

Article

Performance Assessment of Hydrological Models Considering Acceptable Forecast Error Threshold

Qianjin Dong ^{1,2,*} and Fan Lu ³

¹ State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China

² Key Laboratory of Poyang Lake Wetland and Watershed Research, Ministry of Education, Jiangxi Normal University, Nanchang 330022, China

³ Department of Water Resources, China Institute of Water Resources and Hydropower Research, Beijing 100038, China; E-Mail: lf9805320@163.com

* Author to whom correspondence should be addressed; E-Mail: dqjin@whu.edu.cn; Tel.: +86-27-6877-2209.

Academic Editor: Miklas Scholz

Received: 13 August 2015 / Accepted: 29 October 2015 / Published: 4 November 2015

Abstract: It is essential to consider the acceptable threshold in the assessment of a hydrological model because of the scarcity of research in the hydrology community and errors do not necessarily cause risk. Two forecast errors, including rainfall forecast error and peak flood forecast error, have been studied based on the reliability theory. The first order second moment (FOSM) and bound methods are used to identify the reliability. Through the case study of the Dahuofang (DHF) Reservoir, it is shown that the correlation between these two errors has great influence on the reliability index of hydrological model. In particular, the reliability index of the DHF hydrological model decreases with the increasing correlation. Based on the reliability theory, the proposed performance evaluation framework incorporating the acceptable forecast error threshold and correlation among the multiple errors can be used to evaluate the performance of a hydrological model and to quantify the uncertainties of a hydrological model output.

Keywords: assessment; rainfall forecast errors; peak flood forecast errors; reliability analysis; first order second moment (FOSM); bound method; Dahuofang reservoir (DHF)

1. Introduction

The forecast error, which is usually defined as difference between the simulated and observed streamflow, could benefit the assessment of a hydrological model using certain criteria like Root Mean Square Error (RMSE) and the Nash–Sutcliffe coefficient [1]. However, these criteria only represent the combined effects of the model bias and the model uncertainty because only residuals are considered [2]. Consequently, they are unable to indicate how well the hydrological function of a system is maintained [3] and do not account for heteroscedasticity and serial dependence of the residuals [4–6]. As a result, the traditional criteria have limited ability to quantifying the reliability of a hydrological model.

On the other hand, the forecast error does not necessarily cause risk, since the forecast error has an acceptance threshold, which has rarely been discussed in previous studies. An acceptance threshold of forecast error is defined as an interval in which the forecast error is within the interval of the acceptance threshold, the hydrological model is reliable and may have potential application; otherwise, the hydrological model is not accepted. Thus, the acceptable threshold of forecast error could support a much broader perspective for the assessment of hydrological model.

Although the serial dependency of respective forecast error for a hydrological model has been widely studied [4,5,7–14], the research of inter-correlation of forecast errors (e.g., rainfall forecast error and flood forecast error) is limited. Krzysztofowicz [15] presented an analytic numerical method combining the precipitation uncertainty and hydrologic uncertainty to forecast river stages, which is a mixture of two distributions related to occurrence and nonoccurrence of precipitation. Hostache *et al.* [16] assessed the predictive uncertainty of the coupled atmospheric-hydrologic-error correction model, which used a bivariate meta-Gaussian probability density function. The latter enabled the computation of confidence intervals in a normalized space, and the results demoed the validity of the model. Nester *et al.* [17] analyzed the scaling relationship with catchment area between the precipitation forecast error and hydrological simulation error. Yazdi *et al.* [18] proposed a stochastic methodology based on Monte Carlo simulation and multivariate analysis to provide the uncertainty band of the rainfall–runoff model and to calculate the probability of the acceptable forecasts. They discussed the acceptable error for threshold discharge (amplitude error) and the acceptable error for the time (phase error). Dogulu *et al.* [19] focused on the methods predicting model residual uncertainty based on the quantile regression and UNcertainty Estimation based on local Errors and Clustering (UNEEC) and compared the performance of the methods. Although many studies focus on the uncertainties of the hydrological model, the study on the rainfall forecast error and peak flood forecast error is rare.

Considering the classical rainfall–runoff model, the rainfall is transformed to effective rainfall after the interception (e.g., canopy interception, soil detention). Therefore, the term of rainfall forecast error appears because of the uncertainties during runoff yield, such as the simplified calculation of evaporation, canopy interception, soil detention. During the runoff confluence, the peak flood forecast error appears for the simplified physical catchment, which the hydrological model simulated and the complex interaction among the soil–surface–groundwater. Thus, we will explore the inter-correlation between the rainfall forecast error and peak flood forecast error, and try to present an improved approach to assessing the performance of the hydrological model and considering the inter-correlation of forecast errors. Therefore, the following problems we focus on in this study are:

- (1) What is the correlation between the rainfall forecast error and peak flood forecast error, is it statistically significant?
- (2) How to calculate the reliability index of a hydrological model based on the reliability theory, under the context of correlated forecast errors considering two scenarios, single failure mode and double failure mode (see Section 4)?
- (3) What is the effect of different types of the performance function on the reliability index of a hydrological model?
- (4) What is the change law for the reliability index of a hydrological model with regard to the variable correlation between the two forecast errors?

To address these questions, we propose a framework to assess the performance of the hydrological model based on the reliability theory, which has been widely applied in the civil engineering field [20–24].

The reminder of this paper is organized as follows: Section 2 starts with a description of reliability theory, including FOSM and bound methods. Section 3 describes the study area and error samples, and the results are then presented and discussed in Section 4. Finally, the conclusion is drawn in Section 5.

2. Material and Methods

2.1. Simple Introduction to Reliability Theory

Reliability is defined as the probabilistic measure of whether a system meets certain standards [25,26]. The reliability theory analyzes the relationship between the load and the resistance of a problem by using a reliability index and/or a failure probability. In the context of a hydrological model performance assessment problem, the forecast error, which is the output of the hydrological model, could be regarded as the load of the model, and the acceptable threshold could be regarded as the resistance. This is because the acceptable threshold could reflect some features of certain physical catchment, thus may also give reference to assessing the hydrological model. The forecast error (load) should meet the acceptable threshold (resistance) if the hydrological model works well and can be applied for practice. Otherwise, the failure probability of the event can be defined as follows. For example, the event is whether the forecast error of a hydrological model meets the acceptable threshold of the forecast error for certain catchment.

$$p_f = P\{x_r \leq g(x)\} = \int \int_{x_r \leq g(x)} f_{g,r}(G, R) dGdR \quad (1)$$

where x_r is the acceptable error resistance, $g(x)$ is the error load, and $f_{g,r}(G, R)$ is the joint distribution of the error load and the acceptable error resistance. The distribution on $f_{g,r}(G, R)$ can be obtained by using either numeric methods (e.g., Monte-Carlo sampling methods [27,28]) or analytic methods.

