLUE method is parallelized with the MPI. Different processes simultaneously run the same code with the same input data, and only the model parameters are changed, forming a typical single program, multiple data (SPMD) scenario in parallel computing.

In addition, the parallel computing provided by the MPI is based on the processes. Each process uses an independent memory space, so it is able to build a more complex calculation process of each process. However, the disadvantage is that both the opening and destruction of the processes need to occupy computing resources, so the number of opening processes should be limited; it is better to ensure a certain amount of tasks of each process. Therefore, this study chooses the equivalent parallel modeling method to allocate the same amount of computations of each process to maximize the use of computing resources. A specific calculation process is shown in Figure 1, and its brief principle is stated below:

- 1. Create multiple processes according to the sample size and available computing resources, and specify one root process, while the remaining processes are non-root processes.
- 2. Read the model configuration file in the root process, and sample in the model parameter space to generate a group of candidate parameter sets.
- 3. Assign the candidate parameter sets equally to all processes.
- 4. On each process, perform the model run during the calibration period and likelihood value calculation with the sub-candidate parameter sets. Evaluate the likelihood value, and save the parameter sets and forecast results if they are above (or below according to the likelihood function) the threshold value. Among the processes, this step is performed simultaneously.



Figure 1. Flowchart for the parallel Likelihood Uncertainty Estimation (GLUE) method implemented by a Message-Passing Interface (MPI) on a supercomputing system.

2.4. The Xinanjiang Model

The Xinanjiang model is a famous conceptual hydrological model proposed by Zhao et al. [57]. Many researchers have carried out successful practices and developed improvements in many basins around the world. The Xinanjiang model used in this study is integrated in the EasyDHM model system [4], of which the characteristic is, on the basis of retaining the original runoff generation algorithm, to calculate the routing process using the surface information obtained from DEM and

other catchment data and adjust the process using the corresponding parameters. The parameters and their ranges are listed in Table 1. The priori probability distribution of the parameters is assumed as a uniform distribution, as in many other studies.

Parameter	Description		
С	Coefficient of the deep layer evapotranspiration	0.1~0.3	
IMP (%)	Percentage of impervious and saturated areas in the catchment	0~0.7	
WM1 (mm)	Averaged soil moisture storage capacity of the upper layer	5~100	
WM2 (mm)	Averaged soil moisture storage capacity of the lower layer	50~300	
WM3 (mm)	Averaged soil moisture storage capacity of the deep layer	5~100	
В	Exponential of the distribution of the tension water capacity	0.15~0.35	
SM (mm)	Areal mean free water capacity of the surface soil layer	5~100	
EX	Exponent of the free water capacity curve influencing the development of the saturated area	0.5~2	
KG	Outflow coefficients of the free water storage to groundwater relationships	0.05~0.7	
KSS	Outflow coefficients of the free water storage to interflow relationships	0.05~0.7	
KKG	Recession constants for groundwater storage	0.9~0.999	
KKSS	Recession constants for lower interflow storage	0.05~0.95	
PETM	Coefficient of the actual evapotranspiration	0.5~1	
CH_S2M	Coefficient of the river slope	0.1~10	
CH_L2M	Coefficient of the river length	0.5~20	
CH_N2M	Coefficient of the river roughness	0.1~10	
CH_K2M	Coefficient of the river bottom hydraulic conductivity	0.1~10	

Table 1. The parameters and	l their boundaries of th	e Xinanjiang model.
-----------------------------	--------------------------	---------------------

3. Case Study

3.1. Study Area and Data Sets

In this study, the catchment above the Gaoan station in the Jinjiang River Basin is selected as the research area (Figure 2). The Jinjiang River Basin, located in the southwest of the Jiangxi Province, is the second largest tributary of the Ganjiang River, with a total length of 307 km and a total drainage area of 7886 km². The altitude of the Jinjiang River Basin ranges from 18 to 1096 m above sea level. The mountains and low hills are distributed in the northwest, and the plain lays in the east. The Jinjiang River Basin is located in the subtropical area, with a warm and humid climate. The average annual precipitation is approximately 1600 mm. The Gaoan hydrological station is located on the lower reaches of the Jinjiang River, above which the drainage area is 6215 km².



Figure 2. Map of the Jinjiang River Basin that is controlled by the Gaoan station.

The daily precipitation data used in this study come from 20 rain gauges in and around the study area. Other meteorological data, including the temperature, wind speed, relative humidity, and solar radiation, were obtained from the 2 meteorological stations near the study area. The daily streamflow data obtained from the Gaoan hydrological station were provided by the Jiangxi Hydrological Bureau.

