RESEARCH ARTICLE OPEN ACCESS

Cascade process and Pareto rule: application to runoff data of two Mexican rivers (Conchos and Nazas)

J. R. Mercado-Escalante, Pedro A. Guido-Aldana, Waldo Ojeda-Bustamante.
(Mexican Institute of Water Technology;)

ABSTRACT
This paper has two objectives: the first is to show that the annual runoff of rivers such as Conchos and Nazas (Mexico), follow the Pareto rule of 80:20 when classes are ordered from largest to less, and can be compared with cascade processes. The second objective is to show that cascade process produce the core which gives rise to fractional integral and therefore to differential equations of fractional order. Finally, we conclude that the Pareto rule is a first approach to saturation described by the complementary characteristic function, and runoff data provide the order of the temporal derivative. Therefore, cascade processes are manifested ubiquitously in nature, and show us a way to evolve towards the imbalanced and become in distribution mechanisms that turn into a transition that destroys old and build new correlations.
***Keywords* -** Pareto rule, rivers, Cantor process, cascade process, probability density, structure function.

I. INTRODUCTION

The first objective of this paper is to show that the annual runoff of Conchos River reflect the postulate of Pareto when its data are arranged from highest to lowest, and compared with cascade processes. Pareto principle, also known as the 80:20 rules, is named in honor of Vilfredo Pareto, Italian sociologist and economist who enunciated it in 1906. The principle specifies an unequal relationship between inputs and outputs stating that for many phenomena, 20% of invested input is responsible for 80% of the results obtained. Put another way, 80% of consequences stem from 20% of the causes. This simple rule is observed across many disciplines and sectors of society.

Cascade processes shown ubiquitous in nature, and present us a way to evolve towards unbalanced and become into distribution mechanisms that turn into a transition which destroys old correlations and build new. A leading example is the energy cascade in the process of turbulence which distinguishes as a distribution mechanism of kinetic energy and its eventual transformation into heat energy, and is known as Kolmogorov cascade or Richardson.

Cascade processes are described by means of its tree form and the sequence of a generalized Cantor process, Fig. 1. The issue is to split gradually a unit to create ever smaller parts, but in turn increasingly numerous, thereby simultaneously combine the infinitely small with the infinitely numerous, and configure the ways in which passes a flow or a counterflow. Flows in these tree structures may be in the forward direction, as in the case of energy cascade, or in the reverse direction, as in the case of a soliton formation, this last a solitary wave propagating in a nonlinear medium without deformed (Scott, 1844); or in the process of productive accumulation and centralization in the economic sphere. If a process can result in a reverse process, an oscillating behavior is created. Of course there are equivalent descriptions as binary formulation and graphs, which allow relatively easily highlight important properties or characteristics. And we understand that split is similar to bifurcate, [1].

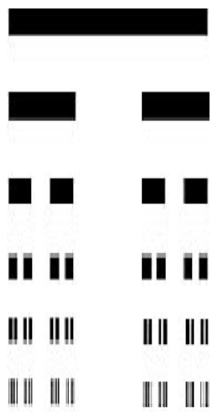

Figure 1. Cantor set, example of lineal fractal.

The transformations of the order and correlations in disorder, or entropy, and new correlations reflect the arrow of time; while the opposite should occur in the reverse process. As Boltzmann visualized, entropy is a measure of disorder or molecular chaos, while the natural evolution occurs through its increment with the trend toward maximum chaos as possible, mechanism through which energy becomes unavailable, because individual correlations within the system are destroyed or mixed; the balance could be called the state of maximum entropy, the chaos; and chaos, the subtle form of order. Evolution towards imbalance and transition shows a theory of "becoming" as Goethe had claimed, [2]; and that, we celebrate as the conjunction between deterministic and random.

Moreover, as a second objective we see that the above tree structures can also be used to construct the most famous and important mathematical functions from the point of view of the approaches, and establish their links with the differential equations. So differential equations will not be seen as a starting point but as an arrival point or second stage of a process of mathematical modeling.

"For science, a phenomenon is ordered if their movements can be explained in a scheme of cause and effect represented by a differential equation". However, correlations open a range of possibilities within which some causes operates in a more decisive way, with more heft, while others do so more weakly. Correlations are established due to the presence of a set of transient causes.

II. METHODS

There are available historical records of annual runoff volumes for Conchos River, a tributary of the Rio Grande, which are led to the storage dam 'La Boquilla' in the Irrigation District 005 Delicias, Chihuahua. Runoff volumes record that we analyzed comprises a period of 80 years (1935-2014), although it would be desirable to have records with 4 digits to provide greater statistical support to our conclusions. Fig. 2 shows a photograph of the storage dam 'La Boquilla'.

