The Corrected Calculation for the Sediment Discharge Based on the Copula Function

WANG Xue [1, a] and LI Yu [2]

[1] Integrated Management Center of Changjiang Water Resources Commission, Wuhan, China

[2] Hydrology Bureau of Changjiang Water Resources Commission, Wuhan, China

Key Words: sediment discharge; corrected calculation; Copula Function

Abstract. Some problems including complex calculation methods and parameter calibration difficulties exist in the current corrected calculation of the sediment discharge. A new method based on the copula function was proposed in this paper to solve them. The results show that the proposed method can improve the precision of calculation without adding any task or difficulty. The method was put forward based on the statistic characteristics of the observed sample and was calculated only by the statistical methods such as copula function and conditional expectation. It can provide a new computing approach and idea for the corrected calculation of the sediment discharge.

1 Introduction

For a long time, the channel cut-and-fill amount which calculated by sediment discharge method was not consistent with that calculated by the volume method[1]. Researches showed that the key of the problem is that we couldn't obtain the concentration of suspended sediment near the bottom layer of bed accurately [2-5]. According to "Code for measurement of suspended sediment in open channels" [6], the measurement range of the suspended sediment is always from the water surface to the 0.8 times depth. The five-point samples method is more accurate, but limited to the instrument structures, the sediment weight which the distance to the river bed less than 0.5 meters, or the concentration was higher, or the particle size was coarser can't be measured. For these reasons, the sediment transport rate has large difference between calculated value and real value, so that the precision of annual sediment discharge is also affected.

In order to modify the existing measurement results, many researchers had carried out relevant research [3-5]. A new method based on the copula function was proposed in this paper. Firstly, marginal distribution function for the mean sediment concentration of five-point and seven-point samples was given by the optimization program. Secondly, joint distribution function was constructed by the Copula function. Finally, the conditional probability distribution for the mean sediment concentration of seven-point samples could be obtained by the formula of the conditional probability once the mean sediment concentration of five-point samples was given. And the correcting value of the final sediment concentration was the expected value of the conditional probability distribution. The results of measurement of suspended sediment in Yichang hydrological station were selected as a case study to test the feasibility and practicality of the method.

2 The Corrected Calculation for the Sediment Discharge Based on the Copula Function

2.1 Sediment Distribution Function Optimization

There is no unanimous final conclusion yet about the distribution form and the distribution function of the mean sediment concentration. 12 types of statistical distribution models which are widely used in the field of hydrology, meteorology and finance are selected in this paper to fit the distribution of the mean sediment concentration [7-8]. The distribution functions of different types are showed in table 1.

Table 1. Distribution functions of different types.

Model Type	Function Name (Abbreviation)
Bounded distribution model	Beta distribution(BET), Power function distribution(POW)
No boundary distribution model	Gumber distribution(GUM), Normal distribution(NOR)
Non-negative distribution model	Exponential distribution(EXP), Gamma distribution(GAM), Lognormal distribution(GNO), Weibull distribution(WBL), P-III distribution (PIII)
Singular distribution model	Generalized extreme value distribution(GEV), Generalized Pareto distribution(GPA), Generalized logical distribution(GLO)

Linear moment method is used to estimate the parameters. K-S method is used to test the deviation degree between the theoretical distribution and the empirical distribution. An optimal distribution function is selected by the criterion of sum of squares of deviations (RMSE Criterion).

$$RMSE = \sqrt{\sum_{i=1}^{n} (p_{ei} - p_i)^2 / n} \tag{1}$$

Where P_i is the theoretical frequency; P_{ei} is the empirical frequency; and n is the number of the sequences.

2.2 Conditional Expectation Based on the Copula Function

The Copula function is an effective method of building a joint distribution. At present, the Copula function has been widely used in multivariate analysis and calculation [9]. Due to the positive correlation of the mean sediment concentration between the five-point and the seven-point samples, the Gumbel-Hougaard Copula function (G-H Copula function) in the Archimedean Copula function family is selected as the joint distribution function. The fitting result of the Copula function is assessed by RMSE criterion. The mathematical expression of the G-H Copula function is:

$$C_\theta(u,v) = \exp\left\{-[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta}\right\}, \quad \theta > 1 \tag{2}$$

Where θ is the parameter of the Copula function, which can be obtained by the relationship to the correlation coefficient τ of Kendall rank.