All the time series data range from 2000 to 2015, in which the period from 2000 to 2011 was regarded as the calibration period, and the period from 2012 to 2015 was regarded as the validation period.

3.2. Hardware and Software Environment

The hardware environment of this study is the "Yuan" supercomputing system of the Supercomputing Center of Chinese Academy of Sciences [58]. It is a new generation supercomputing system affiliated with the general supercomputing environment center of the Chinese Academy of Sciences. The total computing power of the "Yuan" supercomputing system is approximately 2.3 Pflops (the computing power refers to the double precision floating-point peak computing power), and the CPU general-purpose computing capacity is 700 Tflops. The maximum computing power used in this study is approximately 96 Tflops. The user terminal system version of the "Yuan" supercomputing system is CentOS 6.6, the compiler version is GCC-8.3.0, and the MPICH version is MPICH-3.3.1.

3.3. Evaluation Criteria

The indicators used in this study to evaluate the computing performance of the parallel GLUE method are the speedup ratio (Sp) and parallelism efficiency (E), which are given by the following expressions:

$$Sp(n) = \frac{T_s}{T_p} = \frac{t_s + t_p}{t_s + t_p/n}$$
⁽²⁾

$$E = \frac{Sp}{n} = \frac{T_s}{nT_p} \tag{3}$$

where T_s represents the execution time of the serial program, T_p represents the execution time of the parallel program, t_s is the time required to execute the serial part of the program, t_p represents the time required to execute the parallel part of the program, and n represents the number of processes used in the parallel program. The value of Sp increases with increasing n, and the range is [1, n). If Sp = n, the parallel program is considered to have reached the linear speedup ratio (i.e., the ideal speedup ratio). The value range of E is (0,1), and the ideal value of E is 1, in which it is considered that the parallel program has achieved linear speedup, and the processors are fully utilized.

4. Results and Discussion

4.1. Analysis of Parallelism Efficiency

To analyze the computing efficiency of the parallel GLUE method, a GLUE-based Xinanjiang model parameter uncertainty analysis was performed with a sampling scale of 10,000. Different process numbers are set on the "Yuan" supercomputer for serial (1 process) or parallel (more than 1 process) computing. The number of processes ranges from 2 to 400 and is equal to the number of CPU cores allocated in the supercomputer. The operation time and evaluation indexes *Sp* and *E* are listed in Table 2 and Figure 3. The results show that in terms of the calculation time, the number of calculation processes increases from 1 in serial computing to 400, which is the maximum number of the case, and the total time consumption in 10,000 sampling is reduced from 7608 s to 22 s (also shown in Figure 3a). This shows that the parallel method greatly speeds up the operation speed of the GLUE method.

	Processes Number	Task Number of Single Process	Calculation Time (s)	Sp	Ε
Serial	1	10,000	7608	1	1
Parallel	2	5000	3807	1.998	0.999
	4	2500	1933	3.936	0.984
	8	1250	975	7.803	0.975
	20	500	390	19.508	0.975
	40	250	196	38.816	0.970
	50	200	159	47.849	0.957
	80	125	100	76.080	0.951
	100	100	82	92.780	0.928
	200	50	41	185.561	0.928
	400	25	22	345.818	0.865

Table 2. The calculation time and efficiency indexes of GLUE runs with different processes.



Figure 3. The execution time and computational efficiency of the parallel GLUE method with different numbers of allocated processors. (**a**) The impact of the number of processors on the execution time and (**b**) the impact of the number of allocated processors on the speedup ratio (*Sp*) and the parallelism efficiency (*E*).

In addition, as shown in Figure 3b, the speedup ratio (*Sp*) of the parallel GLUE algorithm increases with the increase in the number of processes. The maximum *Sp* reaches 345.818 when the number of processes reaches the maximum of 400. Within the scale of the case, the relationship between the increase in *Sp* and the number of processes is close to 1:1, which shows that the number of tasks of a single process still has room to go further, reflecting the high parallelism and scalability of the parallel GLUE method. The change in the parallelism efficiency (*E*) shows that with the increase in the number of processes, the relative efficiency of the parallel GLUE method always maintains a high level, which is still more than 0.9 at 200 processes and 0.865 at 400 processes. All the above indexes prove the high efficiency of the present parallel GLUE method.