For analysis, data sets are formed, called classes, uniform in size. Distribution of all data available at that uniform collection is studied. Number of classes k can be chosen according to the rule or method of Sturges: $2^{k-1} = n$, [3] , being n the size of the data sample, which in this case is 80 years, and is estimated by its nearest integer. Frequencies are determined by counting the number of data in each class, and it can be derived to account for other forms of frequency, as in the case of relative frequencies, which allows introducing the probability as a measure of belonging to one of the sets-classes. Historical records of these annual volumes are presented in Table 1.

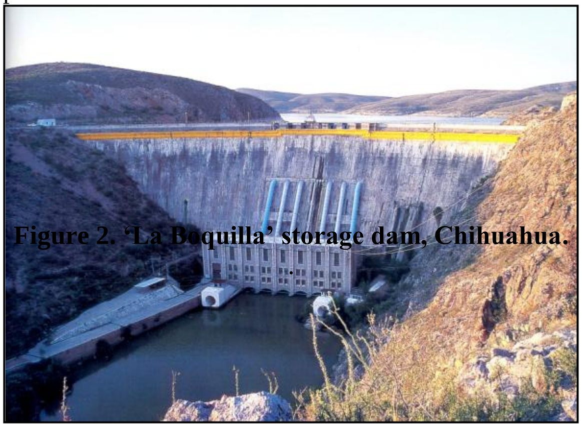

Figure 2. 'La Boquilla' storage dam, Chihuahua.

Table 1. Annual runoff volume in Mm^3 and grouped (Conchos River).

Class number	Class sets, ranges	Class centers (Mm^3)	*Frequencies*			
			Observed	Cumulated	Relative	Cum. Relative
1	0-500	250	12	12	15.3	15.3
2	500-1000	750	29	41	36.1	51.4
3	1000-1500	1250	21	62	26.4	77.8
4	1500-2000	1750	6	68	6.9	84.7
5	2000-2500	2250	8	76	11.1	95.8
6	2500-3000	2750	2	78	1.4	97.2
7	3000-3500	3250	2	80	2.8	100.0

A second way to group sets-classes, such as Pareto's proposal, is to sort the sets decreasingly according to their size, from largest to smallest, and in a similar way as cascade processes. This grouping is shown in Table 2.

Pareto assumption representation is made in Fig. 3, showing in the horizontal the frequencies in their new order, from highest to lowest, so that the greatest frequency is closer to the origin than the smallest. Fig. 3 shows that the highest frequency is around 20% and the cumulated relative frequencies grow from there about 20%, up to 100%, thus making a distance of the order of 80%. So, the famous proportion of the efficiency of 80:20 enunciated by Vilfredo Pareto (1848-1923), Marquis, engineer and specialist in mathematical economics is illustrated.

Table 2. Frequencies ordered from largest to smaller for annual runoff in Conchos River.

	Frequencies			
Class number	Observed	Cumulated	Relative	Cum. Relative
N2	29	29	36.25	36.25
N3	21	50	26.25	62.50
N1	12	62	15.00	77.50
N5	8	70	10.00	87.50
N4	6	76	7.50	95.00
N6	2	78	2.50	97.50
N7	2	80	2.50	100.00
	80		100.00	