Assuming that the failure probability is based on a specific performance criteria, the limit state corresponding to Equation (1), which represents a boundary between safe and unsafe regions on the parametric space, is:

$$Z = G(x_1, x_2, \dots, x_n) = x_r - g(x_1, x_2, \dots, x_n) = 0 \quad (2)$$

Equation (2) is the performance function of a hydrological model based on forecast error information. The hydrological model is either reliable and its accuracy is acceptable or the hydrological model is unreliable. Solutions of Equation (2) are reference points.

If Z follows the normal distribution $N(\mu_Z, \sigma_Z)$, the reliability index β can be obtained as follows.

$$\beta = \frac{\mu_Z}{\sigma_Z} \tag{3}$$

Thus, the relationship between β and p_f (Ganji & Jowkarshorijeh, 2012) is as follows:

$$p_f = \Phi(-\beta) = 1 - \Phi(\beta) \tag{4}$$

where $\Phi()$ is a standard normal probability distribution function.

Equation (4) is derived under the condition that the performance function follows a normal distribution. However, this situation does not valid for a large set of random variables with arbitrary distributions. Hence, many approximation methods for calculating the reliability index have been developed, such as first order second moment (FOSM) method [29], second order second moment (SOSM) method [30], second order fourth moment (SOFM) method [31], and responding surface method (RSM) [32]. The FOSM method is adopted in this study since it is the simplest and most convenient method in the reliability theory among the above-mentioned methodologies [33].

2.2. FOSM Method

FOSM is a well-known method used in water resources engineering, which is originally developed to evaluate the safety of the structure systems. The FOSM method takes the first two statistical moments of a linear approximation of the performance function and attempts to find the minimal distance from the given nominal point to the tangent hyper-plane. This distance provides a measure of reliability for evaluation of hydrological model through the reliability index.

Without loss of generality, consider a function G of interest of several random variables X_1, X_2, \dots, X_n , with mean $\mu_X = (\mu_{X_1}, \mu_{X_2}, \dots, \mu_{X_n})$ and standard deviation $\sigma_X = (\sigma_{X_1}, \sigma_{X_2}, \dots, \sigma_{X_n})$ (X is the basic variables in the reliability theory). G transforms the random variable into a random variable, $Z = G(X_1, X_2, \dots, X_n)$, which is the performance function in the reliability theory. The linearized form of Z about the mean value μ_X is defined as follows:

$$Z \approx G_X(\mu_X) + \sum_{i=1}^n \frac{\partial G_X(\mu_X)}{\partial X_i} (X_i - \mu_{X_i}) \tag{5}$$

in which $\frac{\partial G_X(\mu_X)}{\partial X_i}$ is the gradient vector of G and describes the sensitivity of Z to the input variable X , and the mean (the first moment) and variance (the second moment) of Z can be computed as follows, respectively:

$$\mu_Z \approx G_X(\mu_X) \tag{6}$$

$$\sigma_Z^2 \approx \sum_{i=1}^n \left[\frac{\partial G_X(\mu_X)}{\partial X_i} \right]^2 \sigma_{X_i}^2 \tag{7}$$

The FOSM method requires that the input variables X are statistically independent. If input variables are dependent, the transformation of input variables is needed. The common transformation in engineering is orthogonal transformation, which is detailed in Shinozuka [34], other transformation includes Rosenblatt transformation [35], Nataf transformation [36], Hermite polynomials transformation [37], and so on. The orthogonal transformation was adopted in this study to deal with the correlated input variables because of its robust feature and compact formulation.

2.3. Ditlevsen’s Bounds Method

Ditlevsen’s bounds [38] method is widely used for calculating system failure probability. This method is often reasonably accurate and computationally more efficient than Monte-Carlo simulation. Considering a general structural system, the failure of the system E is modeled as the union of all m possible failure modes with the failure events $E_i (i = 1, \dots, m)$, $E = E_1 \cup E_2 \cup \dots \cup E_m = \cup_{i=1}^m E_i$. Suppose the joint density probability function of the basic variables X is $f_X(x)$, then the failure probability of the systems E is:

$$p_f = Pr(E) = \int_E f_X(x) dx \tag{8}$$

in which $Pr(\)$ denotes the failure probability.

Generally, the structure system includes series structure, parallel structure and mix of series and parallel structure, according to the relationship between the system and its members. The series structure is a structure in which the failure of any member of the system will lead to an overall failure of the system. The parallel structure, on the other hand, is a structure in which the system fails only if all its members fail. Denote the failure probability of E_i as p_{fi} ; the failure probability of system p_f can be obtained according to Cornell [39], which is named wide bounds approach.

$$\max_{1 \leq i \leq m} p_{fi} \leq p_f \leq 1 - \prod_{i=1}^m (1 - p_{fi}) \tag{9}$$

where m is the number of event E_i when it is failure.

Denoting the failure probability of two failure modes failing simultaneously as $p_{fij} = Pr(E_i E_j)$, the failure probability of the systems p_f , according to the narrow bound [38,40], is

$$p_{f1} + \sum_{i=2}^m \max \left(p_{fi} - \sum_{j=1}^{i-1} p_{fij}, 0 \right) \leq p_f \leq \sum_{i=1}^m p_{fi} - \sum_{i=2}^m \max_{j < i} p_{fij} \tag{10}$$

For the parallel case, the bound for the failure probability of systems could be expressed as:

$$\prod_{i=1}^q p_{fi} \leq p_f \leq \min_{1 \leq i, j \leq q} p_{fij} \tag{11}$$

where q is the failure status numbers of the total failure events $E_U = \cup_{i,j=1}^m (E_i \cap E_j)$.

3. Study Area

The Dahuofang (DHF) Reservoir is located in the middle-to-lower course of the Hun River, between $41^{\circ}31' \text{ N}$ – $42^{\circ}15' \text{ N}$ and $120^{\circ}20' \text{ E}$ – $125^{\circ}15' \text{ E}$, in Liaoning Province, northeast of China (see Figure 1). It is a large-scale hydraulic project, mainly used for flood defense and irrigation. It controls an area of 5437 km^2 . The area has a mean annual rainfall of over 800 mm, generally concentrating in July and August, and the floods are driven mainly by storms. The DHF model was used to forecast floods with 3-h time steps, which is a hybrid conceptual rainfall–runoff model, consisting of two parts. The runoff yield part has an eight-parameter excess infiltration runoff model, based on the Horton curve [41] and the double-layered infiltration curve for reductive calculation, while using parabolas to describe the upper amount of water storage and the distribution of double-layered infiltration. The runoff confluence part has an eight-parameter empirical unit hydro-graph convergence model with variable intensity and variable confluence velocity, which utilizes exponents and triangular function production to describe the empirical unit hydro-graph, and antecedent rainfall, reflecting the velocity variations in confluence. The structure of the DHF model is shown in Figure 2, and its parameters are listed in Table 1.