As a traditional method of model parameter optimization, a parameter calibration algorithm based on various optimization algorithms also needs many repeated calculations. At present, many studies have applied parallel computing technology to improve the speed of the parameter calibration algorithm. Table 3 shows some cases of the application of parallel computing technology in parameter calibration. Although the parallel scale and efficiency improvement range of these studies are quite different, they all show the same trend, in that, within a certain number of processes, the number of opening processes and the speedup ratio are close to a linear relationship, and the parallel efficiency is close to 1. However, if the number of opening processes continues to increase, the increase in the speedup ratio and the parallel efficiency will dramatically decrease.

This is because in the area of parallel computing, the efficiency will be affected by I/O, inter-process communication, resource contention, load balancing and other factors. With the increase in the number of processes, the impact of these factors will increase significantly. At the same time, due to the influence of the inherent principle of the optimization algorithm, many inter-process communications must be carried out to compare the fitness of different populations and generate new populations. This communication intensive scenario is more likely to be affected by the efficiency of inter-process

communication when the number of open processes increases, which ultimately affects the scalability of parallel computing.

	Rouholahnejad et al. [59]	Zhang et al. [36]	Her et al. [37]	Ercan et al. [60]
Hardware	personal computer	cluster computer	cluster computer	cluster computer
Parallel tool	not mentioned	MPI	MATLAB built-in	not mentioned
Maximum processes	24	400	8	256
Maximum Sp	~10	110	6.17	~175
Parallelism efficiency	~0.4	0.275	0.77	~0.68

Table 3. Cases that use parallel technology in parameter calibration.

In contrast, the parallel GLUE algorithm in this study needs very little inter-process communication during the calculations, which is only during parameter distribution and result collection. Compared with parallel parameter calibration, the calculation is less affected by inter-process communication, and the results of the parallelism efficiency analysis in this study also prove this. The *E* value is still more than 0.9 when allocating 200 processes and 0.865 when allocating 400 processes, which reflects the high scalability of the parallel GLUE algorithm. In a similar study of the parallel GLUE algorithm, Kan et al. [52] found that the speedup ratio of the parallel GLUE algorithm performed on one 8-core CPU of a single computer is more than 10 when the sampling scale is 8000 or below; however, the speedup ratio decreases rapidly when the sampling scale exceeds 8000, and it is only 1.72 when the sampling scale is 1,024,000. This is because a single computer is very limited in CPU computing power, memory capacity, memory bandwidth and hard disk reading speed. When the computing scale is expanded, the resource contention between processes is very obvious, which seriously affects the efficiency of the parallel algorithm.

4.2. Impacts of Sampling Scale and Likelihood Threshold on the Number of Behavioral Parameter Sets

To analyze the parameter uncertainty of the Xinanjiang model based on the parallel GLUE method, this study mainly investigated the influence of the likelihood function threshold and the parameter sets sampling scale on parameter uncertainty. The threshold of the likelihood function in this study is a total of 11 levels with a range of 0.60~0.80 and an interval of 0.02. For each level, a total of 10.2 million, 1.02 million and 102,000 parameter sets were sampled from the parameter space. The maximum computing resources occupied are 100 nodes and 2400 logical CPUs, and the total computing power is approximately 96 Tflops. The sampling method is the Latin hypercube sampling (LHS) method [61].

The number of behavioral parameter sets and their proportion in the total sample scale under 3 different parameter sets sampling scales and 11 different NSE likelihood thresholds are shown in Table 4 and Figure 4. It is shown that the numbers of behavioral parameter sets obtained by different sampling sizes under the same likelihood function threshold are different, but the proportions of the behavioral parameter sets in each sample size are basically the same. For example, when the threshold is 0.62, the numbers of behavioral parameter sets when the sampling scales are 10.2 million, 1.02 million and 0.102 million are 7105, 724 and 79, respectively, but the proportions of the behavioral parameter sets remain steady, the larger the sample size is, the more behavioral parameter sets there are. On one hand, a larger number of behavioral parameter sets can better reflect the posterior distribution of parameters, which is the distribution of parameters under the situation that the likelihood value is above the threshold, under a certain threshold but, on the other hand, can also provide more forecasting members for future hydrological ensemble forecasting research.

