A comparative study of the methods for estimating streamflow at ungauged sites

Y.M. Xiao^a, J. Xia^a, L. Lin^a and J.M. Wang^a

^a School of Water Resources and Hydroelectric Engineering, Wuhan University, Wuhan, China Email: <u>ymxiao@163.com</u>

Abstract: Streamflow records are most fundamental data for water resources planning, hydropower design and management, water supply, irrigation, flood control, and maintaining ecosystem health. However, gauged streamflow data usually are not available to conduct such applications at rural or remote areas. Although many methods have been developed to estimate streamflow at ungauged sites, such as hydrological drainage area ratio method and regional flow duration curves (FDC) method, accuracy of estimated streamflow is still poor. In this study, an improved FDC method was developed and applied to an ungauged basin in the central south of China. Results show that the improved FDC have better performances than other traditional methods in the study area. Meanwhile, the proposed method in this study is more reasonable and interpretable than other methods in predicting regional streamflow at ungauged sites. By considering three more smaller percentiles, i.e. 1%, 2% and 5%, further analysis shows that derived empirical FDC has a better performance in the ungauged sites in this study. It suggests that inclusion of smaller quantiles can improve the accuracy of estimated streamflow both in high or low flow regimes.

Keywords: Ungauged site, index gauge, flow duration curve, streamflow prediction

1. INTRODUCTION

In early 2003, the International Association of Hydrological Sciences (IAHS) launched a decade-long initiative of predictions in ungauged basins (PUB). The main purpose of this program is to study new hydrological theories and methods to solve the problems of lack of sufficient gauged streamflow data for better water resources management in developing countries (e.g., Yang et al., 2012; Salaudeen et al., 2016). Most of the hydrological stations in the world are located in Europe, America and Japan, while the stations are relatively sparse in less developed areas, especially in China where streamflow gauges in rural and remote areas are extremely sparse. Therefore, study on how to estimate streamflow at ungauged sites in China are very important.

Up to now, several methods have been proposed to estimate streamflow at ungauged sites. Among them, methods based on the flow duration curve (FDC) are the most popular and practical one, which was firstly proposed by Mimikou in 1985 (Mimikou et al., 1985). The FDC represents a percentage greater than or equal to one of the studied flow series and is a graphical visualization of a streamflow variability (e.g., Ssegane et al., 2012; Vogel et al., 1994 and 1995, and Nruthya et al., 2015). The shape of the FDC are influenced by hydrometeorological characteristics of the catchment and thus spatial variability of different percentile flows were attributed to catchment characteristics. Yang et al., (2012) used an area ratio method to predict FDC at ungauged basin in China, i.e. Mishui River, but the accuracy was low. It suggests that more studies have to be conducted to improve the capability of the FDC methods for estimating streamflow in the Mishui River basin.

In this study, a new method for calculating ensemble average errors in the FDC methods that was used by Yang et al. (2012) was developed to improve the practicability of in the humid region of China. The new method assume that the ensemble average errors are weighed in proportion to geographical proximity catchment characteristics of the sites in the Mishui River basin. Then the ensemble average errors were applied to modify the synthetic flow duration curves at the validation station. Hydrological, climatological, topographical attributes of catchments have been also considered for the analysis. Results show that model was most sensitive to elevation followed by drainage area, mean annual precipitation. Effectiveness of the presented methods was assessed. Conclusions based on the study are presented and discussed.

2. STUDY AREA AND DATA

2.1 Study area and data

The studied area is Mishui River, which is located in Hunan Province in the central of China and covers an area of approximately 9972 km². The Mishui River is a tributary of the Xiangjiang River. The Mishui catchment lies within humid subtropical climate regions with warm rainy seasons and cold dry seasons. Average annual surface air temperature in the region is 17.8° C. The maximum and minimum temperatures were 40.2 °C and -11.9° C. The elevation of the catchment ranges from 49m in the northwest to 2093 m in the southeast. The rainfall in the region is dominated by northeast monsoon from March to October. Annual rainfall varies from 1435 mm to 1700 mm per year. In the study area, there are 5 gauged stations with at least 5 complete years of daily runoff and rainfall observations from 1981-1985. The catchment is considered as a pristine catchment without any regulation (i.e. not significantly affected by regulation structures or water extraction). Figure 1 shows the distribution of gauging stations in the study area. The basic information of the five stations (Wulipai, Ganxi, Anren, Kenkou and Longjiashan station) are listed in the Table 1.

