Entropy Parameter *M* in Modeling a Flow Duration Curve

Yu Zhang [1], Vijay P. Singh [1,2,*] and Aaron R. Byrd [3]

[1]   Department of Biological and Agricultural Engineering, Texas A&M University,
     College Station, TX 77840, USA; zhangyu199002@tamu.edu
[2]   Zachry Department of Civil Engineering, Texas A&M University, College Station, TX 77843-2117, USA
[3]   Hydrologic Systems Branch, Coastal and Hydraulics Laboratory, Engineer Research Development Center,
     U.S. Army Corps of Engineers, Vicksburg, MS 39181, USA; Aaron.R.Byrd@erdc.dren.mil
*   Correspondence: vsingh@tamu.edu; Tel.: +1-979-845-7028

Abstract: A flow duration curve (FDC) is widely used for predicting water supply, hydropower, environmental flow, sediment load, and pollutant load. Among different methods of constructing an FDC, the entropy-based method, developed recently, is appealing because of its several desirable characteristics, such as simplicity, flexibility, and statistical basis. This method contains a parameter, called entropy parameter *M*, which constitutes the basis for constructing the FDC. Since *M* is related to the ratio of the average streamflow to the maximum streamflow which, in turn, is related to the drainage area, it may be possible to determine *M* a priori and construct an FDC for ungauged basins. This paper, therefore, analyzed the characteristics of *M* in both space and time using streamflow data from 73 gauging stations in the Brazos River basin, Texas, USA. Results showed that the *M* values were impacted by reservoir operation and possibly climate change. The values were fluctuating, but relatively stable, after the operation of the reservoirs. Parameter *M* was found to change inversely with the ratio of average streamflow to the maximum streamflow. When there was an extreme event, there occurred a jump in the *M* value. Further, spatially, *M* had a larger value if the drainage area was small.

Keywords: flow duration curve; Shannon entropy; entropy parameter; modeling; spatial and dynamics characteristic

1. Introduction

A flow duration curve (FDC) is usually constructed empirically by plotting discharge against the percentage of time the discharge is equaled or exceeded during the year. Discharge from a gauge station can be daily, weekly, or monthly. The timescale of discharge depends on the use of FDC. For example, weekly discharge may be adequate for water supply, daily discharge for hydropower, and monthly discharge for sediment load and pollutant load [1,2]. Nonparametric methods use the record of discharge for the whole period for constructing an FDC and make no probabilistic statements about a given calendar or water year, because all the years of record are combined together into a whole period, so a return period cannot be assigned.

The methods for predicting an FDC are either deterministic or stochastic. For a given year of streamflow record at a station, an annual flow duration curve (AFDC) can be constructed [3,4]. With AFDCs of all the years at a given station, at each exceedance probability discharge percentiles can be determined given a return period. This leads to a final FDC with probabilistic statements by assigning return periods to individual AFDCs.

Singh [5] related dimensionless discharge with drainage area and constructed an exponential form of FDC using a deterministic model. Vogel and Fennessey [4] used an AFDC to define

recurrence intervals for FDCs. Cigizoglu and Bayazit [6] modeled FDCs by introducing stream flow as a product of two variables, which represented the periodic and stochastic components. Castellarin et al. [7] developed a five-parameter stochastic model which combined annual flow distribution and standardized the daily flow distribution of the basin to simulate FDC and AFDC percentiles for the whole period of record. All of these studies made a series of assumptions because of statistical components, such as variables that are independent and identically distributed. Singh et al. [8] introduced Shannon entropy theory for modeling FDC, where the entropy of discharge or the probability density function (PDF) was used to express the uncertainty of flow. This method needs no fitting for the whole period of discharge record and no assumption about daily flow. This method contains an entropy parameter M which plays a fundamental role in the derivation of FDC. The objective of this paper was, therefore, to further study the temporal and spatial characteristics of the entropy parameter M in the entropy-based method and apply the method to 73 sites in Brazos River basin, Texas, USA.

2. Materials and Methods

The derivation of the FDC and the study area are described in this section. For the dataset, codes, and software information used in this paper, please see the Supplementary Materials. For the derivation of the FDC, first, the entropy of discharge is introduced, then the constraints for the probability density function (PDF) are determined. Second, entropy maximizing is conducted by using the method of Lagrange multipliers and solved numerically. Third, the cumulative probability distribution function (CDF) is embedded in the process and a relationship between the discharge and exceedance period is derived.

2.1. Derivation of FDC

The derivation of FDC using Shannon entropy is detailed in Singh et al. [8]. For the sake of completeness, a brief synopsis is given here. The Shannon entropy of discharge (Q) or $f(Q)$ [$H(Q)$] can be expressed as:

$$H = -\int_{Q_{min}}^{Q_{max}} f(Q) \ln[f(Q)] dQ \tag{1}$$

where Q_{min} and Q_{max} are the minimum and maximum discharges, respectively, and $f(Q)$ is the PDF of Q. The objective is to derive $f(Q)$ by maximizing H for which two constraints are defined as:

$$C1 = \int_{Q_{min}}^{Q_{max}} f(Q) dQ \tag{2}$$

$$C2 = \int_{Q_{min}}^{Q_{max}} Q f(Q) dQ = \overline{Q} = Q_m \tag{3}$$

where Q_m is the mean discharge. Entropy maximizing is done using the method of Lagrange multipliers:

$$L = -\int_{Q_{min}}^{Q_{max}} f(Q) ln f(Q) dQ - (\lambda_0 - 1)\left(\int_{Q_{min}}^{Q_{max}} f(Q) dQ - C1\right) - \lambda_1\left(\int_{Q_{min}}^{Q_{max}} Q f(Q) dQ - C2\right) \tag{4}$$

where L is the Lagrangian function, and λ_0 and λ_1 are the unknown Lagrangian multipliers. Differentiating Equation (4) with respect to $f(Q)$ and equating the derivative to zero yield the PDF of Q as:

$$f(Q) = exp(-\lambda_0 - \lambda_1 Q) \tag{5}$$

Substitution of Equation (5) in Equations (2) and (3) yields the solution for λ_0 and λ_1:

$$\lambda_0 = -ln\lambda_1 + ln[exp(-\lambda_1 Q_{min}) - exp(-\lambda_1 Q_{max})] \tag{6}$$

$$-\frac{1}{\lambda_1} - \frac{Q_{min}exp(-\lambda_1 Q_{min}) - Q_{max}exp(-\lambda_1 Q_{max})}{exp(-\lambda_1 Q_{min}) - exp(-\lambda_1 Q_{max})} = -\overline{Q} \tag{7}$$

The entropy parameter M is defined as $\lambda_1 Q_{max}$.

In order to construct an FDC, a relation between the CDF of Q and time needs to be hypothesized. A possible form of CDF can be expressed as:

$$F(Q) = 1 - a\left(\frac{t}{T}\right)^b \tag{8}$$

where a and b are coefficients, t is the number of days that discharge is being equaled or exceeded, and T is the total number of days for a year. Parameters a and b can be estimated by empirical fitting and it is hoped that they will be relatively stable.

Differentiating Equation (8) we obtain:

$$dF(Q) = f(Q)dQ = -ab\left(\frac{1}{T}\right)^b t^{b-1}dt \tag{9}$$

Substituting Equation (5) into Equation (9), integrating from Q to Q_{max}, replacing the term $exp(\lambda_0)$ from Equation (6) and replacing $\lambda_1 Q_{max}$ with M, the final FDC is obtained as:

$$\frac{Q}{Q_{max}} = -\frac{1}{M}ln\left\{exp(-M) - \left[exp(-M) - exp\left(-\frac{MQ_{min}}{Q_{max}}\right)\right]a\left(\frac{t}{T}\right)^b\right\} \tag{10}$$

Equation (10) contains Q_{max}, and Q_{min} which are known from observations, and M which can be calculated using Equation (7).

2.2. Study Area

The entropy parameter M was determined from observations and its space-time characteristics were then investigated. It was also related to the drainage area. Then, FDC was constructed and its reliability was assessed.

The study area was Brazos River basin (Figure 1) which extends from Eastern New Mexico to Southeastern Texas, up to the Gulf of Mexico. The basin has a length of approximately 1219 km and a width varying from about 133 km in the High Plains in the upper basin to a maximum of 210 km in the vicinity of the city of Waco, to about 19 km near the city of Richmond in the lower basin. The basin drainage area is approximately 116,550 square kilometers, with about 111,370 square kilometers in Texas and the remainder in New Mexico [9]. There are 73 gauging stations with discharge records 50 years long that were analyzed in this paper. Daily maximum, minimum and mean discharges; and reservoir and gauge station information were collected from the USGS website (https://waterdata.usgs.gov/nwis).
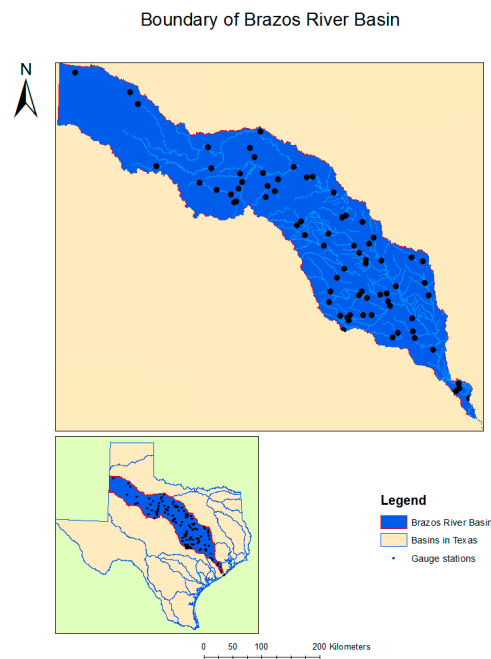
Boundary of Brazos River Basin

Figure 1. Brazos River basin.

3. Results and Discussion

3.1. Flow Duration Curve Estimation

The entropy parameter M is defined as $\lambda_1 Q_{max}$, where λ_1 can be obtained from Equation (7) by numerical solution with the observed Q_{max}, Q_{min} and $\overline{Q}$.