In this study, the DHF's 31 floods from 1951 to 1996 were analyzed. The frequency of floods in this period is listed in Table 2, which shows the years having more than two floods [42]. Here, the rainfall forecast errors denotes the errors of effective precipitation for the set of 31 floods, and corresponding peak flood forecast error.

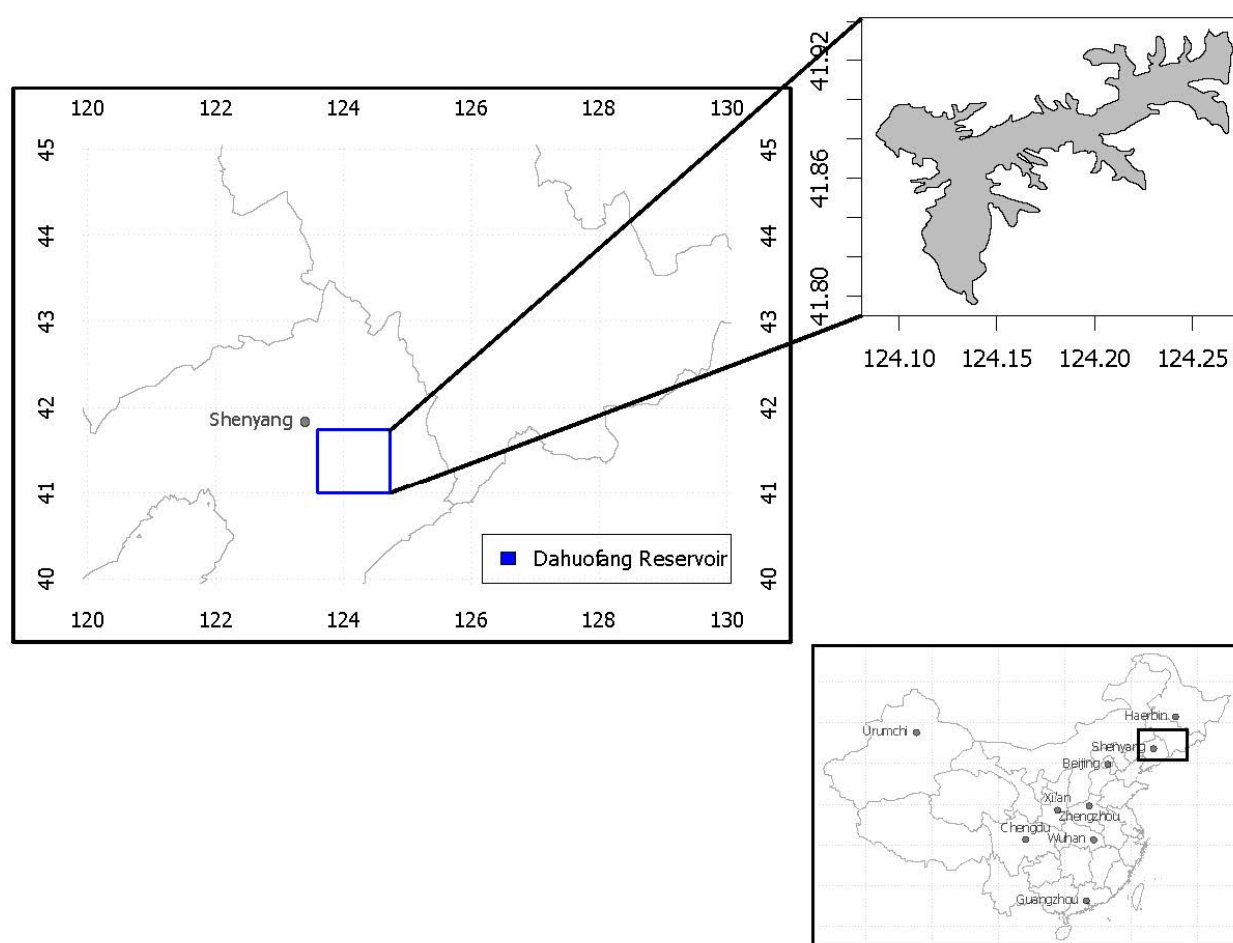


Figure 1. Location of DHF Reservoir.

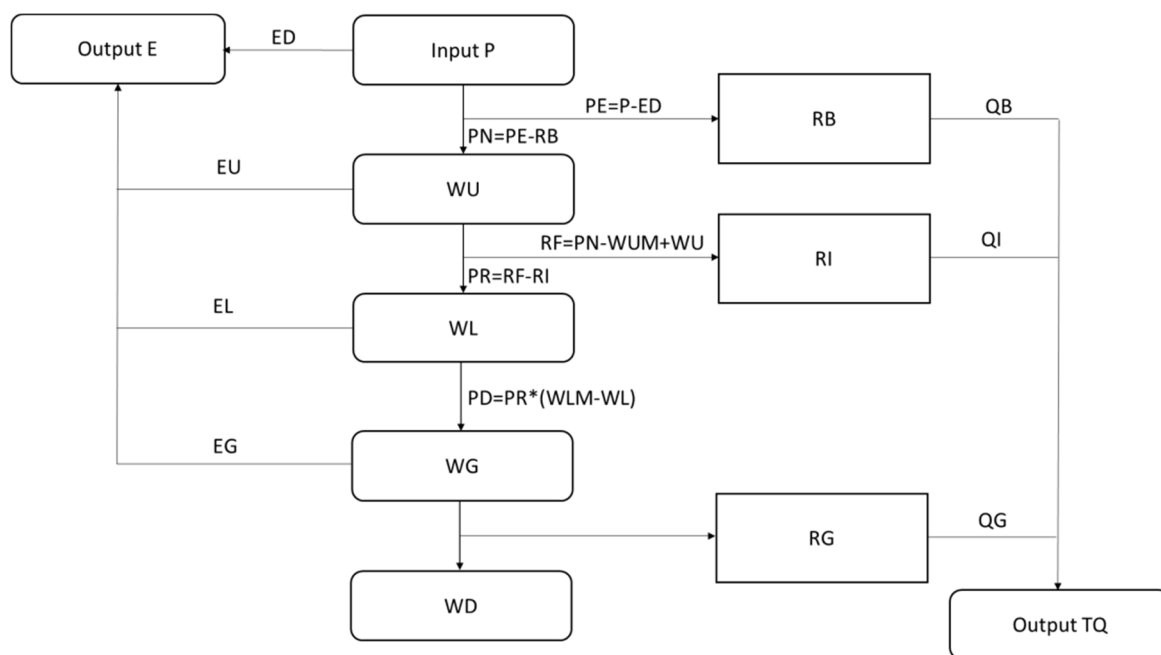


Figure 2. Flowchart of the DHF hydrological model.