Sampling Scale (Million)	10.2		1.02		0.102	
Threshold Value	Number of Behavioral Parameter Sets	Proportion (%)	Number of Behavioral Parameter Sets	Proportion (%)	Number of Behavioral Parameter Sets	Proportion (%)
0.80	0	0	0	0	0	0
0.78	2	$1.96 imes 10^{-5}$	0	0	0	0
0.76	38	3.73×10^{-4}	3	2.94×10^{-4}	0	0
0.74	128	1.26×10^{-3}	12	1.17×10^{-3}	1	9.80×10^{-4}
0.72	376	3.69×10^{-3}	41	4.02×10^{-3}	1	9.80×10^{-4}
0.70	808	7.92×10^{-3}	81	7.94×10^{-3}	10	9.80×10^{-3}
0.68	1590	1.56×10^{-2}	162	1.59×10^{-2}	19	1.86×10^{-2}
0.66	2772	2.72×10^{-2}	278	2.73×10^{-2}	24	2.35×10^{-2}
0.64	4794	4.70×10^{-2}	477	4.68×10^{-2}	43	4.22×10^{-2}
0.62	7105	0.0697	724	0.0701	79	0.0774
0.60	11,533	0.113	1147	0.112	130	0.127

Table 4. The number of behavioral parameter sets obtained under different sampling sizes and Nash–Sutcliffe efficiency (NSE) threshold values.



Figure 4. The proportion of behavioral parameter sets under different sampling sizes and likelihood threshold values. The black, red and blue lines refer to the sampling sizes of 10.2 million, 1.02 million and 0.102 million, respectively.

The results also show that the larger the sample size, the more likely it is to obtain the behavioral parameter sets under the better likelihood function threshold conditions. As shown in Table 4, the test performed with 10.2 million samples obtained 2 behavioral parameter sets at the likelihood function threshold of 0.78, but no behavioral parameter set was obtained in the 1.02 million and 102,000 samples at the same threshold condition. Moreover, the results show that if the behavioral parameter sets are obtained under the condition of a higher likelihood function threshold, the parameters can better reflect the posterior distribution of the model parameters when they approach the global optimal solution, which is the best simulation performance of all the possible parameter sets and can also provide high-quality forecast members for future hydrological ensemble forecasting.

Figure 4 also shows that the proportion of behavioral parameter sets increases exponentially with the decrease in the likelihood function, which is NSE in this study, threshold. In the existing research on parameter uncertainty using the GLUE method, Li et al. [28] and Xue et al. [62] have studied the relationship between the likelihood function threshold in the GLUE method and the proportion of behavioral parameter sets. It is found that there is a good linear relationship between the likelihood function of behavioral parameter sets. Note that the main focus of these studies is on a relatively macro range where the proportion of behavioral parameter sets is greater than 0.1%. In this study, due to the large scale of parameter sets sampling, we mainly investigate the micro interval where the proportion of behavioral parameter sets is less than 0.1%. This shows that there is a

different relationship between the threshold value and the proportion of behavioral parameter sets around the tipping point of 0.1%, which is a linear relationship in the range of more than 0.1% and an exponential relationship in the interval of less than 0.1%.

4.3. Impact of the Sampling Scale and Likelihood Threshold on the Parameter Posterior Distribution

Figure 5 illustrates the comparison of the posterior distribution of the sensitivity parameters derived from the experiment in Section 4.2 for the sampling scale of 10.2 million and the likelihood thresholds of 0.66 and 0.74. There are several aspects that can be interpreted from the results. First, the types of sensitivity parameters remain consistent in different threshold situations and thus in the sampling situations (plot not given). For both threshold values, the Xinanjiang model obtained the same 9 sensitive parameters through the parallel GLUE method. Second, the posterior distribution of some sensitivity parameters changes when the threshold value changes, while those of others remain unchanged. As shown in Figure 5, some of the sensitivity parameters, i.e. IMP, WM3, KKSS and PETM, get more concentrated posterior distributions when the threshold value increases from 0.66 to 0.74. However, the sampling scale has no impact on the posterior distributions. Third, the scatter plots in Figure 5 show that there is no linear relationship among the sensitive parameters, regardless of the threshold values or the sampling scale.



Figure 5. The posterior distribution and the joint relationships of the sensitivity parameters of the Xinanjiang model generated by the parallel GLUE method with the likelihood threshold value of 0.66 and 0.74. The scatter plots show the joint relationships of the sensitivity parameters within behavioral parameter sets. The plots in the diagonal show the posterior distribution of the sensitivity parameters. The blue and red dots indicate the parameter values when the threshold values are 0.66 and 0.74, respectively.

5. Conclusions

In this study, an MPI-based parallel computing accelerated GLUE method was proposed and implemented in a supercomputing system. A smaller scale experiment was performed with different processors to test the computational efficiency of the parallel GLUE method. A larger scale experiment was performed to analyze the impact of the sampling size and the threshold value on the parameter uncertainty of the Xinanjiang model. The following conclusions were drawn from the study:

- 1. The MPI-based parallel GLUE method can significantly improve the calculation efficiency and shorten the calculation time. For the sampling size of 100,000, the parallelism efficiency remains at 0.865, and the speedup ratio is close to a linear speedup ratio, even if the allocated processors are up to 400. The results show the great scalability of the present method.
- 2. The proportion of behavioral parameter sets remains unchanged as the sampling scale changes. Within the interval of less than 0.1%, the proportion of behavioral parameter sets and the threshold value have an exponential relationship. The larger the sampling scale, the more likely it is that behavioral parameter sets are obtained at high threshold values.
- 3. The threshold values impact the posterior distribution of the sensitivity parameters. Some parameters may yield more concentrated posterior distributions when the threshold is higher.