Using Equation (10) and the observed data, an FDC was constructed using the entropy parameter M, as shown in Figure 2.
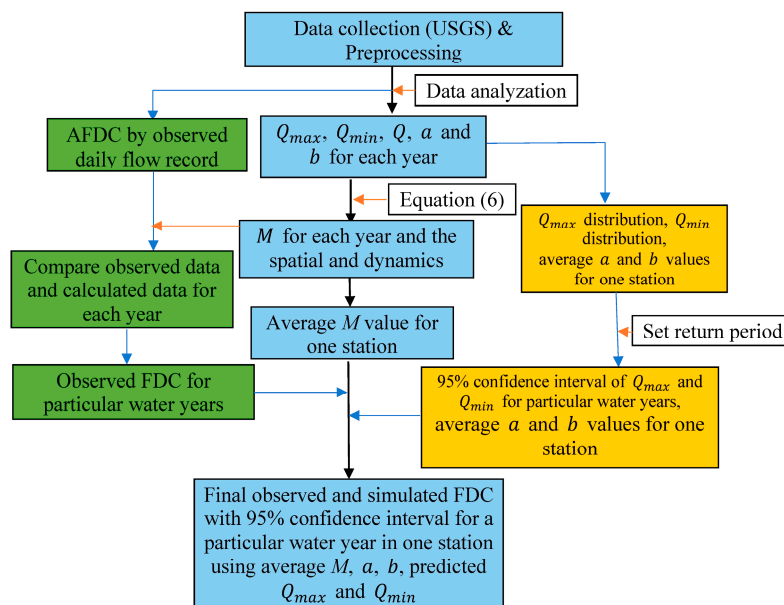

Figure 2. Schematic of the FDC construction.

First, the FDC of a specific year for a station was analyzed. Taking station 08093100 as an example, for 2009, M, calculated from Equation (7), equaled 10.47. After constructing the FDC for observations, parameters a and b were calculated using Equation (8) as $a = 1.021$ and $b = 0.778$. Substituting M, Q_{max}, Q_{min}, a, and b in Equation (10), we estimated the FDC. The correlation coefficient (R^2) between the observed and estimated FDCs was 0.969, which showed a good agreement, as shown in Figures 3 and 4.

Figure 3. Relationship between F(Q) and t/T of station 08093100 in 2009.

Figure 4. FDC of station 08093100 in 2009.

Second, the FDC was predicted for a particular hydrologic year using average values of M, a, and b for one station. For station 08093100, a, b, and M were calculated for each year and their histograms were constructed, as shown in Figures 5 and 6, and then their average values were estimated for the station. For the prediction of FDC, we needed to estimate Q_{max}, Q_{min}, and $\overline{Q}$ first by fitting the gamma distribution to each data set, as shown in Figures 7–9. For return periods of 1.3-year, 1.4-year, and 1.8-year, the estimated Q_{max}, Q_{min}, and $\overline{Q}$ with 95% confidence intervals were calculated, as shown in Table 1. The observed hydrologic years of 1.3-year, 1.4-year, and 1.8-year return periods were 2003, 2009, and 1994. The reason why we chose these years is that we wanted to focus on simulation for the recent years using parameters for a station. In addition, it showed that not all the stations followed good fitting, which is explained at the end of this section. Then, FDCs were predicted and compared with observed FDCs. The R^2 values of the predicted and observed FDCs were 0.979, 0.969, and 0.960,

respectively. Figures 10–12 show that 95% intervals covered most of the observed data. The same was done for other stations in the basin.

Table 1. Q_{max}, Q_{min}, a, b, and M for different water years for station 08093100.

Water Year	Year	Q_{max}	Q_{min}	LI Q_{max}	LI Q_{min}	UI Q_{max}	UI Q_{min}	a	b	M	R^2
1.3	2003	121.26	0.21	44.89	0.08	302.58	0.54				0.979
1.4	2009	169.09	0.3	68.16	0.12	395.75	0.7	1.02	0.89	9.88	0.969
1.8	1994	257.14	0.46	114.31	0.2	558.27	0.99				0.96

Note: LI means lower interval, UI means upper interval, discharge unit is m^3/s.

Figure 5. Relative frequency of parameter a at station 08093100.

Figure 6. Relative frequency of parameter b at station 08093100.
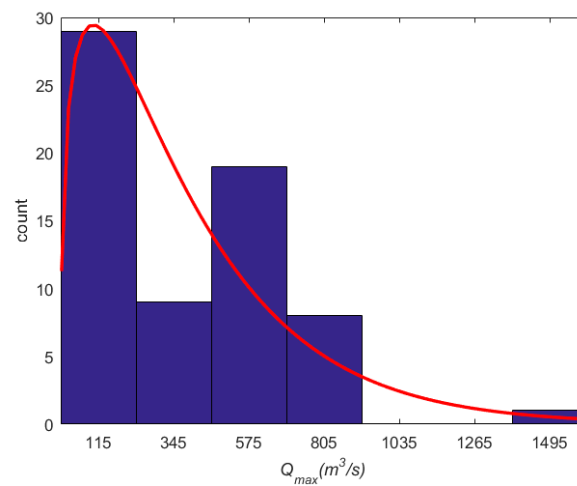
Figure 7. Gamma distribution fitting of the maximum discharge for station 08093100.
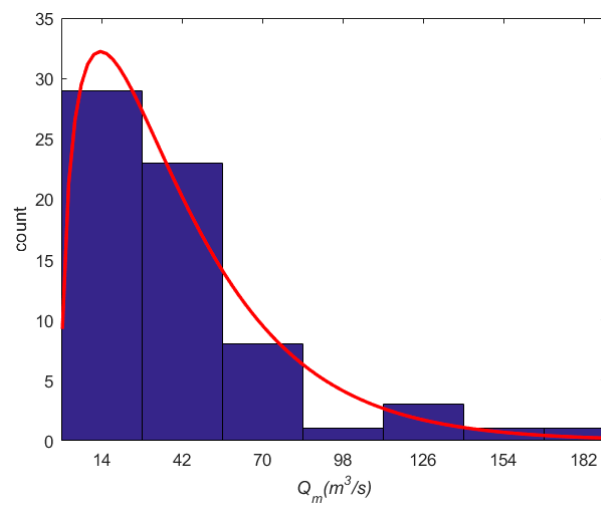