Daily evaporation data is only available at Wulipai station. The spatial and temporal distribution of the rainfall is uneven. On average, 58.5% of annual total runoff is in the period of March to June. While the dry period, i.e. July to September, only has about 20% of the annual runoff.

Q:4-	Area	Site 1	role	
Sile	$(km^2)_{\circ}$	Latitude Longitude		
Ganxi	9972.	113°00′	27°05′。	calibration
Longjiashan	4515.	113°27′	26°52′.	calibration
Anren	1815.	113°18′	26°41′.	calibration
Wulipai	814.	113°44′	26°29′.	calibration
Kenkou	692.	113°45′	26°41′.	validation

Table 1. Gauging sites in the study area



Figure 1. Distribution of Gauging stations across Mishui River basin

2.2 Geomorphologic characteristics

The key hydrologic and climatic characteristics of each gauged stations are listed in Table 2. Five catchment descriptors are used in this study including drainage area (F in km²), mean catchment elevation (H in m), catchment slope in degree (S in degree), mean annual rainfall (P in mm) and main stream length (L in km). Values of all the descriptors. The morphoclimatic attributes are obtained using ArcGIS. Drainage area reflects the runoff volume that can be generated from rainfall, watershed slope, and elevation contributes in the runoff momentum of a particular flood magnitude (Salaudeen et al., 2016).

Table 2. Summary of the key hydrologic and climatic characteristics of all the gauged stations

Station name	Record Length(years)	$F(km^2)$	L(m)	S(degree)	H(m)	P(mm)
Ganxi	5	9972	241	3.5127	417	1434.38
Longjiashan	5	4515	124	4.7425	578	1615.36
Anren	5	1815	59.6	2.732	382	1581.74
Wulipai	5	814	74.8	6.6975	892	1661.68
Kenkou	5	692	82	5.6585	760	1669.02

The empirical flow duration curves at each gauged station are derived from daily streamflow data in descending order by using the Weibull distribution to estimate the probability of exceedance. The empirical flow duration curves at each gauged station are established at intervals of 10%, i.e. drawing a curve with nine points from the 10%~90%. Following previous study by Yu et al. (2000), points with very small probability of exceedances (e.g. 1%) are used to characterize the big flow value and to get a better calibration. Therefore, two analyses are conducted in this study with two different groups of percentiles to characterize empirical FDC. The first group, i.e. group A, 9 exceedance percentages from 10% to 90% with an interval of 10% are chosen. The second group, i.e. group B, the same as group A but with three more exceedance percentages, i.e. 1%, 2%, and 5%. Differences in these two groups indicate whether it is necessary to include big flood events to improve the accuracy of estimated streamflow in the study area using the proposed new method comparing with previous one.

3. METHODOLOGY and RESULTS

In order to generate the flow duration curves in ungauged basins obtained from the linear regression, four of five gauged stations (Wulipai, Ganxi, Anren, and Longjiashan stations) are chosen for deriving the regional flow duration curve. The Kenkou station is left for validation. The FDC represent the integrated impacts of climate, geology, geomorphology, soils and vegetation on streamflow magnitude. In previous research, the regionalized flow duration curve was established with a set of regression equations relating the flows of 9 exceedance percentages (10%-90%) with the catchment area (F) as the only descriptor, i.e.,

$$Q_p = C + D \times F \tag{1}$$

where Q_p is the flow of p exceedance percentage, C and D are the constants, and F is the catchment area. Coefficients of determination, i.e. R², is lesser than 0.95 when p \geq 70%. Therefore, the simulation accuracy is poor in dry period.

3.1 Area ratio method

Area ratio method is traditionally used to estimate streamflow based on related information from gauged sites to ungauged sites. But the relationship between catchment area and Q_p is logarithmic. The regional model was obtained from the nonlinear regression as equation (2). The coefficients a and b are shown in the Table 3.