Table 1. Parameters description of the DHF hydrological model.

Parameters	Description	Parameters	Description
P	precipitation	RB	surface runoff
ED	pan evaporation	RI	Interflow runoff
PE	effective precipitation	RG	groundwater runoff
PN	infiltrative precipitation	RF	infiltration
WU	surface upper layer water storage	PR	infiltration after detaining interflow runoff
WUM	storage capacity of upper layer	WG	groundwater storage
WL	surface lower layer water storage	WD	deep groundwater storage
WLM	storage capacity of lower layer	PD	lower infiltration
EU	upper layer evaporation	EL	lower layer evaporation
EG	groundwater evaporation	QB	surface streamflow
QI	interflow streamflow	QG	groundwater streamflow

Table 2. Number of floods in specified year for DHF reservoir.

Year	Number of Floods
1964	5
1985	4
1995	3
1996	3

The accuracy uses the criteria of qualified rate, and the correlation between the forecast errors are shown in Table 3. Table 3 shows that the DHF hydrological model has good performance in reproducing volumes and peak floods in the DHF river basin because the DHF model is especially developed for this river basin. It is also seen from the Table 3 that the errors of rainfall forecast and flood forecast are statistically significant correlated, where the qualified rate is calculated by counting the errors which is within the range of 20% observed rainfall or flood samples, and dividing the total number of respective

error samples. The correlation coefficients between two forecast errors are calculated according to statistics theory. Comparing different correlation coefficients between rainfall forecast error and flood forecast error, it seems statistically more nonlinear than linear association for these two forecast errors because the former have larger coefficients, but for the convenient of calculation, only the linear correlation is considered in this study.

Table 3. Accuracy of floods and correlation of forecast errors for DHF reservoir.

Items	Qualified Rate (%)	Correlation of Forecast Errors		
		Pearson	Kendall	Spearman
Rainfall forecast	93.55	0.40 *	0.50 **	0.68 **
Flood forecast	80.65			

Note: * The correlation is significant at the level 0.05; ** the correlation is significant at the level 0.01.

To determine the distribution of forecast errors for DHF hydrological model, a normal statistical test namely the Lilliefors test [43,44] has been conducted and the result is shown in Figure 3. It is observed that the rainfall forecast errors and flood forecast errors are almost normally distributed, although there are some outliers on the upper tail of the error samples. The Lilliefors test on the distribution of the error samples shows that we could not reject the null hypothesis that these errors come from a normal family at the 5% significant level. This result is consistent with the conventional hypothesis that the errors follow a normal distribution, but it is very obvious that the distributions of errors samples do not come from a standard normal distribution because they deviate the origin of the axis. Thus, the probability density functions (pdfs) of error samples are obtained with the rainfall forecast errors $e_r \sim N(2.09, 9.12)$ and the flood forecast error $e_f \sim N(-3.13, 294.11)$ through normal moments fitting method.

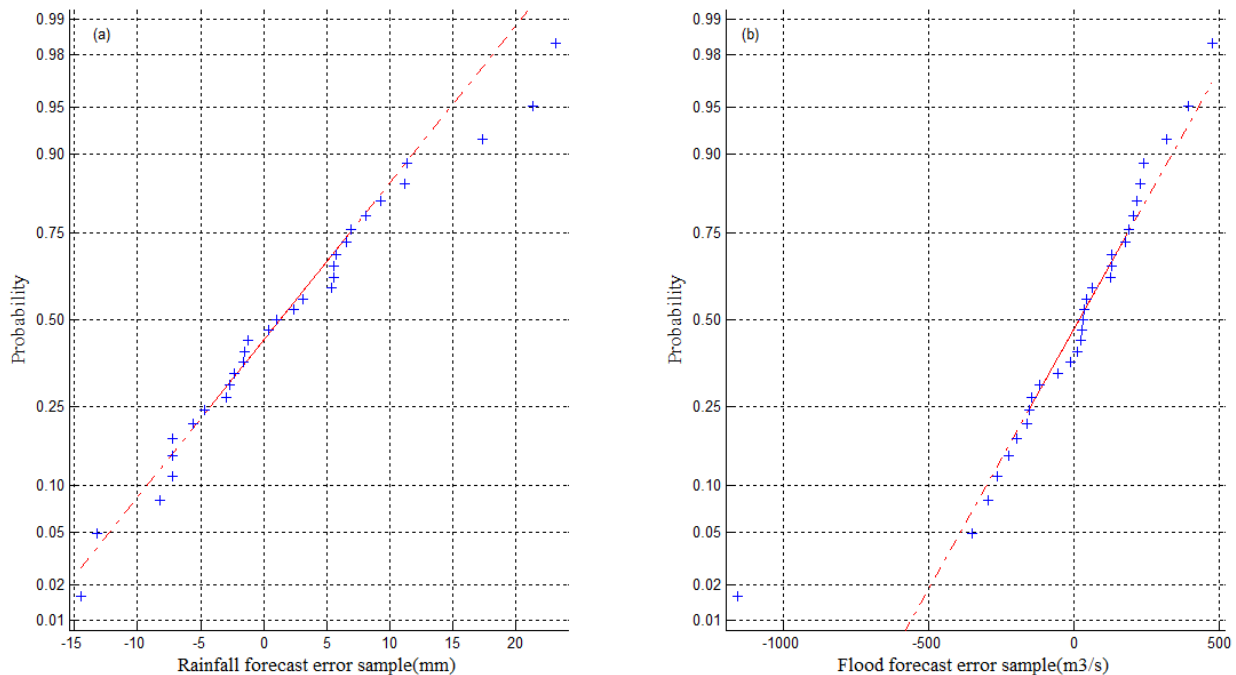


Figure 3. Normal distribution test for (a) rainfall forecast error sample; and (b) flood forecast error sample.

To identify the relationship between the forecasted items such as rainfall or flood peaks and the forecast error, Figure 4 is obtained through the standardized pairs of the forecasted rainfall and the rainfall forecast error, and pairs of the forecasted flood peak and the flood forecast error. The standardized forecast or forecasted samples are conducted by the forecast or forecasted samples minus their respective mean, and then are divided by their respective standard errors. Figure 4 shows that the two pairs of rainfall and peak flood seems to be stochastic with the two correlation coefficients of two pairs being both less than 0.2, thus it could be regarded as independent pairs.