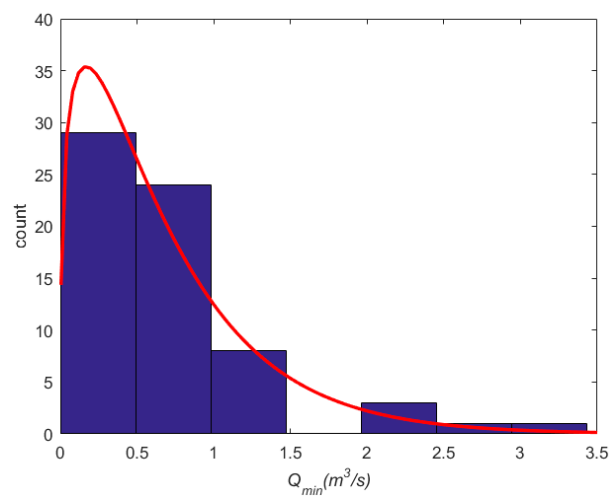**Figure 8.** Gamma distribution fitting of the mean discharge for station 08093100.

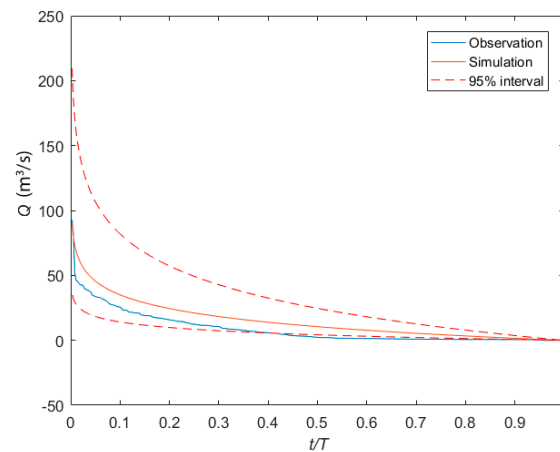**Figure 9.** Gamma distribution fitting of the minimum discharge for station 08093100.

Figure 10. Estimation of FDC for 2003 using average *M*, *a*, and *b* values of station 08093100.
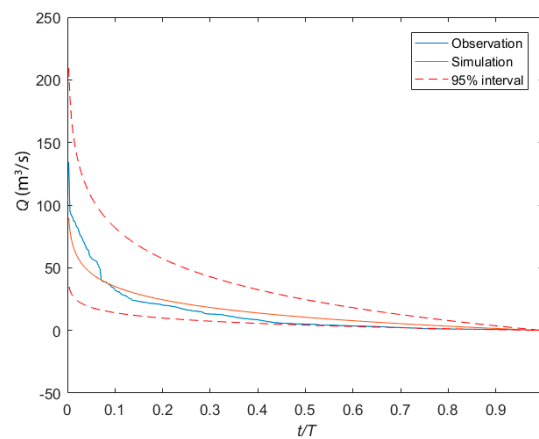

Figure 11. Estimation of FDC for 2009 using average *M*, *a*, and *b* values of station 08093100.
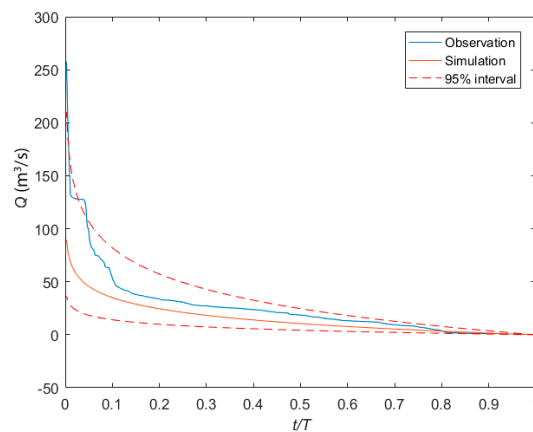

Figure 12. Estimation of the FDC for year 1994 using average *M*, *a*, and *b* values of station 08093100.

It was observed that the predicted FDCs fit well at most of the stations when discharges were relatively small, but were slightly poorer in the parts having large discharge values. Prediction for each year showed that R^2 was not always good. Figure 13 showed a good fit for the relationship with the ratio of $\overline{Q}$ and Q_{max}. When $\overline{Q}/Q_{max} \geq 0.10$, $R^2 \geq 0.90$. Further investigation could focus on making adjustments for better FDC prediction.
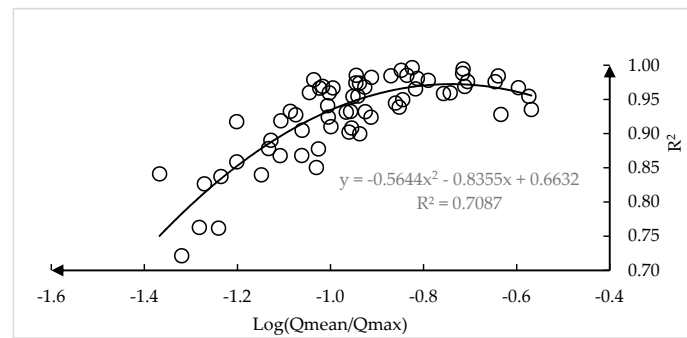
Figure 13. Relationship of R^2 and $\overline{Q}/Q_{max}$ for station 08093100.