The expected value for the mean sediment concentration of the seven-point samples can be obtained according to the conditional expectation equation $E(y|x)$ when the mean sediment concentration of the five-point samples is known. The calculation equation of $E(y|x)$ as follows:

$$E(y|x) = \int_{-\infty}^{+\infty} y F'_{Y|X}(y|x)dy = \int_{-\infty}^{+\infty} yc(u,v)f_Y(y)dy = \int_0^1 c(u,v)F_Y^{-1}(v)dv \tag{3}$$

The numerical solution for the equation was calculated by the composite Simpson equation because the analytical solution cannot be obtained directly.

2.3 Calculation Steps

The conditional expected value for the mean sediment concentration of the seven-point samples can be obtained according to the conditional probability equation and the conditional expectation equation based on the Copula function when the mean sediment concentration of the five-point samples is known. Then this value can be used as the final correcting value to achieve the purpose of correcting the mean sediment concentration. The calculation steps as follows:

(1) The mean sediment concentration is calculated firstly. Weighting the point velocity V_η at the same relative depth η in the same vertical of monitoring points I by the area Ω, which is $V_\eta = \sum_i K_i V_{\eta-i} = \sum_i \frac{\Omega_i}{\Omega} V_{\eta-i}$. The sediment concentration CS_η is obtained by weighting the area and velocity, which is $CS_\eta = \sum_i K_i' CS_{\eta-i} = \sum_i \frac{\Omega_i}{\Omega} \frac{V_{\eta-i}}{V_\eta} CS_{\eta-i}$. Then the mean sediment concentration of the five-point samples CS_5 and the seven-point samples CS_7 are calculated by the weights of the depth.

(2) Optimizing the marginal distribution and calculating the parameters of the joint distribution function. The marginal distribution functions $F_{CS_5}(\bullet)$ and $F_{CS_7}(\bullet)$ can be selected from the distribution functions to be optimized according to the calculated series of *CS5* and *CS7* and the goodness evaluation index. Meanwhile, the parameter θ of the joint distribution $C_\theta(u,v)$ needs to be determined.

(3) Discrete interval. The value range [*CS7max*, *CS7min*] is divided into *n* equal division according to maximum value *CS7max* and minimum value *CS7min* of *CS7* data series. And each division is denoted by $h = (CS_{7\max} - CS_{7\min})/n$. Then the middle point of the subinterval [*CS7(k)*, *CS7(k+1)*] is denoted by $x_{k+1/2} = x_k + h/2$.

(4) Calculating the integral function. $u = F_{CS_5}(x_{CS_5})$ and $v = F_{CS_7}(x_{CS_7})$, as well as $F_Y^{-1}(v)$ which is the inverse function of *y* and the Copula function $c(u,v)$ are obtained by given the mean sediment concentration for five-point x_{CS_5} and seven-point samples x_{CS_7}, and the marginal distributions $F_{CS_5}(\bullet)$ and $F_{CS_7}(\bullet)$.

(5) The final correcting value is calculated by Simpson equation. For the range of *i-th*, the formula $f(v_i) = c(u,v_i)F_Y^{-1}(v_i)$ can be obtained. Then, the expected value for the mean sediment concentration of the seven-point samples $E(y|x)$ can be obtained according to the composite Simpson equation and the numerical integration method when the mean sediment concentration of the five-point samples is known. The calculation equation as follows:

$$E(y|x) = \int_0^1 c(u,v)F_Y^{-1}(v)dv = \frac{1}{6}[f(v_0) + 4\sum_{i=0}^{n-1} f(v_{i+1/2}) + 2\sum_{i=0}^{n-1} f(v_i) + f(v_n)] \quad (4)$$

3 Case Study

3.1. Data

The difficulty and the workload of the precision measurement with multi-line and multi-point are great. The more close to the sample points of the bottom bed, the higher requirements for the operation normatively and precisely. It is hard for a great deal of observation for a long time, and it is difficult to collect the sample in the general stations either. In view of these, 65 samples nearby bottom bed between 1973 to 1977 years in Yichang hydrological station were used for analysis as sample data. The point integration method of 10 lines and 7 points was used as the sampling method of suspend sediment. The sampling positions of the point integration method are water surface, 0.2*h*, 0.6*h*, 0.8*h*, 0.9*h*, 0.5m above the river bed, 0.1m to 0.2m above the river bed. The first five measuring points are the five-point samples in this paper, and *h* is the depth of water.