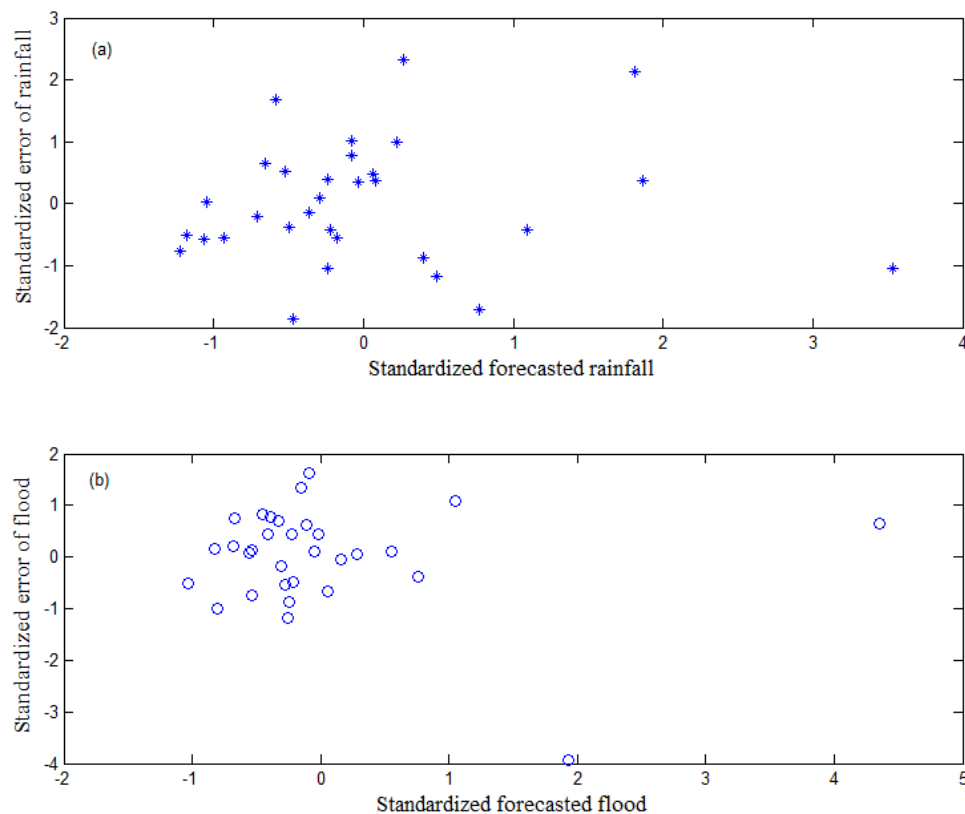


Figure 4. Relationship of (a) standardized forecasted rainfall and standardized error of rainfall; and (b) standardized forecasted flood and standardized error of flood.

4. Results and Discussion

The following sections discussed two modes of reliability indexes that considered the acceptable error threshold, especially the single and double failure modes. The single failure mode deals with the rainfall forecast error or flood forecast error separately, while the double mode deals with the failure of jointly rainfall forecast error and flood forecast error.

4.1. Single Failure Mode

To assess the performance of a hydrological model, 20% of the standard deviations of observed samples in rainfall forecasts and flood forecasts are regarded as their acceptable error threshold (forecast error resistance), according to the Standard for hydrological information and hydrological forecasting of China [45]. The reliability indexes β calculated by Equation (3) and the failure probabilities p_f calculated by Equation (4) of the two errors are shown in Figure 5. We note that since the performance

functions are normal and the respective forecast error resistance is constant, the calculation of the β and p_f is therefore easy in this case. The variations in failure probabilities and reliability indexes, using different proportions (e.g., 5%, 10%, 15%, 20%, 25%, 30% and 35%) of the standard deviations of observed samples as the acceptable error threshold, are also shown in Figure 5, where the solid lines are the rainfall forecast error, the dashed lines are the flood forecast error. The horizontal axis represents different proportions of the standard deviations of observed rainfall or peak floods, and the left and right vertical axis represent the failure probability and the reliability index, respectively.

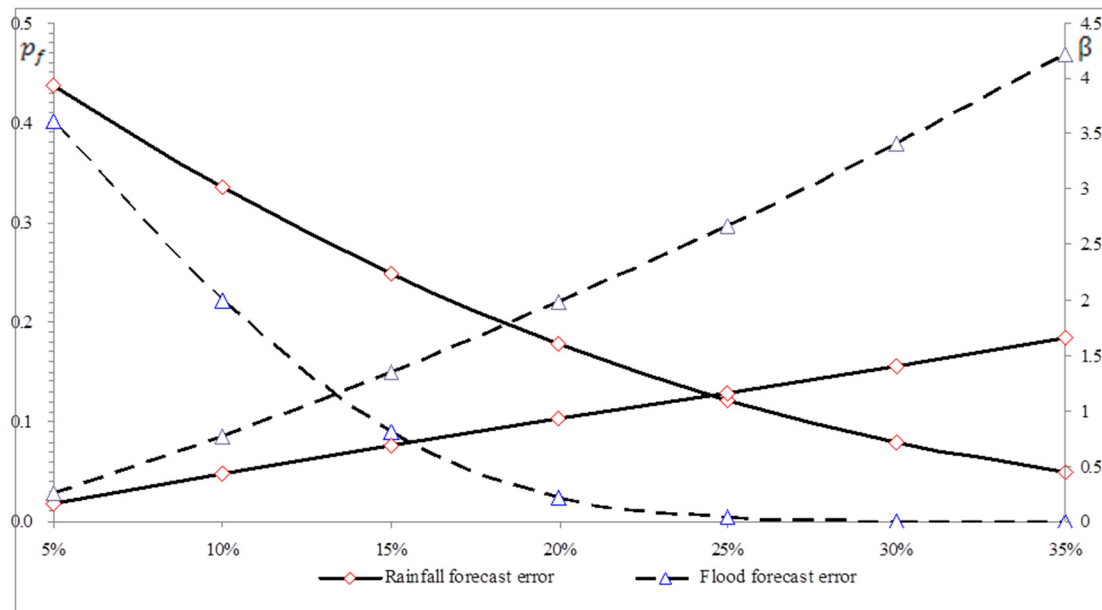


Figure 5. Failure probabilities and reliability indexes of rainfall forecast errors and flood forecast errors for DHF reservoir.

Figure 5 shows that when the acceptable error threshold increases, the failure probabilities of the two errors decrease, and the reliability indexes increase. This point justifies that an increasing acceptable error threshold is equivalent to increasing resistances of error. The failure probability of flood forecast error decreases nonlinearly with respect to the acceptable error threshold up to 35% of the standard deviations of the observed flood samples. Obviously, different acceptable thresholds have different failure probabilities of flood forecast error, and the curve shows more flat after the 20% threshold, indicating that the failure probability of flood forecast error may not be sensitive when the threshold reaches to certain proportion. At the same time, the reliability index of flood forecast error changes nonlinearly with different acceptable thresholds. However, the failure probability and reliability index of rainfall forecast error changes almost linearly with different threshold of rainfall forecast error. As a result, the reliability indexes of these two forecast error are 0.94 for rainfall forecast error and 1.38 for flood forecast error respectively, corresponding failure probability are 0.17 for the former and 0.085 for the latter. It shows that the flood forecast errors is more reliable than the rainfall forecast errors in DHF hydrological model, because the former has larger reliability index, since the reliability index is a relevant index with the inclusion of the variance of the observed samples.