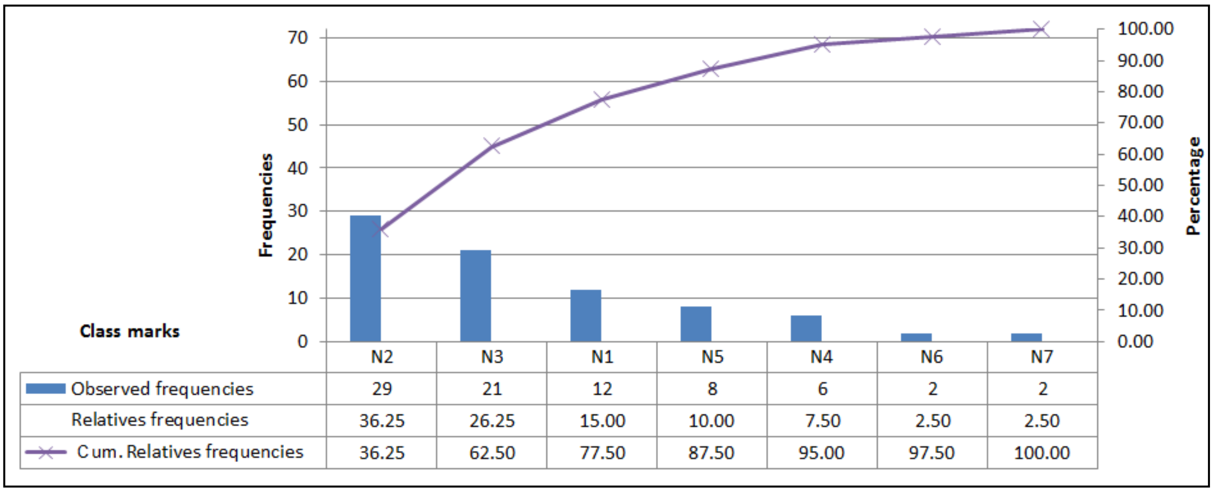

Figure 3. Results of applying Pareto statement to Conchos River runoff data.

2.1 Probability density

Moreover, for ungrouped data we perform a statistical analysis that allows us to find a probability density that give a structure to data, and its moment generating function, and then compared it with the ogive of pooled data.

We seek for the first four central moments and found the following values, where the first is zero, $[5.3089 \times 10^5, 4.4842 \times 10^8, 1.1205 \times 10^{12}]$. Then we look for the symmetry parameters (skewness) and the kurtosis, themselves happen to be, respectively $[1.1593, 4.0259]$; so it has an asymmetric distribution of thicker tails at right and quite leptokurtic, and that far exceeds the kurtosis of the Gaussian, so it is quite away from normal. Nor approaches the Gumbel that have a kurtosis of $12/5$, although its asymmetry is $1.14\ldots$ and can be considered of the same order to that found. We note that these two distributions are usually used to analyze this type of data.

Subsequently, we seek for classification parameters according to Pearson method, and found the following values $[-8.2823\ 10^5, -0.49421]$. As these two negative values we conclude that the probability density can be classified as Type I, or as Type Beta density.

Then we seek for the two shape parameters and are respectively $[-0.00027470, 2.7058]$, so it is a decreasing density, which is shifted to the left with respect to the center line, with reference to logistics.

Next, we calculate the diffusion coefficient by means of the variance of the probability density and the value of $0.029326$is found, so the areal velocity of diffusion is very slow.

Furthermore, we found Hurst exponent and the fractal dimension of the runoff graph, which are respectively $[0.50014, 1.4999]$, where the average of the two must be unity; therefore both represent a competition among them with zero-sum, so the gain of one results in the loss of the other, so that the sum thereof be zero; but also, being $H > 1/2$ it is classified as persistent.

Then, being the classification as Type I and its shape parameters, we find the graph of the probability density, which is shown in Fig. 4.

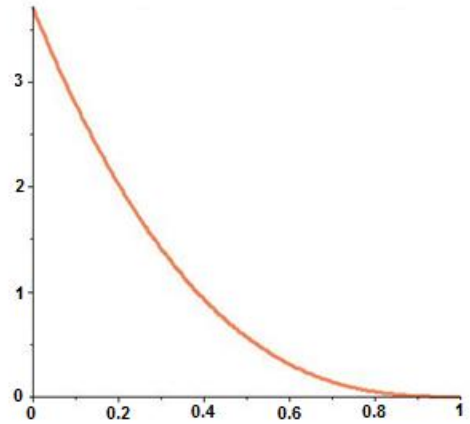

Figure 4. Probability density of Conchos River runoff.

Additionally, it is interesting to compare it with that outlined in grouped data which is obtained from the bar chart with ogive. You can imagine that with the gradual choice of more classes' number and therefore getting smaller sets this graph may evolve to that shown in Fig. 4. Ogive is presented in Fig. 5 which shows also its asymmetrical nature with thicker tails right.

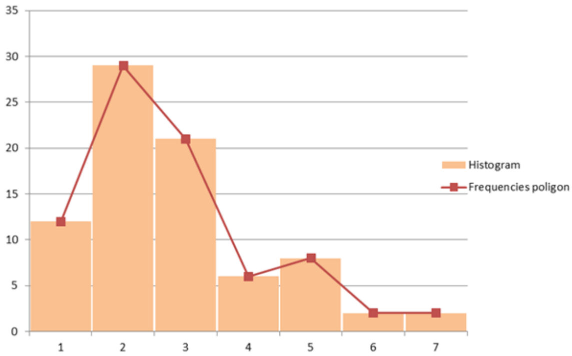

Figure 5. Histogram and frequencies polygon for Conchos River runoff.

Next, we find the moment generating function linked to the probability density previously found, and its graph can be seen in Fig. 6.