3.2 Marginal distribution Optimization

After the distribution functions is verified by the K-S test method, the goodness of fit for the marginal distribution is tested by the RMSE indicator. The results are showed in table 2.

Table 2. The marginal distribution of the mean sediment concentration of five-point samples and seven-point samples.

Distribution Function	Five-point Samples (%)		Seven-point Samples (%)	
	RMSE	K-S	RMSE	K-S
BET	3.43	7.82	3.15	7.21
POW	2.98	6.49	3.06	6.84
GUM	3.9	8.89	3.7	8.02
NOR	7.72	14.95	7.46	14.21
EXP	2.59	5.95	2.71	6.11
GAM	2.07	5.04	1.89	5.05
GNO	1.98	4.27	1.98	3.88
WBL	3.23	5.67	2.85	4.62
PIII	1.72	3.61	1.60	4.09
GEV	3.08	7.17	3.15	6.65
GPA	1.96	4.18	1.76	3.84
GLO	2.93	6.39	3.03	5.77

Note: When n=65, $\alpha = 0.05$, the critical value of K-S Test is $1.36 / \sqrt{65} \times 100 = 16.87\%$

The detailed data shown in Table 2 reveal that, whether the mean sediment concentration sample sequences belong to the five-point or seven-point samples, the values of the statistics D_{KS} for the 12 kinds of distribution functions are less than the critical value 16.87% with the confidence level α=0.05. In other words, the 12 kinds of distribution functions all can describe the probability distribution of the sample sequences. In order to choose the function which fitting best from the given distribution functions, the RMSE criterion is adopted. The smaller the values of the RMSE, the higher the fitting degree of the distribution function to samples. PIII distribution is chosen as the marginal distribution of the mean sediment concentration for the five-point and seven-point samples in this paper for two reasons. First, the RSME values of GNO, GPA and PIII in the 12 kinds of distribution functions are all small. Second, the PIII distribution is the most popular distribution function used in hydrologic frequency analysis.

3.3 Rationality Analysis

Based on the joint distribution, the conditional probability accumulative frequency curves of the mean sediment concentration for the seven-point samples were drawn in Figure 1. Five cases of the mean sediment concentration for the five-point samples were considered which were P=10% (Exceptionally large sediment), P=25% (Large sediment), P=50% (Middle sediment), P=75% (Less sediment) and P=90% (Exceptionally less sediment).
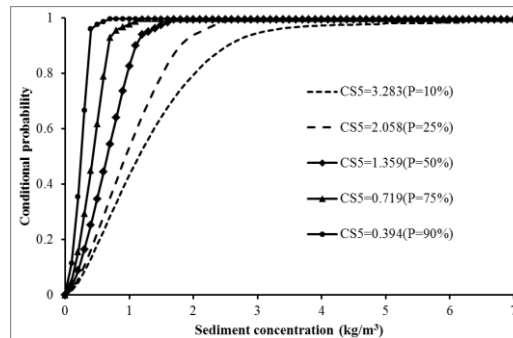

Figure 1. The conditional probability distribution of the mean sediment concentration of seven-point samples.

Figure 1 show that, with longitudinal comparison, when the mean sediment concentration of the five-point samples is a certain value, the corresponding value of the mean sediment concentration of the seven-point samples is not unique. The value appeared randomly in a certain rang and the probability for the appearance is different. The difference showed in the figure is the changes of the curvature of the accumulative frequency curves. The reason for the difference is the uncertainty of

the changes of the water and sediment conditions. With horizontal comparison, while the decrease of the mean sediment concentration of the five-point samples, the corresponding curvature and slope of the conditional probability accumulative frequency curves of the mean sediment concentration of the seven-point samples all change more. The slope of the curves is larger and the span tended to average state is smaller. The reason for the change is the good relativity between five-point and seven-point samples. The smaller the mean sediment concentration of the five-point samples, the more little probability to appear lager value of the mean sediment concentration of the seven-point samples. So the corresponding accumulative curves tend to smooth.

The mean sediment concentration of five-point samples was corrected by using the mean sediment concentration of the seven-point samples as real value, according to the corrected calculation based on the Copula function. The results of the corrected calculation were showed in Table 3.

Table 3. The result of corrected calculation of the sediment concentration based on Copula function.