4.2. Double Failure Mode

4.2.1. Unknown Performance Function

The reliability index of the DHF hydrologic model can be calculated by combining the rainfall and flood forecast errors. Since the physical mechanism of the two errors affecting the hydrologic model is unclear, an unknown performance function is first considered. In reliability theory, the wide and narrow bounds methods [38–40] were used to calculate reliability indexes of serial and parallel structural systems. Here, the serial or parallel structured error represent the relationship to the reliability of the hydrological model, where the forecasted rainfall and flood are both used to calculate the forecast error. Taking into account the correlation between the two failure modes (here simply using the linear correlation coefficient between the two error samples as coefficient for the two failure modes for the convenient of calculation), the results are shown in Table 4.

Table 4. System failure probability of rainfall forecast errors and flood forecast errors for DHF reservoir (with bounds method).

Structure Types	Correlation Coefficient	Wide Bounds Method	Narrow Bounds Method
Serial	0.40	[0.17, 0.24]	[0.22, 0.22]
Parallel		[0.015, 0.085]	[0.017, 0.035]

Table 4 indicates that the two errors have larger failure probability in serial structure than in parallel structure. This is due to the mutual effects of the two errors in a serial structure, where the failure of any error will result in failure of the whole system. In the parallel structure, on the other hand, the failure of one error does not imply the failure of the system since it only fails when both errors fail. Table 4 shows the system failure probability is increased comparing to the single failure mode of rainfall forecast error or flood forecast error with the serial structure type for these two forecast errors, however, the system failure probability with the parallel structure type is decreased comparing to the single mode, which is because the structure type of forecast error changes the interaction of forecast errors on the failure probability of combined system. That is to say, it is easier to failure with serial structure because of the additional load, while more difficult to failure with parallel structure because of the additional resistance. Table 4 also shows that the interval of the wide bounds method contains the interval of the narrow bounds method. Hence, in practice, the narrow bounds method is generally used to avoiding the overly wide range gathered from the wide bounds method and the result obtained by the narrow bounds may be regarded more accurate because of its narrow variance from the perspective of hydrology.

4.2.2. Known Performance Function

Five types of performance functions (Equation (12)) are considered in this study.

$$Z_a = e_R - (e_r + e_f) \quad (12a)$$

$$Z_b = e_R - (e_r - e_f) \quad (12b)$$

$$Z_c = e_R - (-e_r + e_f) \quad (12c)$$

$$Z_d = e_R - e_r \times e_f \tag{12d}$$

$$Z_e = e_R - e_f / e_r \tag{12e}$$

where Z_a, Z_b, Z_c, Z_d, Z_e denote different performance functions. e_R is systemic resistance under combined effect of rainfall forecast error and flood forecast error, e_r is rainfall forecast error load, and e_f is flood forecast error load.

The calculation of the failure probability of the system was performed by Monte-Carlo sampling methods [27,28]. In the calculation, e_R is constant and obtained by the production of the 20% standard deviation of respective observed samples. Figure 6 shows the variation of the reliability indexes with the Pearson coefficient and the system resistance.

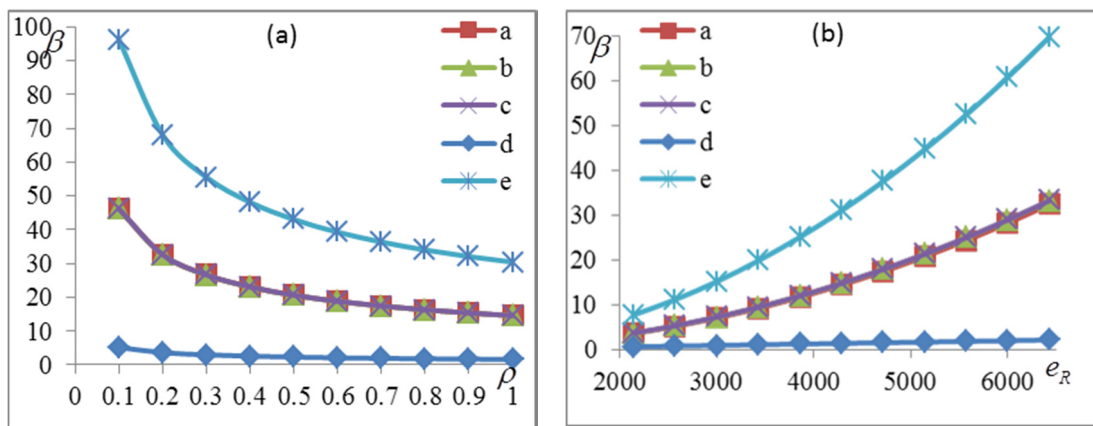


Figure 6. Relationship between system reliability indexes and (a) different correlation coefficients and (b) different system resistances of two errors of DHF reservoir.

Figure 6a shows that the system reliability indexes varies with different correlation coefficients, and the system reliability indexes decrease with the increasing positive correlation coefficients between two error samples for all five types performance function. Due to the potential relationship between the system reliability indexes and the correlation of two forecast errors [46], the derivative of system reliability index with respect to the correlation may be negative. Thus, the system reliability index is a decrease function of correlation coefficient when it varies in the interval (0, 1]. Different forms of performance function have great impact on the system reliability indices. Form Equation (12d) has the smallest β at the same Pearson coefficient, while form Equation (12e) has the largest. Forms Equations (12a)–(12c) have almost the same change curves with different coefficients, which may due to the reason that the values of loads of Equations (12a)–(12c) are far small than the resistances. It is also clear that the resistance e_R has great impact on system reliability indices from the Figure 6b, since the system reliability indices are increased with the resistance for the two error samples at the same resistance. Form of Equation (12e) shows the largest system reliability index, while form Equation (12d) shows the smallest one. Forms of Equations (12a)–(12c) behave very similarly. The form of Equation (12d) could be chosen as the form of performance function according to the principle of worst. As a result, the reliability index of system was 1.39 with the performance function form Equation (12d) under the correlation coefficient 0.40.

4.3. Discussion

4.3.1. The Acceptable Threshold Value of the Performance Function

The performance assessment of a hydrological model may be easy problem by comparing the simulated and observed outputs. Based on this logic, many performance assessment approaches are presented, such as the NSE [1], RMSE and the R or R^2 . Most of them are based on the residuals of the model for the simulated hydrological processes [47,48]. The residuals could have acceptable threshold, which would be resistance from the perspective of reliability theory. Therefore, based on the reliability theory, the assessment of performance of hydrological model is conducted to (1) considering the acceptable forecast error threshold and (2) including the correlation among the forecast errors, which will add a new insight to the forecast error and for the assessment of a hydrological model, and the traditional criteria are still very useful from different aspects to quantify the performance of a hydrological model.