3.2. Time Variability of M

The stream flow changes because of natural and anthropogenic factors, such as reservoir operation and climate change. First, we mapped the locations of reservoirs in the basin and analyzed the impact of reservoir on the time variability of *M* values. Reservoir locations, in part, are shown in Figure 14. As an example, we picked three stations, 08093100, 08099500, and 08093360, which were downstream of Whitney reservoir, Proctor reservoir, and Aquilla reservoir, respectively. The *M* values of these stations are shown in Figure 15a–c. For station 08093100, before 1951, the *M* value fluctuated, while after 1951 it was relatively stable because of the impact of the Whitney reservoir operation. The mean *M* value was 11.15 for the whole period, while the mean *M* value after 1951 was 9.88. It can be seen that the reservoir operation had a 12.85% influence on the *M* values for this station. However, our interest was in the period after 1951. Stations 08099500 and 08093360 had the same situation as did station 08093100, that is, the *M* values were fluctuating before the reservoir operation, but were stable thereafter. These stations were affected by the reservoirs by 189.15% and 43.82%, respectively. Similarly, there were other reservoirs in the basin which had an impact on the stations downstream of the reservoirs. For further analysis, we just chose record periods after the reservoir impact. After removing the impact of reservoirs, it was observed that the *M* values were relatively stable with time. At some stations, however, the *M* values jumped or fluctuated in some particular years.
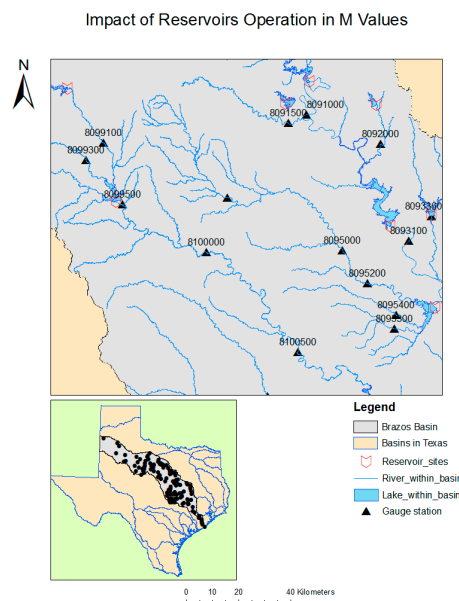

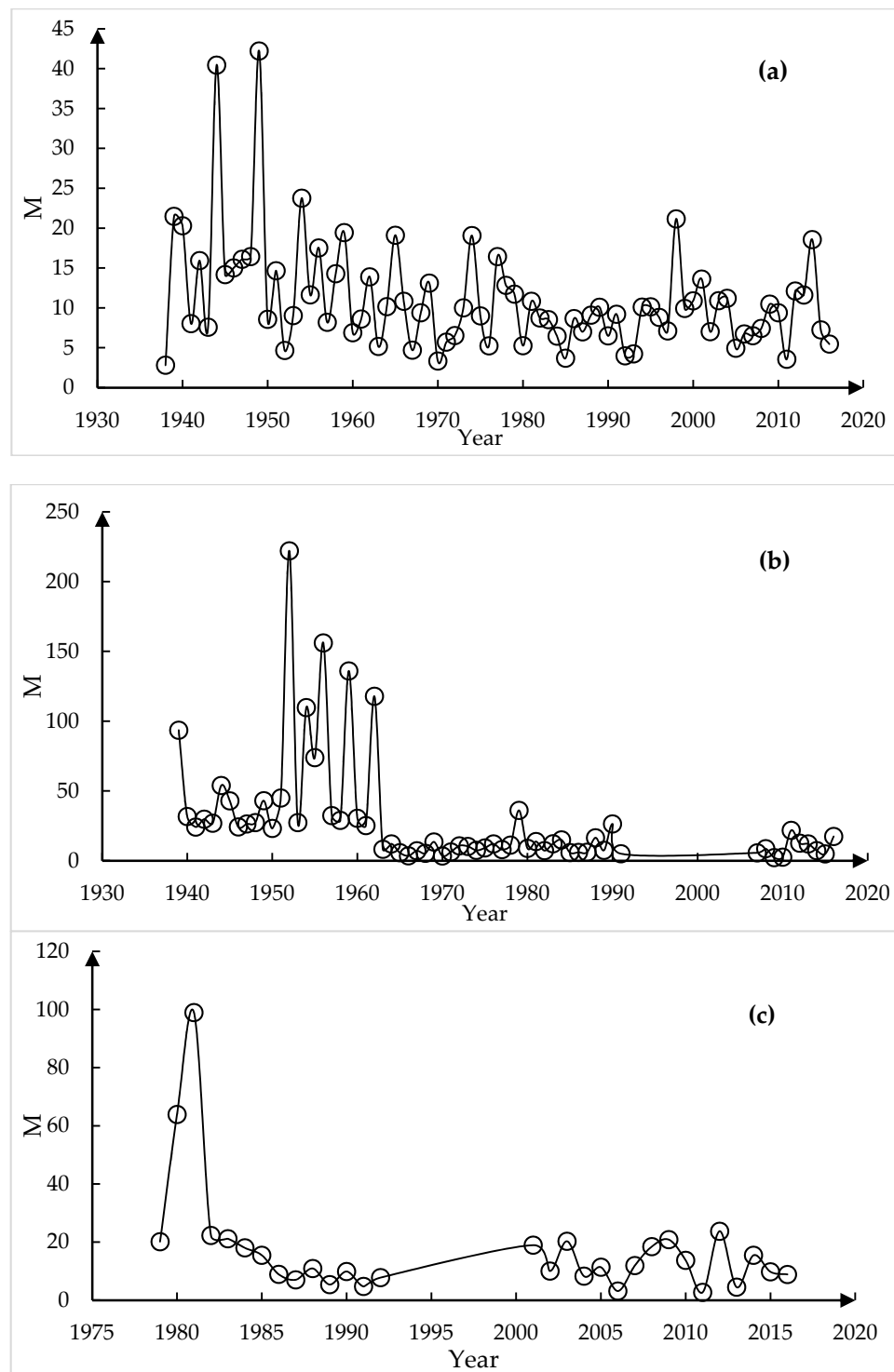**Figure 14.** Locations of reservoirs and stations (middle part of the basin, the scalar applies to the basin panel).