$\ln Q_{p} = a \times \ln F + b \qquad (2)$	Table 3. The coefficient of Eq2 of p			
Cross validation experiment was performed	exceedance percentage			
by assuming Kenkou site as an ungauged site.	p.	b.	a	
The chieving function is defined as the	0.01	-0.8131	1.024017.	
The objective function is defined as the	0.02.	-0.73365.	0.960464	
relative bias of FDCs $\triangle Q_p$:	0.05.	-0.94661	0.969170	
$\Delta Q_{\rm p} = \frac{Q_{\rm p} - q_{\rm p}}{q_{\rm p}} \tag{3}$	0.1	-1.02471	0.945667.	
where $\wedge \Omega_n$ is the flow of n exceedance	0.2	-1.08279	0.909015.	
percentage in the observed flow duration	0.3.	-1.14992	0.886883.	
curve $q_{\rm p}$ is the flow of p exceedance	0.4	-1.13251	0.849082	
percentage in the simulated flow	0.5	-1.13089	0.821204	
duration curve (Figure 2)	0.6	-1.12212	0.790903.	
	0.7.	-1.20694	0.786167	
	0.8	-1.34161	0.792596	
	0.9	-1.34571	0.759425	
40.00%				



Figure 2. The relative bias of FDCs

Figure 2 shows that performance of area ration method varies significantly at different sites, but the average fractional error of FDCs is small. It suggests that the flow duration curve method can be used in the study area for estimating streamflow at ungauged sites but it needs further calibration or improvement.

The synthetic flow duration curve was validated at the Kenkou stations as shown in Figure 3 and 4, in which the synthetic flow duration curves are very close to the actual flow duration curves, especially in B group. This allows us to compare the performance of selected model as a function of different duration intervals and to assess how the results vary in high-flows and low-flows (i.e. different streamflow regimes). Results shows analysis of group B are better than group A. It should be noted, however, that the method only fits well in the small and medium flow segments, bias at two points of $Q_{1\%}$ and $Q_{5\%}$ are relatively large.



Xiao et al., A comparative study of the methods for estimating streamflow at ungauged sites

3.2 Improved method of area ratio

In order to improve the consistency of synthetic flow duration curve at ungauged station, many methods are proposed. For regional FDC, we consider to fit the shape of FDC at different magnitudes. After fitting the shape, the regional FDC may still have some differences with empirical FDC, but it can better fit with the measured FDC by adjusting a magnification coefficient. The magnification coefficient is determined by a linear relationship of regional FDC's mean with the area of basin F. However, this is unable to improve the performance dramatically. Additionally, there are only 5 stations, 4 of which are for establishing linear relationship, therefore, the correlated points are significantly less than normal, and not suitable for sparse station network area.

In this study, a new modified method is proposed for the study area. With the help of widely available remote sensing data, basin area, average elevation, slope of basin, river length and other terrain data can be easily obtained from digital elevation model (DEM) database. Therefore, more descriptor of the shape of FDC are considered in this study including watershed topography as well as hydrological and meteorological conditions.

Data of four stations (Ganxi, Wulipai, Anren, Longjiashan) are used to establish the regional FDC for ungauged station Kengkou. The error of regional FDC is corrected by area ratio, height ratio, precipitation ratio, slope ratio, and area-precipitation ratio of four stations. The identified set of attributes and their corresponding results are shown in the Table 4.

This study found that drainage area (F), mean catchment elevation (H) and mean annual rainfall (P) are the three most important factors to predict different flow quantiles. Results obtained are in good agreement with observed data, especially good for the method modified by ratio of elevation, which shows significant improvement comparing with other method (e.g., Yang et al., 2012) (see Figure 5 and 6). Errors of different sites are derived from information on FDCs and flow records of similar hydrological sites in the study area. But similar sites play different roles as they have different catchment, topography and climatic characteristics. Errors in the synthetic flow duration curves were weighed in proportion to geographical proximity or catchment area of the different similar sites in study area to arrive at their ensemble average errors. Then the ensemble average errors were applied to modify the synthetic flow duration curves at the Kenkou station.

Estimated streamflow quantiles at the ungauged site using developed methods were compared with historical observed streamflow quantiles (see Table 4).