Author Contributions: Conceptualization, H.W. and X.L.; methodology, Z.Y.; software, Z.Y.; validation, Z.Y.; resources, W.L.; data curation, W.L.; writing—original draft preparation, Z.Y.; writing—review and editing, W.L.; visualization, Z.Y.; supervision, H.W.; project administration, H.W.; funding acquisition, X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This paper was jointly supported by the National Key R&D Program of China (2017YFB0203104), the National Natural Science Fund (51709273) and the project of Power Construction Corporation of China (DJ-ZDZX-2016-02).

Acknowledgments: The results described in this paper are obtained on the China National Grid (http://www.cngrid.org)/China Scientific Computing Grid (http://www.scgrid.cn).

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

References

- 1. Wen, X.; Liu, Z.; Lei, X.; Lin, R.; Fang, G.; Tan, Q.; Wang, C.; Tian, Y.; Quan, J. Future changes in Yuan River ecohydrology: Individual and cumulative impacts of climates change and cascade hydropower development on runoff and aquatic habitat quality. *Sci. Total Environ.* **2018**, *633*, 1403–1417. [CrossRef] [PubMed]
- 2. Yin, Z.; Liao, W.; Lei, X.; Wang, H.; Wang, R. Comparing the Hydrological Responses of Conceptual and Process-Based Models with Varying Rain Gauge Density and Distribution. *Sustainability* **2018**, *10*. [CrossRef]
- Lei, X.; Tian, Y.; Liao, W.; Bai, W.; Jia, Y.W.; Jiang, Y.Z.; Wang, H. Development of an AutoWEP distributed hydrological model and its application to the upstream catchment of the Miyun Reservoir. *Comput. Geosci.* 2012, 44, 203–213. [CrossRef]
- 4. Lei, X.; Liao, W.; Wang, Y.; Jiang, Y.; Wang, H.; Tian, Y. Development and Application of a Distributed Hydrological Model: EasyDHM. *J. Hydrol. Eng.* **2014**, *19*, 44–59. [CrossRef]
- Wang, M.; Lei, X.; Liao, W.; Shang, Y. Analysis of changes in flood regime using a distributed hydrological model: A case study in the Second Songhua River basin, China. *Int. J. Water Resour. Dev.* 2018, 34, 386–404. [CrossRef]
- Bartholmes, J.C.; Thielen, J.; Ramos, M.H.; Gentilini, S. The european flood alert system EFAS—Part 2: Statistical skill assessment of probabilistic and deterministic operational forecasts. *Hydrol. Earth Syst. Sci.* 2009, *13*, 141–153. [CrossRef]
- 7. Thiboult, A.; Anctil, F.; Boucher, M.-A. Accounting for three sources of uncertainty in ensemble hydrological forecasting. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 1809–1825. [CrossRef]
- 8. Krueger, T.; Page, T.; Hubacek, K.; Smith, L.; Hiscock, K. The role of expert opinion in environmental modelling. *Environ. Model. Softw.* **2012**, *36*, 4–18. [CrossRef]
- 9. Beck, M.; Krueger, T. The epistemic, ethical, and political dimensions of uncertainty in integrated assessment modeling. *Wiley Interdiscip. Rev. Clim. Chang.* **2016**, *7*, 627–645. [CrossRef]