Figure 15. (**a**) *M* dynamics at station 08093100; (**b**) *M* dynamics at station 08099500; and (**c**) *M* dynamics at station 08093360.

Second, we determined the effect of climate change on the *M* values. *M* was defined as the Lagrange multiplier λ_1 times Q_{max}, as expressed by Equation (6), which relates it to Q_{max}, Q_{min}, and $\overline{Q}$. Though Equation (6) is slightly complicated, it can be simplified by setting Q_{min} equal to zero, which can usually be assumed to be near zero (it is true at most of the stations in the Brazos River

basin). Then we found that M had an inverse relation with the ratio of $\overline{Q}$ and Q_{max}, as shown in figures plotting M and the ratio (Figures 16 and 17)
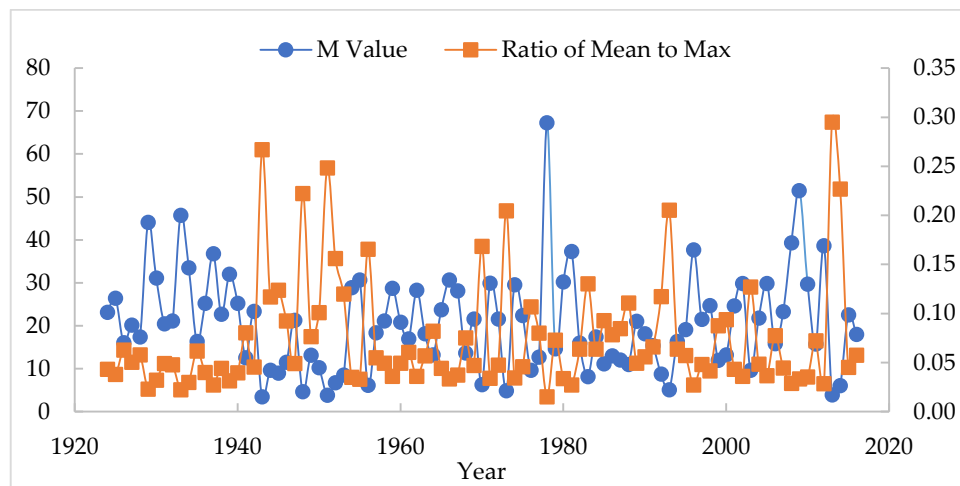

Figure 16. Correlation between M values and the ratio of $\overline{Q}$ and Q_{max} at station 08089000.
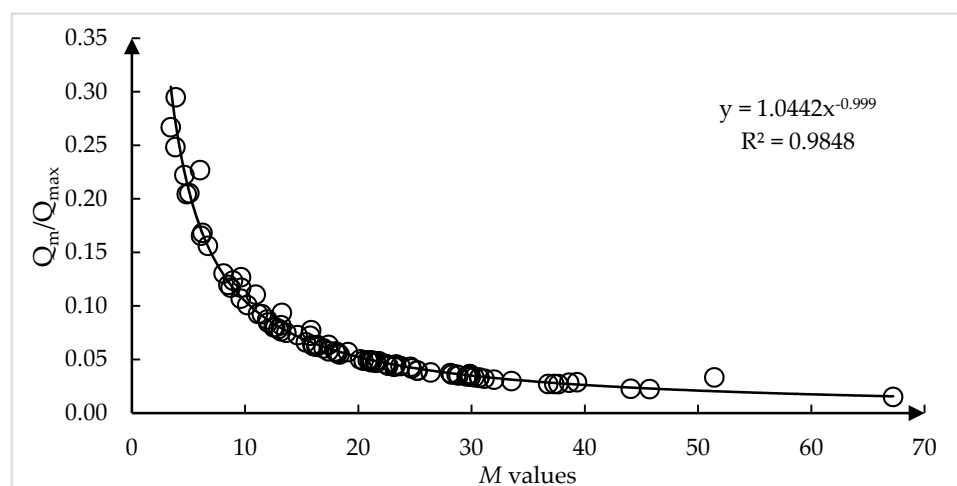

Figure 17. Powered relationship between M value and ratio of $\overline{Q}$ and Q_{max} at station 08089000.