In this study, the acceptable threshold of forecast error is only determined by the observed samples of rainfall or flood, it is an empirical and conventional approach because of the scarcity of physical foundation on allowance error limit. The 20% of observed sample variance is adopted according to the legally bind standard (GAQSIQC), thus as a guideline, it is also recommended to use this acceptable threshold in practice based on the standard, however, it requires further works including the sensitivity analysis of performance function of a hydrological model to determine the critical value because the resistance directly impacts the evaluation of performance of hydrological models, and many factors may influence the acceptable threshold of forecast error, not only the precision of the hydrological model, which will involve the uncertainties in hydrological system, but also the recognition limit of the human on forecast error itself which remain unexplored.

In the form of the performance function, only linear function between the load and the resistance is considered, which is popular and simplest way in the literatures appeared. Other forms could be considered after analyzing the relevant features between the load and the resistance. In addition, the nonlinear correlation between the basic variables such as rainfall forecast error and flood forecast error need to explore in deep research in the future.

4.3.2. Correlation between the Basic Variables of the Performance Function

The issue of correlation among the basic variables of the performance function has been widely discussed in the structure engineering, and many approaches have been developed to deal with this issue in the reliability theory, which presents a convenient tool for correlated variables because of many mature approaches such as orthogonal transformation [34] and Rosenblatt transformation [35], since many performance function in practice are complex, and it is difficult to obtain the analytical solution for the performance function, thus many approximation methods have been proposed, including the FOSM, SOSM and SOFM, and so on.

The correlation between the rainfall forecast error and flood forecast error is a challenging work since few literature has discussed the physical mechanism of the correlation [17,18]. Results are reasonable because the correlation is statistically significant through the case study of DHF reservoir, but the correlation between the variables of the rainfall forecast error and flood forecast error may not the same

correlation between the modes of two failure modes, and hence, only the linear correlation is considered for the two forecast errors in the convenient method. However, whether the rainfall forecast error influence the flood forecast error or not requires further investigation.

5. Conclusions

By considering the forecast error may not cause risk instantly and the statistically significant correlation between the rainfall forecast error and the flood forecast error, an assessment framework based on the reliability theory on the performance of the hydrological model is presented. Compared with conventional criteria such as NSE [1] and RMSE, this framework used not only the residual information but also the threshold of the residuals. Although the determination of the threshold is empirical, the reliability index gives another insight for the evaluation of performance of hydrological models and it remains a difficult task to compare the reliability index to the common criteria like RMSE, NSE [1] and R^2 because the former add information on the threshold of forecast error. The correlation between the two errors and two failure modes is also considered through the reliability theory. Although the physical mechanism of the correlation between the two errors is not clear in detail, it is necessary for dealing with the correlation when it is statistically significant. Future work could be focused on the error propagation from the rainfall forecast error to the flood forecast error, and using the reliability index to guide the practice for presenting criteria for evaluating the performance of hydrological models. Through the case study of DHF hydrological model, the following conclusion could be obtained:

- (1) The sensitivity of performance function on different acceptable threshold of forecast error are different, and the variation of the flood forecast error seems to change nonlinearly with their acceptable threshold and the rainfall forecast error changes almost linearly with their acceptable threshold.
- (2) The correlation between the rainfall forecast error and the flood forecast has great impact on the evaluation of performance of hydrological model, and the reliability index of the hydrological model system decreases when the correlation increases positively. The resistance has great impact on the evaluation of performance of a hydrological model, which means that larger resistance indicates higher reliability index.

Acknowledgments

This research is financially supported by the National Science Foundation of China (Grant No. 51439007, 51190094, 50909073), the National Basic Research Program of China (“973” Program, Grant No. 2013CB036406), the Collaborative Innovation Center for Major Ecological Security Issues of Jiangxi Province and Monitoring Implementation (Grant No. JXS-EW-00), and the Fundamental Research Funds for the Central Universities (Grant No. 2042014kf0292).

Author Contributions

Qianjin Dong carried out the calculation, result analysis and drafted the manuscript, which was revised by both authors. All authors read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Nash, J.E.; Sutcliffe, J.V. River flow forecasting through conceptual models—Part I: A discussion of principles. *J. Hydrol.* **1970**, *10*, 282–290.
2. Willems, P. Parsimonious rainfall–runoff model construction supported by time series processing and validation of hydrological extremes—Part 2: Intercomparison of models and calibration approaches. *J. Hydrol.* **2014**, *510*, 591–609.
3. Van Werkhoven, K.; Wagener, T.; Reed, P.; Tang, Y. Sensitivity-guided reduction of parametric dimensionality for multi-objective calibration of watershed models. *Adv. Water Resour.* **2009**, *32*, 1154–1169.
4. Vrugt, J.A.; Diks, C.G.H.; Gupta, H.V.; Bouten, W.; Verstraten, J.M. Improved treatment of uncertainty in hydrologic modelling: Combining the strengths of global optimization and data assimilation. *Water Resour. Res.* **2005**, *41*, doi:10.1029/2004WR003059.
5. Neumann, M.B.; Gujer, W. Underestimation of uncertainty in statistical regression of environmental models: Influence of model structure uncertainty. *Environ. Sci. Technol.* **2008**, *42*, 4037–4043.
6. Willems, P. A time series tool to support the multi-criteria performance evaluation of rainfall–runoff models. *Environ. Model. Softw.* **2009**, *24*, 311–321.
7. Sorooshian, S.; Dracup, J. Stochastic parameter estimation procedures for hydrologic rainfall–runoff models: Correlated and heteroscedastic error cases. *Water Resour. Res.* **1980**, *29*, 1185–1194.
8. Sorooshian, S. Parameter estimation of rainfall–runoff models with heteroscedastic streamflow errors—The noninformative data case. *J. Hydrol.* **1981**, *52*, 127–138.
9. Kelly, K.S.; Krzysztofowicz, R. A bivariate meta-Gaussian density for use in hydrology. *Stoch. Hydrol. Hydraul.* **1997**, *11*, 17–31.
10. Xu, C.Y. Statistical analysis of a conceptual water balance model, methodology and case study. *Water Resour. Manag.* **2001**, *15*, 75–92.
11. Montanari, A.; Brath, A. A stochastic approach for assessing the uncertainty of rainfall–runoff simulations. *Water Resour. Res.* **2004**, *40*, doi:10.1029/2003WR002540.
12. Mantovan, P.; Todini, E. Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology. *J. Hydrol.* **2006**, *330*, 368–381.
13. Kavetski, D.; Fenicia, F.; Clark, M.P. Impact of temporal data resolution on parameter inference and model identification in conceptual hydrological modelling: Insights from an experimental catchment. *Water Resour. Res.* **2011**, *47*, doi:10.1029/2010WR009525.
14. Evin, G.; Thyer, M.; Kavetski, D.; McInerney, D.; Kuczera, G. Comparison of joint *versus* postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. *Water Resour. Res.* **2014**, *50*, 2350–2375.
15. Krzysztofowicz, R. Integrator of uncertainties for probabilistic river stage forecasting: Precipitation-dependent model. *J. Hydrol.* **2001**, *249*, 69–85.