P observed		simulated	simulated modified simulated modified		simulated modified
1	i observed	sinnulated	by ratio of F	by ratio of H	by ratio of F and P
1%	144	121.38	125.58	146.36	125.99
2%	107	96.34	95.50	110.39	95.77
5%	64.9	62.44	59.95	65.49	60.05
10%	47	44.76	43.27	47.73	43.35
20%	34.4	30.84	30.41	36.39	30.52
30%	25.2	22.87	22.37	26.72	22.45
40%	20.7	18.61	17.98	22.30	18.06
50%	17.6	15.58	15.03	19.05	15.10
60%	15	13.05	12.66	16.24	12.72
70%	12.1	10.41	10.04	13.07	10.10
80%	9.7	7.96	7.63	10.22	7.68
90%	7.94	6.35	5.87	8.09	5.92

Table 4. The results of observed and simulated flows at different exceedance probability

The performance of different methods in predicting FDC for ungauged site was further investigated. Error in quantile estimation was determined by comparing estimated quantiles with observed corresponding quantile to the ungauged site, in terms of four performance measurements: Nash Sutcliffe Error (Nash), Relative-bias (R-bias), Root Mean Square Error (RMSE), and Coefficient of Determination(R²). The results are shown in Table 5. All the methods have good performance for the Kenkou site. Relatively, method considering elevation based weights is the best result comparing with other methods in the study.



Figure 6. The synthetic flow duration curves modified by different methods at the Kenkou station (B)

Table	5.	Performance	of the	developed	methods
1 ant	J •	1 ci iorinanec	or the	ueveloped	memous

Method	Nash	RMSE	R-bias(%)	R ²
simulated without modified	0.966895	0.125705	11.69387	0.995502
Simulated modified by ratio of F	0.972233	0.150311	14.13677	0.99879
simulated modified by ratio of H	0.998449	0.056213	-4.89209	0.99976
Simulated modified by ratio of F and P	0.973499	0.14653	13.77745	0.999035

4. DISCUSSION AND CONCLUSIONS

This study assessed different methods for the prediction of flow duration curve at an ungauged basin in the central south of China. An improved method based on the flow duration curve was developed and applied. Results shows that the proposed method is better than other models and is also more reasonable and interpretable in study area. By considering three more smaller percentiles, i.e. 1%, 2% and 5%, derived empirical FDC has a better performance in the ungauged sites in this study. It suggests that inclusion of smaller quantiles can improve the simulation results both in high or low flow regimes.

ACKNOWLEDGMENTS

The paper is funded by National Natural Science Foundation of China search projects (grant agreement number 51279141 and 41571028).

Xiao et al., A comparative study of the methods for estimating streamflow at ungauged sites

REFERENCES

- Mimikou, M., and Kaemaki, S. (1985). Regionalization of Flow Duration Characteristics. *Journal of Hydrology*,82:77-91.
- Nruthya K., and Srinivas, V. V. (2015). Evaluating methods to predict streamflow at ungauged sites using regional flow duration curves: a case study. *Aquatic Procedia*,4:641-648.
- Razaqa, S. A., Shahida, S. et al., (2016). Prediction of flow duration curve in ungauged catchments using genetic expression programming. *Procedia Engineering*, 154:1431-1438.
- Ssegane H., Tollner E.W., Mohamoud Y.M., et al. (2012). Advances in variable selection methods II: Effect of variable selection method on classification of hydrologically similar watersheds in three Mid-Atlantic ecoregions. *Journal of Hydrology*, 438-439: 26-38.
- Vogel, R., and Fennessey, N. (1994). Flow-duration curves. I: new interpretation and confidence intervals. Journal of Water Resources Planning & Management, 120 (4):485-504.
- Vogel, R., and Fennessey, N. (1995). Flow duration curves II: a review of applications in water resources planning. *Journal of the American Water Resources Association* 31(6): 1029–1039.
- Yang, B, Liu, Y., and Wang G. (2012). Improved regionalized flow duration curve based runoff simulation of ungauged basin. *Water Resources and Hydropower Engineering*, 43: 14-17. (*In Chinese*)
- Yu, P. S., and Yang, T. C. (2000). Using synthetic flow duration curves for rainfall-runoff model calibration at ungauged site. *Hydrological Process*, 14(6):117-133.