- Melsen, L.A.; Teuling, A.J.; Torfs, P.J.J.F.; Zappa, M.; Mizukami, N.; Mendoza, P.A.; Clark, M.P.; Uijlenhoet, R. Subjective modeling decisions can significantly impact the simulation of flood and drought events. *J. Hydrol.* 2019, 568, 1093–1104. [CrossRef]
- 11. Hassanzadeh, Y.; Afshar, A.A.; Pourreza-Bilondi, M.; Memarian, H.; Besalatpour, A.A. Toward a combined Bayesian frameworks to quantify parameter uncertainty in a large mountainous catchment with high spatial variability. *Environ. Monit. Assess.* **2019**, *191*. [CrossRef] [PubMed]
- 12. Liu, Y.R.; Li, Y.P.; Huang, G.H.; Zhang, J.L.; Fan, Y.R. A Bayesian-based multilevel factorial analysis method for analyzing parameter uncertainty of hydrological model. *J. Hydrol.* **2017**, *553*, 750–762. [CrossRef]
- 13. Pang, B.; Yue, J.; Huang, Z.; Zhang, R. Parameter uncertainty assessment of a flood forecasting model using multiple objectives. *J. Flood Risk Manag.* **2019**, 12. [CrossRef]
- Zadeh, F.K.; Nossent, J.; Woldegiorgis, B.T.; Bauwens, W.; van Griensven, A. Impact of measurement error and limited data frequency on parameter estimation and uncertainty quantification. *Environ. Model. Softw.* 2019, 118, 35–47. [CrossRef]
- Huang, X.; Liao, W.; Lei, X.; Jia, Y.; Wang, Y.; Wang, X.; Jiang, Y.; Wang, H. Parameter optimization of distributed hydrological model with a modified dynamically dimensioned search algorithm. *Environ. Model. Softw.* 2014, *52*, 98–110. [CrossRef]
- Bomhof, J.; Tolson, B.A.; Kouwen, N. Comparing single and multi-objective hydrologic model calibration considering reservoir inflow and streamflow observations. *Can. Water Resour. J.* 2019, 44, 319–336. [CrossRef]
- Fang, W.; Huang, S.; Ren, K.; Huang, Q.; Huang, G.; Cheng, G.; Li, K. Examining the applicability of different sampling techniques in the development of decomposition-based streamflow forecasting models. *J. Hydrol.* 2019, 568, 534–550. [CrossRef]
- 18. Zhou, Q.; Chen, L.; Singh, V.P.; Zhou, J.; Chen, X.; Xiong, L. Rainfall-runoff simulation in karst dominated areas based on a coupled conceptual hydrological model. *J. Hydrol.* **2019**, *573*, 524–533. [CrossRef]
- 19. Beven, K.; Binley, A. Future of distributed models: Model calibration and uncertainty prediction. *Hydrol. Process.* **1992**, *6*, 279–298. [CrossRef]
- 20. Kuczera, G.; Parent, E. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: The Metropolis algorithm. *J. Hydrol.* **1998**, *211*, 69–85. [CrossRef]
- 21. Van Griensven, A.; Meixner, T. Methods to quantify and identify the sources of uncertainty for river basin water quality models. *Water Sci. Technol.* **2006**, *53*, 51–59. [CrossRef] [PubMed]
- 22. Abbaspour, K.C.; Johnson, C.A.; van Genuchten, M.T. Estimating uncertain flow and transport parameters using a sequential uncertainty fitting procedure. *Vadose Zone J.* **2004**, *3*, 1340–1352. [CrossRef]
- 23. Kong, X.; Li, Z.; Liu, Z. Flood Prediction in Ungauged Basins by Physical-Based TOPKAPI Model. *Adv. Meteorol.* **2019**, 2019. [CrossRef]
- Xiang, Y.; Li, L.; Chen, J.; Xu, C.-Y.; Xia, J.; Chen, H.; Liu, J. Parameter Uncertainty of a Snowmelt Runoff Model and Its Impact on Future Projections of Snowmelt Runoff in a Data-Scarce Deglaciating River Basin. *Water* 2019, 11. [CrossRef]
- Smith, K.A.; Barker, L.J.; Tanguy, M.; Parry, S.; Harrigan, S.; Legg, T.P.; Prudhomme, C.; Hannaford, J. A multi-objective ensemble approach to hydrological modelling in the UK: An application to historic drought reconstruction. *Hydrol. Earth Syst. Sci.* 2019, 23, 3247–3268. [CrossRef]
- 26. Blasone, R.-S.; Madsen, H.; Rosbjerg, D. Uncertainty assessment of integrated distributed hydrological models using GLUE with Markov chain Monte Carlo sampling. *J. Hydrol.* **2008**, *353*, 18–32. [CrossRef]
- 27. Stedinger, J.R.; Vogel, R.M.; Lee, S.U.; Batchelder, R. Appraisal of the generalized likelihood uncertainty estimation (GLUE) method. *Water Resour. Res.* **2008**, *44*. [CrossRef]
- 28. Li, L.; Xia, J.; Xu, C.-Y.; Singh, V.P. Evaluation of the subjective factors of the GLUE method and comparison with the formal Bayesian method in uncertainty assessment of hydrological models. *J. Hydrol.* **2010**, *390*, 210–221. [CrossRef]
- 29. Jin, X.; Xu, C.-Y.; Zhang, Q.; Singh, V.P. Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model. *J. Hydrol.* **2010**, *383*, 147–155. [CrossRef]
- Wu, Q.; Spiryagin, M.; Cole, C.; McSweeney, T. Parallel computing in railway research. *Int. J. Rail Transp.* 2020, *8*, 111–134. [CrossRef]
- 31. Peng, X.; Yu, P.; Chen, G.; Xia, M.; Zhang, Y. CPU-accelerated explicit discontinuous deformation analysis and its application to landslide analysis. *Appl. Math. Model.* **2020**, *77*, 216–234. [CrossRef]