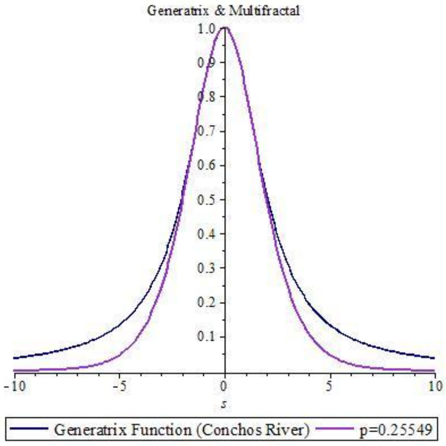

Figure 6. Moment generating function linked to the density of Fig. 4.

Now, we seek for a cascade process of binomial type that approximates the generating function and found a value of 0.25549, for the probability of success p that in the binomial, the same information dimension of the generating function produce. It is inferred that adjusting appropriately the parameters p and $q = 1 - p$ it can be approach the generatrix function of confluent type and can be described as a cascade process of binomial type.

2.2 Nazas River

Besides, we also have historical records for annual runoff volumes of Nazas River, which are gathered by the storage dam 'Lazaro Cardenas' in the irrigation district 017 of the Lagunera Region, Coahuila-Durango, Fig. 7.

Figure 7. 'Lazaro Cardenas' storage dam in Nazas River, Durango.

Flows drained record that we analyzed covers a period of 86 years (1929-2014). Historical records of annual runoff of river Nazas are presented below in Table 3.

Again, the second way to group classes, as proposed by Pareto, is ordered sets decreasingly according to its size, from larger to smaller, and similarly to cascade processes, as shown in Table 4.

Information is plotted, placing in the horizontal the relative frequencies in the new order, from highest to lowest, and Fig. 8 is obtained.

Therefore, we have been reflected in the two rivers, Conchos and Nazas, the statement of Pareto efficiency: 80% of the consequences stem from 20% of the causes. Or 80% of results come from 20% of efforts and time spent.

Table 3. Annual runoff volumes in Mm^3 and grouped for Nazas River.

Class number	Class (Mm^3)	Class centers	*Frequencies*			
			Observed	Cumulated	Relatives	Cum. Rel.
1	0 – 500	250	15	15	17.44	17.44
2	500 – 1000	750	27	42	31.40	48.84
3	1000 – 1500	1,250	18	60	20.93	69.77
4	1500 – 2000	1,750	13	73	15.12	84.88
5	2000 – 2500	2,250	9	82	10.47	95.35
6	2500 – 3000	2,750	1	83	1.16	96.51
7	3000 – 3500	3250	3	86	3.49	100.00

Table 4. Frequencies sorted from highest to lowest for annual runoff volumes in Nazas River.

Class marks	Frequencies	*Frequencies*		
		Cumulated	Relatives	Cumulated Rel.
2a.	27	27	31.40	31.40
3a.	18	45	20.93	52.33
1a.	15	60	17.44	69.77
4a.	13	73	15.12	84.88
5a.	9	82	10.47	95.35
7a.	3	85	3.49	98.84
6a.	1	86	1.16	100.00

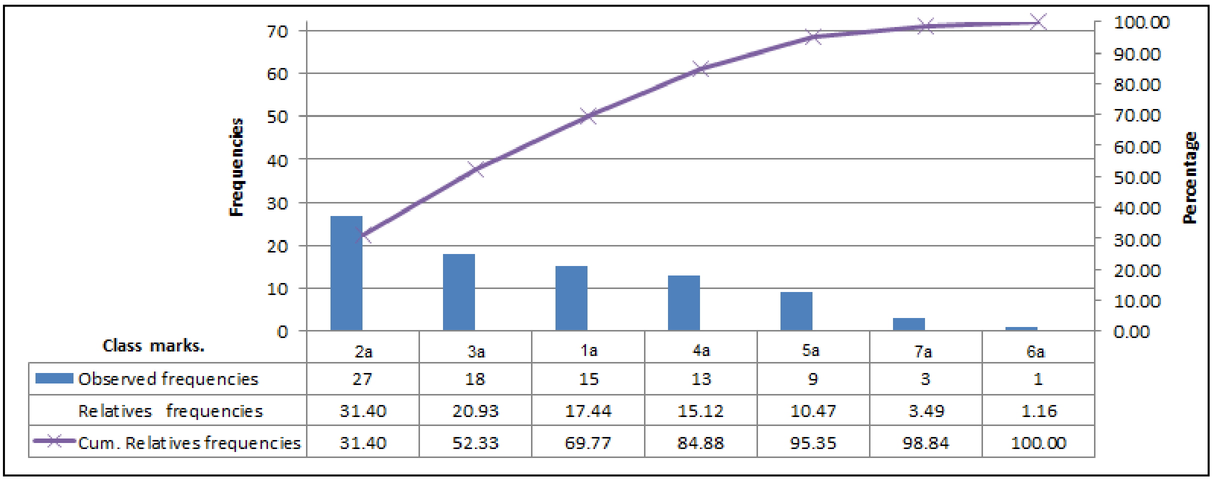

Figure 8. Results of applying Pareto sentence to runoff data of Nazas River.