16. Hostache, R.; Matgen, P.; Montanari, A.; Montanari, M.; Hoffmann, L.; Pfister, L. Propagation of uncertainties in coupled hydro-meteorological forecasting systems: A stochastic approach for the assessment of the total predictive uncertainty. *Atmos. Res.* **2011**, *100*, 263–274.
17. Nester, T.; Komma, J.; Viglione, A.; Blöschl, G. Flood forecast errors and ensemble spread—A case study. *Water Resour. Res.* **2012**, *48*, doi:10.1029/2011WR011649.
18. Yazdi, J.; Salehi-Neyshabouri, S.A.A.; Golian, S. A stochastic framework to assess the performance of flood warning systems based on rainfall–runoff modeling. *Hydrol. Process.* **2014**, *28*, 4718–4731.
19. Dogulu, N.; López, P.; Solomatine, D.P.; Weerts, A.H.; Shrestha, D.L. Estimation of predictive hydrologic uncertainty using the quantile regression and UNEEC methods and their comparison on contrasting catchments. *Hydrol. Earth Syst. Sci.* **2015**, *19*, 3181–3201.
20. Han, S.W.; Wen, Y.K. Method of reliability-based seismic design. 1: Equivalent nonlinear systems, *J. Struct. Eng. ASCE* **1997**, *123*, 256–263.
21. Wen, Y.K. Reliability and performance-based design. *Struct. Saf.* **2001**, *23*, 407–428.
22. Mohammadzadeh, S.; Ahadi, S.; Nouri, M. Stress-based fatigue reliability analysis of the rail fastening spring clip under traffic loads. *Laitin Am. J. Solids Struct.* **2014**, *11*, 993–1011.
23. Rakoczy, A.M.; Nowak, A.S. Resistance factors for lightweight concrete members. *ACI Struct. J.* **2014**, *111*, 103–111.
24. Silva, J.E.; Garbatov, Y.; Soares, G.C. Reliability assessment of a steel plate subjected to distributed and localized corrosion wastage. *Eng. Struct.* **2014**, *59*, 13–20.
25. Loucks, D.P.; Stedinger, J.R.; Haith, D.A. *Water Resources System Planning and Analysis*; Prentice-Hall: Upper Saddle River, NJ, USA, 1988.
26. Ganji, A.; Jowkarshorijeh, L. Advance first order second moment (AFOSM) method for single reservoir operation reliability analysis: A case study. *Stoch. Environ. Res. Risk Assess.* **2012**, *26*, 33–42.
27. Augusti, G.; Baratta, A.; Casciati, F. *Probabilistic Methods in Structural Engineering*; Chapman and Hall: London, UK, 1984.
28. Rubinstein, R.Y.; Kroese, D.P. *Simulation and Monte-Carlo Method*; Wiley: New York, NY, USA, 2007.
29. Hasofer, A.M.; Lind, N.C. Exact and invariant second-moment code format. *J. Eng. Mech. Div. AMSE* **1974**, *100*, 111–121.
30. Rackwitz, R.; Fiessler, B. Structural reliability under combined random load sequences. *Comput. Struct.* **1978**, *9*, 489–494.
31. Xu, L.; Cheng, G.D. Discussion on: Moment methods for structural reliability. *Struct. Saf.* **2003**, *25*, 192–199.
32. Faravelli, L. Response surface approach for reliability analysis. *J. Struct. Mech. ASCE* **1989**, *115*, 2763–2781.
33. Maskey, S.; Guinot, V. Improved first-order second moment method for uncertainty estimation in flood forecasting. *Hydrol. Sci. J.* **2003**, *48*, 183–196.
34. Shinozuka, M. Basic analysis of structural safety. *J. Struct. Div. ASCE* **1983**, *109*, 721–740.
35. Rosenblatt, M. Remark on a multivariate transformation. *Ann. Math. Stat.* **1952**, *23*, 470–472.
36. Liu, P.L.; der Kiureghian, A. Multivariate distribution models with prescribed marginal and covariances. *Probab. Eng. Mech.* **1986**, *1*, 105–112.

37. Schoutens, W. *Stochastic Processes and Orthogonal Polynomial*; Springer-Verlag: New York, NY, USA, 2000.
38. Ditlevsen, O. Narrow reliability bounds for structural systems. *J. Struct. Mech.* **1979**, *7*, 453–472.
39. Cornell, C.A. Bounds on the reliability of structural systems. *J. Struct. Div. ASCE* **1967**, *93*, 171–200.
40. Kounias, E.G. Bounds for the probability of a union with applications. *Ann. Math. Stat.* **1968**, *39*, 2154–2158.
41. Horton, R.E. The role of infiltration in the hydrologic cycle. *Trans. Am. Geophys. Union* **1933**, *14*, 446–460.
42. Zhai, S.Y. Risk Analysis of Flood Forecasting and Flood Regulation under Different Limited Level. Master's Thesis, Nanjing Hydraulic Research Institute, Nanjing, China, 2010.
43. Lilliefors, H. On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *J. Am. Stat. Assoc.* **1967**, *62*, 399–402.
44. Lilliefors, H. On the Kolmogorov-Smirnov test for the exponential distribution with mean unknown. *J. Am. Stat. Assoc.* **1969**, *64*, 387–389.
45. General Administration of Quality Supervision, Inspection and Quarantine of P.R.C. (GAQSIQC). *Standard for Hydrological Information and Hydrological Forecasting*; (GB/T 22482-2008); China Water & Power Press: Beijing, China, 2008.
46. Hohenbichler, M.; Rackwitz, R. First-order concepts in system reliability. *Struct. Saf.* **1983**, *1*, 177–188.
47. Kamruzzaman, M.; Shahriar, M.S.; Beecham, S. Assessment of Short Term Rainfall and Stream Flows in South Australia. *Water* **2014**, *6*, 3528–3544.
48. Callegari, M.; Mazzoli, P.; de Gregorio, L.; Notarnicola, C.; Pasolli, L.; Petitta, M.; Pistocchi, A. Seasonal River Discharge Forecasting Using Support Vector Regression: A Case Study in the Italian Alps. *Water* **2015**, *7*, 2494–2515.

© 2015 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 license (<http://creativecommons.org/licenses/by/4.0/>).