Serial Number	Project	Sample Average	Characteristic Value				
			P=10%	P=25%	P=50%	P=75%	P=90%
①	Five-point method (kg/m³)	1.539	3.283	2.058	1.359	0.719	0.394
②	Seven-point method (kg/m³)	1.630	3.543	2.194	1.585	0.803	0.410
③	Corrected method (kg/m³)	1.624	3.510	2.182	1.429	0.747	0.406
Compare ① - ②	Absolute error (kg/m³)	0.092	0.260	0.136	0.226	0.084	0.016
	Relative error(%)	5.948	7.920	6.608	16.630	11.683	4.061
Compare ③ - ②	Absolute error (kg/m³)	0.006	0.033	0.012	0.156	0.056	0.004
	Relative error(%)	0.380	0.920	0.552	9.830	6.936	1.024

The data shown in Table 3 reveal that the measure errors after corrected calculation all reduced at different degree viewing from the sample characteristic value. Before the corrected calculation, the absolute error for the sample average of the mean sediment concentration between five-point and seven-point samples is 0.092kg/m³, and the relative error is 5.948%. After the corrected calculation, the absolute error for the sample average between the two methods reduces to 0.006kg/m³, and the relative error reduces to 0.38%. When P=10%, the decreasing range is the biggest with the relative error reduced from 7.92% to 0.92%. When P=90%, the decreasing range is the smallest with the relative error reduced from 4.061% to 1.024% and the decreasing range is over 3%.

4 Conclusions

Some problems including complex calculation methods and parameter calibration difficulties exist in the current correcting calculation of the sediment discharge. A new method based on the Copula function was proposed in this paper to solve them. The measurement data of suspended sediment in Yichang hydrological station was selected as a case study. Some conclusions obtained as follows:

(1) The water and suspend sediment conditions in Yichang station have well consistency, so the PIII distribution which widely applied in hydrologic frequency analysis is selected to fitting the sediment concentration samples. The results show that the goodness of fit is well. It is reasonable and credible to build the joint distribution for the mean sediment concentration of the five-point and seven-point samples based on G-H Copula function.

(2) The results of the sediment discharge calculated by five-point samples are smaller, compared to that of the seven-point samples. However, the difference between values for the two samples is not big with the correction coefficient between [1.052, 1.061]. Under the present observed conditions and technical conditions, the five-point samples still can meet the requirements of the Chinese code for suspended sediment measurement.

(3) The proposed method in this paper can improve the precision of calculation without adding any task or difficulty. The method based on the statistic characteristics of the observed sample is calculated only by the statistical methods such as Copula function, conditional expectation and numerical integration. It can provide a new approach and idea for the corrected calculation of the sediment discharge.

Acknowledgments

References

[1] DUAN Guang-lei, PENG Yan-bo, GUO Man-jiao. Comparative Analysis on Riverbed Erosion and Deposition Amount Calculated by Different Methods [J]. Journal of Yangtze River Scientific Research Institute, 2014, 31(2): 108-113. (In Chiness)

[2] HAN Qi-wei. Error Analysis of Coarse Suspended Load Measuring [J]. Journal of CHINA hydrology, 2008, 28(1): 1-6. (In Chiness)

[3] BU Ce. The correction method for annual sediment load of the measured suspended sediment [J]. Journal of CHINA hydrology, 1982, (6): 1-8. (In Chiness)

[4] WANG Gui-Dao. The preliminary analysis on the experiment of near bed suspended sediment [J]. Journal of CHINA hydrology, 1985, (1): 27-31. (In Chiness)

[5] YUAN Yuan, ZHANG Xiao-feng DUAN Guang-lei. Modifying the channel erosion and deposition amount of Yichang-Jianli reach calculated by sediment discharge method [J]. Journal of Hydroelectric Engineering, 2014, 33(4): 163-169. (In Chiness)

[6] GB50159-92. Code for measurement of suspended sediment in open channels[S]. (In Chiness)

[7] LIU Lu-liu, JIANG Tong, XU Jin-ge, et al. Research on the hydrological processes using Multi-GCMs and Multi-scenarios in the Xijiang River basin [J]. Journal of Hydraulic Engineering, 2012, 43(12): 1413-1421. (In Chiness)

[8] JIANG Tong, SU Bu-da, MARCO Gemmer. Trends in precipitation extremes over the Yangtze River basin [J]. Advances in Water Sciences, 2008, 19(5): 650-655. (In Chiness)

[9] LI Tian-yuan, GUO Sheng-lian, LIU Zhang-jun, et al. Design flood estimation methods for cascade reservoirs [J]. Journal of Hydraulic Engineering, 2014, 45(6): 641-648. (In Chiness)