- Wang, X.; Feng, L.; Zhao, H. Fast image encryption algorithm based on parallel computing system. *Inf. Sci.* 2019, 486, 340–358. [CrossRef]
- 33. Huang, F.; Tie, B.; Tao, J.; Tan, X.; Ma, Y. Methodology and optimization for implementing cluster-based parallel geospatial algorithms with a case study. *Cluster Comput.* **2020**, *23*, 673–704. [CrossRef]
- 34. Zhang, S.; Li, M.; Chen, Z.; Huang, T.; Li, S.; Li, W.; Chen, Y. Parallel Spatial-Data Conversion Engine: Enabling Fast Sharing of Massive Geospatial Data. *Symmetry* **2020**, *12*. [CrossRef]
- 35. Vrugt, J.A.; Nuallain, B.O.; Robinson, B.A.; Bouten, W.; Dekker, S.C.; Sloot, P.M.A. Application of parallel computing to stochastic parameter estimation in environmental models. *Comput. Geosci.* 2006, *32*, 1139–1155. [CrossRef]
- Zhang, X.; Beeson, P.; Link, R.; Manowitz, D.; Izaurralde, R.C.; Sadeghi, A.; Thomson, A.M.; Sahajpal, R.; Srinivasan, R.; Arnold, J.G. Efficient multi-objective calibration of a computationally intensive hydrologic model with parallel computing software in Python. *Environ. Model. Softw.* 2013, 46, 208–218. [CrossRef]
- 37. Her, Y.; Cibin, R.; Chaubey, I. Application of Parallel Computing Methods for Improving Efficiency of Optimization in Hydrologic and Water Quality Modeling. *Appl. Eng. Agric.* **2015**, *31*, 455–468. [CrossRef]
- Kan, G.; Lei, T.; Liang, K.; Li, J.; Ding, L.; He, X.; Yu, H.; Zhang, D.; Zuo, D.; Bao, Z.; et al. A Multi-Core CPU and Many-Core GPU Based Fast Parallel Shuffled Complex Evolution Global Optimization Approach. *IEEE Trans. Parallel Distrib. Syst.* 2017, 28, 332–344. [CrossRef]
- 39. Huo, J.; Liu, L.; Zhang, Y. An improved multi-cores parallel artificial Bee colony optimization algorithm for parameters calibration of hydrological model. *Future Gener. Comput. Syst.* **2018**, *81*, 492–504. [CrossRef]
- 40. Apostolopoulos, T.K.; Georgakakos, K.P. Parallel computation for streamflow prediction with distributed hydrologic models. *J. Hydrol.* **1997**, *197*, *1–24*. [CrossRef]
- 41. Vivoni, E.R.; Mascaro, G.; Mniszewski, S.; Fasel, P.; Springer, E.P.; Ivanov, V.Y.; Bras, R.L. Real-world hydrologic assessment of a fully-distributed hydrological model in a parallel computing environment. *J. Hydrol.* **2011**, 409, 483–496. [CrossRef]
- 42. Li, T.; Wang, G.; Chen, J.; Wang, H. Dynamic parallelization of hydrological model simulations. *Environ. Model. Softw.* **2011**, *26*, 1736–1746. [CrossRef]
- 43. Wang, H.; Fu, X.; Wang, G.; Li, T.; Gao, J. A common parallel computing framework for modeling hydrological processes of river basins. *Parallel Comput.* **2011**, *37*, 302–315. [CrossRef]
- 44. Liu, J.; Zhu, A.X.; Liu, Y.; Zhu, T.; Qin, C.-Z. A layered approach to parallel computing for spatially distributed hydrological modeling. *Environ. Model. Softw.* **2014**, *51*, 221–227. [CrossRef]
- 45. Zhang, F.; Zhou, Q. Parallelization of the flow-path network model using a particle-set strategy. *Int. J. Geogr. Inf. Sci.* **2019**, *33*, 1984–2010. [CrossRef]
- Sanders, B.F.; Schubert, J.E.; Detwiler, R.L. ParBreZo: A parallel, unstructured grid, Godunov-type, shallow-water code for high-resolution flood inundation modeling at the regional scale. *Adv. Water Resour.* 2010, *33*, 1456–1467. [CrossRef]
- Wang, X.; Shangguan, Y.; Onodera, N.; Kobayashi, H.; Aoki, T. Direct Numerical Simulation and Large Eddy Simulation on a Turbulent Wall-Bounded Flow Using Lattice Boltzmann Method and Multiple GPUs. *Math. Probl. Eng.* 2014, 2014. [CrossRef]
- 48. Wang, Y.; Yang, X. Sensitivity Analysis of the Surface Runoff Coefficient of HiPIMS in Simulating Flood Processes in a Large Basin. *Water* **2018**, *10*. [CrossRef]
- 49. Liu, Q.; Qin, Y.; Li, G. Fast Simulation of Large-Scale Floods Based on GPU Parallel Computing. *Water* **2018**, *10*. [CrossRef]
- 50. Hwang, H.T.; Park, Y.J.; Sudicky, E.A.; Forsyth, P.A. A parallel computational framework to solve flow and transport in integrated surface-subsurface hydrologic systems. *Environ. Model. Softw.* **2014**, *61*, 39–58. [CrossRef]
- Kuffour, B.N.O.; Engdahl, N.B.; Woodward, C.S.; Condon, L.E.; Kollet, S.; Maxwell, R.M. Simulating coupled surface-subsurface flows with ParFlow v3.5.0: Capabilities, applications, and ongoing development of an open-source, massively parallel, integrated hydrologic model. *Geosci. Model. Dev.* 2020, *13*, 1373–1397. [CrossRef]
- 52. Kan, G.; He, X.; Ding, L.; Li, J.; Hong, Y.; Liang, K. Heterogeneous parallel computing accelerated generalized likelihood uncertainty estimation (GLUE) method for fast hydrological model uncertainty analysis purpose. *Eng. Comput.* **2019**, *36*, 75–96. [CrossRef]