Upon calculating M, the effect of climate change was determined. Studies on the impact of climate change on river discharge show that different parts of the basin have different impacts [10,11]. Discharge in a river can increase or decrease due to the impact of climate change and so can the ratio of $\overline{Q}$ and Q_{max}. Taking station 08089000 as an example, it can be seen from Figure 3 that the relation between M and the ratio had a correlation coefficient of −0.74, indicating that a high ratio is usually related to a low M value. At the same time, it was noticed that the M value had a dramatic jump in 1978 when Tropical Storm Amelia happened and caused a large storm in Texas [12]. It can be seen from Figure 18 that, in 1978, where there was an impact of the storm, there was a jump in the M value. This showed how M values reflected the change in flow characteristics related to the weather.
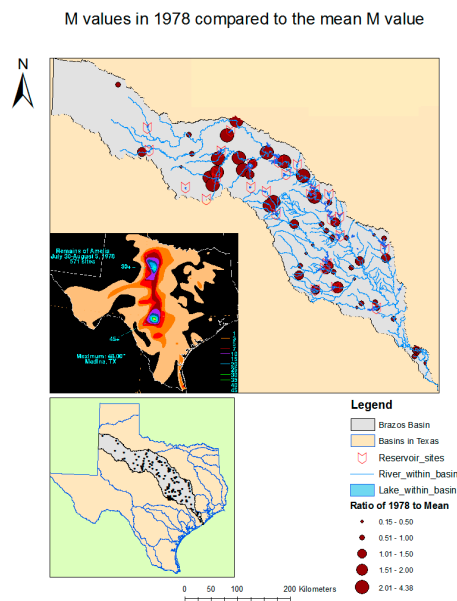
M values in 1978 compared to the mean M value

Figure 18. *M* values in 1978 compared to the mean *M* value.

The next step was to determine what other characteristics could be related to the M values, because the final goal was to apply this method to ungauged basins.

3.3. Spatial Variability

After calculating the *M* values for 73 stations and considering the impact of reservoirs, the mean *M* value was computed for each station. It was found that the *M* values ranged from 8.14 to 123.72. The lowest value occurred at gauge 08116650, which is located in the downstream part of the basin, and the highest value occurred at gauge 08086290, which is located in the middle-upper part of the basin. It can be seen from the map that most of the area in the upstream part had higher *M* values, higher than 55, the middle part had a range from 45 to 55, and the downstream areas had *M* less than 45. This showed a trend of decreasing *M* values from the upstream to the downstream part. It seems that the *M* values changed spatially because the drainage area changed, as shown in Figure 19, where if there was a small drainage area, then there was a large *M* value contour.
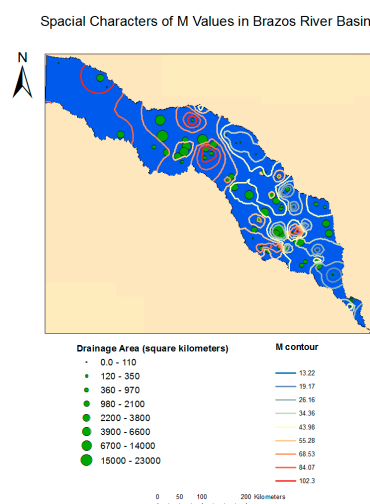
Spacial Characters of M Values in Brazos River Basin

Figure 19. Spatial features of drainage areas and *M* values.

Fuller [13] developed a relation between Q_{max} and $\overline{Q}$ as:

$$\frac{Q_{max} - \overline{Q}}{\overline{Q}} = 1.5A^{-0.3} \tag{11}$$

where A is the drainage area (square kilometers). This relationship indicates that the ratio of $\overline{Q}$ and Q_{max} would increase with an increase in the drainage basin size. Since M has an inverse relation with the ratio of $\overline{Q}$ and Q_{max}, M also has an inverse relation with basin size, which can be reflected by the correlation coefficient -0.536 and the plot of M versus the drainage size (drainage area) in Figure 20.
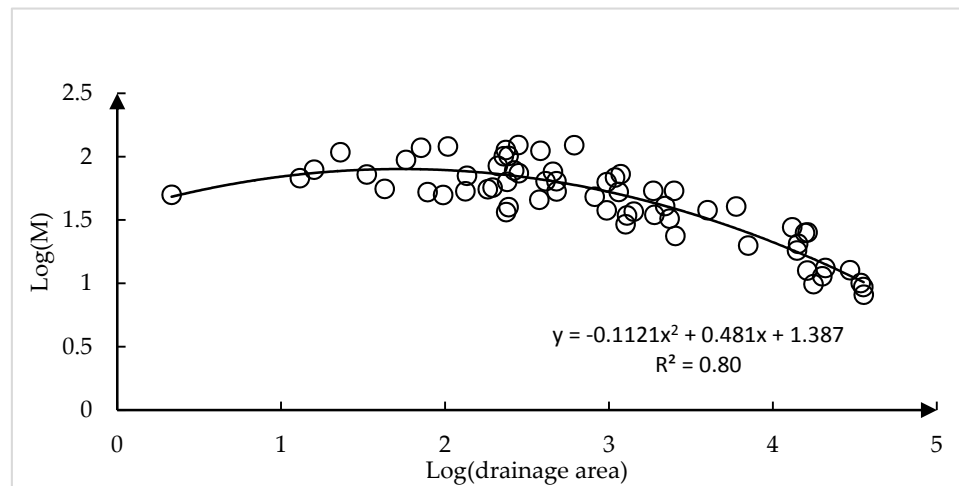

Figure 20. Relationship between the drainage area and M values.