- 53. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models part I—A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290. [CrossRef]
- Thakur, R.; Gropp, W.D. Improving the performance of collective operations in MPICH. In *Recent Advances in Parallel Virtual Machine and Message Passing Interface*; Dongarra, J., Laforenza, D., Orlando, S., Eds.; Springer: Berlin, Germany, 2003; Volume 2840, pp. 257–267.
- 55. Thakur, R.; Rabenseifner, R.; Gropp, W. Optimization of collective communication operations in MPICH. *Int. J. High Perform. Comput. Appl.* **2005**, *19*, 49–66. [CrossRef]
- Buntinas, D.; Mercier, G.; Gropp, W. Implementation and shared-memory evaluation of MPICH2 over the nemesis communication subsystem. In *Recent Advances in Parallel Virtual Machine and Message Passing Interface*; Mohr, B., Traff, J.L., Worringen, J., Dongarra, J., Eds.; Springer: Berlin, Germany, 2006; Volume 4192, pp. 86–95.
- 57. Zhao, R.-J.; Zuang, Y.-L.; Fang, L.-R.; Liu, X.-R.; Zhang, Q.-S. XINANJIANG MODEL. In Proceedings of the Hydrology Forecast Symposium, Oxford, UK, 15–18 April 1980; pp. 351–356.
- Xiao, H.; Wu, H.; Chi, X. SCE: Grid Environment for Scientific Computing. In *Networks for Grid Applications*; Vicat-Blanc Primet, P., Kudoh, T., Mambretti, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; Volume 2, pp. 35–42.
- 59. Rouholahnejad, E.; Abbaspour, K.C.; Vejdani, M.; Srinivasan, R.; Schulin, R.; Lehmann, A. A parallelization framework for calibration of hydrological models. *Environ. Model. Softw.* **2012**, *31*, 28–36. [CrossRef]
- 60. Ercan, M.B.; Goodall, J.L.; Castronova, A.M.; Humphrey, M.; Beekwilder, N. Calibration of SWAT models using the cloud. *Environ. Model. Softw.* **2014**, *62*, 188–196. [CrossRef]
- 61. McKay, M.D.; Beckman, R.J.; Conover, W.J. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **2000**, *42*, 55–61. [CrossRef]
- 62. Xue, C.; Chen, B.; Wu, H. Parameter Uncertainty Analysis of Surface Flow and Sediment Yield in the Huolin Basin, China. *J. Hydrol. Eng.* **2014**, *19*, 1224–1236. [CrossRef]



© 2020 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/).