3.4. Test for Ungauged Stations

We used station 08098290, assuming it as an ungauged station to test for the reliability of applying the function. First, following the schematic in Figure 2 and using the records from the station, we obtained the M value as the true value. Second, we estimated the M value using Equation (12):

$$\log(M) = -0.112[\log(A)]^2 + 0.481\log(A) + 1.387 \tag{12}$$

where A is the drainage area in square kilometers. The M value derived from records of observed data was 13.26. The M value simulated from the function was 14.23, which had a 7.31% difference. Third, we used both M values to form an FDC, compared to the empirical FDC, respectively, and calculated R^2 for both sides. Using the calculated M value led to a mean $R^2 = 0.91$, which ranged from 0.70 to 0.99, and simulated M led to mean $R^2 = 0.89$ which ranged from 0.68 to 0.95 which had a 2.20% difference with the calculated one. At last, we applied Equation (12) to all the stations in the basin and got simulated M for all the stations. The mean $R^2 = 0.86$ for the basin ranged from 0.58 to 0.93, while the calculated M from the records led to an $R^2 = 0.88$ and ranged from 0.61 to 0.95, which showed a mean difference between the results from the calculated and simulated M of 2.32%. Those test results indicated that the function can be applied to other ungauged stations.

4. Conclusions

This study analyzed in time and space the entropy parameter M which is basic to the entropy-based method for constructing the flow duration curve. Upon analysis of 73 stations in the basin, M ranged from 8.14 to 123.72, and was apparently impacted by anthropogenic and natural factors. Temporal patterns changed because of reservoir operation and flow characteristics. At the same time, M changed spatially with the drainage area. By analyzing the spatial and temporal

characteristics of M, a relation between M and drainage area was developed, a log-based function was fitted as y $= -0.112x^2 + 0.388x + 1.567$, which can be used in other basins. For most of the years, the average M yielded a good agreement between predicted and observed FDCs, where the mean R^2 was 0.92. Some years did not have good fit, especially in large discharge parts of the FDC; the reason why this occurred should be studied further. The procedure of applying the entropy parameter M for modeling the FDC can be extended to other basins. Further studies such as the adaptation to other basins, and improvement for the goodness of fit should be investigated.

Author Contributions: Vijay P. Singh, Yu Zhang, and Aaron R. Byrd, conceived the idea and designed the experiments; Yu Zhang performed the experiments; Yu Zhang analyzed and Vijay P. Singh supervised the data; Aaron R. Byrd contributed to discussion and analysis; Yu Zhang and Vijay P. Singh wrote the paper; and Aaron R. Byrd helped with revision. All three authors, Yu Zhang, Vijay P. Singh, and Aaron R. Byrd, contributed to the paper throughout its preparation.

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

References

1. Atieh, M.; Taylor, G.; Sattar, A.M.; Gharabaghi, B. Prediction of Flow Duration Curves for Ungauged Basins. *J. Hydrol.* **2017**, *545*, 383–394. [CrossRef]
2. Atieh, M.; Gharabaghi, B.; Rudra, R. Entropy-Based Neural Networks Model for Flow Duration Curves at Ungauged Sites. *J. Hydrol.* **2015**, *529*, 1007–1020. [CrossRef]
3. LeBoutillier, D.V.; Waylen, P.R. A stochastic model of flow duration curves. *Water Resour. Res.* **1993**, *29*, 3535–3541. [CrossRef]
4. Vogel, R.M.; Fennessey, N.M. Flow-duration curves. I: New interpretation and confidence intervals. *J. Water Resour. Plan. Manag.* **1994**, *120*, 485–504.
5. Singh, K.P. Model flow duration and streamflow variability. *Water Resour. Res.* **1971**, *7*, 1031–1036. [CrossRef]
6. Cigizoglu, H.K.; Bayazit, M. A generalized seasonal model for flow duration curve. *Hydrol. Process.* **2000**, *14*, 1053–1067. [CrossRef]
7. Castellarin, A.; Vogel, R.M.; Brath, A. A stochastic index flow model of flow duration curves. *Water Resour. Res.* **2004**, 40. [CrossRef]
8. Singh, V.P.; Byrd, A.; Cui, H. Flow duration curve using entropy theory. *J. Hydrol. Eng.* **2013**, *19*, 1340–1348. [CrossRef]
9. Wurbs, R.A.; Bergman, C.E.; Carriere, P.E.; Walls, W.B. *Hydrologic and Institutional Water Availability in the Brazos River Basin*; Technical and Special Reports; Texas Water Resources Institute: College Station, TX, USA, 1988.
10. Christensen, N.S.; Wood, A.W.; Voisin, N.; Lettenmaier, D.P.; Palmer, R.N. The effects of climate change on the hydrology and water resources of the Colorado River basin. *Clim. Chang.* **2004**, *62*, 337–363. [CrossRef]
11. Tao, B.; Tian, H.; Ren, W.; Yang, J.; Yang, Q.; He, R.; Lohrenz, S.E. Increasing Mississippi river discharge throughout the twenty-first century influenced by changes in climate, land use and atmospheric CO_2. In Proceedings of the AGU Fall Meeting Abstracts, San Francisco, CA, USA, 15–19 December 2014.
12. Roth, D. *Texas Hurricane History*; National Weather Service: Camp Springs, MD, USA, 2010.
13. Fuller, W.E. *Flood Flows*; Transactions of the American Society of Civil Engineers: New York, NY, USA, 1914; Volume 77, pp. 564–617.