2.3 Cascades

As second goal, cascade processes can be used to define the fractional integral. Cascade is now a sequence of mathematical operations that do the traits, progressively smaller, and in turn, progressively more numerous, so that in the limit the infinitely small is reached with the infinitely numerous and that the product of these two determine the result.

Instead of the pair of numbers $(3,2)$ by Cantor process, a function of the resolution is determined that for this case is defined by $f(h)=h^{-\nu}, \nu>0, h>1$. Procedure of fractionate is performed by means of the first factor in the following expression $C_{\nu,n}$; while the second represents the progressive growth of traits. Thus in each stage of generalized Cantor process, the calculation outlined in (1) is performed,

$$C_{\nu,n} = \left((-1)^n D_h^n\left(h^{-\nu}\right)\right)\cdot(n+1)D_p^{-n}(p)\Big|_{p=h} \qquad (1)$$

So that for the first stage we have $C_{\nu,1} = \frac{(\nu)_1}{1!}h^{-\nu+1}$; then in general for the nth stage of this process $C_{\nu,n} = \frac{(\nu)_n}{n!}h^{-\nu+n}$ is obtained, where $(\nu)_n = \frac{\Gamma(\nu+n)}{\Gamma(\nu)}$ are the Pochhammer symbols expressed by successive factors, either in terms of the Euler gamma function. Resolutions sequence is discretized as $h_n = \frac{n}{x}, x > 0$; through $\Gamma_n(\nu) = \frac{(1)_n}{(\nu)_{n+1}}n^\nu$, we obtain C_n and its limit C_ν, $C_n = \frac{x^{\nu-1}}{\Gamma_n(\nu)((\nu/n)+1)} \to \frac{x^{\nu-1}}{\Gamma(\nu)}$. Fractional integral is defined by the convolution ${}_0D^{-\nu}\varphi(x) = C_\nu * \varphi$; therefore, with the generalized Cantor process, applied to potential functions, nucleuses which define the fractional integral are constructed [4].

Additionally, from parameters obtained since Conchos River runoff data, we can do a multifractal construction, at least with two options. The first as follows: based on the data we can define a structure function of Kummer type.

In the second, the study focuses on changes that show graphs of runoff volumes, which look very irregular, so it makes no sense to think about introducing a slope, as limit value of relative changes relating to increments of its arguments.

Then, we consider the called fractional Brownian motion (fBm) defined from its variations or fluctuations. Those fluctuations are stationary, with two statistical parameters that are independent of the index (timeless): the mean and the variogram; their distributions are Gaussian and also are statistically self-affine, which manifests as "Hurst effect." Auto covariance function of the fluctuations depends only on increases or lags. So, for large lags, the correlation function approaches one of the above types, and therefore, again, based on the data, Hurst exponent is estimated, which is considered as a measure of independence of time series and a way to distinguish fractals series, [5]. In Fig. 9 runoff data are shown, from 1935 to 2014.

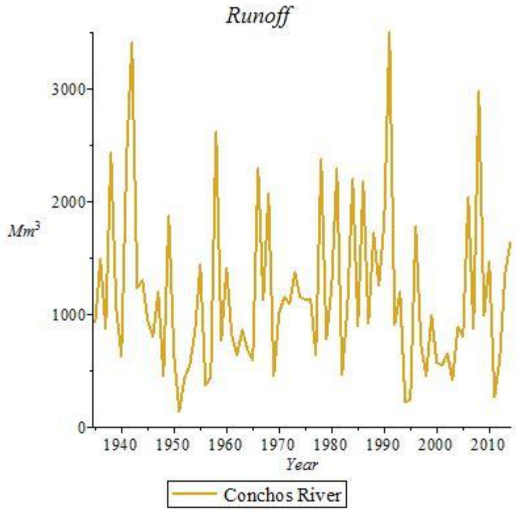

Figure 9. Runoff variation in Conchos River, 1935 – 2014.

It is important to note that correlation function tails decay slowly and thus represent long-range correlations, [6]; and, the larger the Hurst exponent the slower the fall and the longest the tail, [7].

Once you found the Hurst exponent from the data, settle fractional evolution equation given by the differential equation (2), whose order is specified by $\gamma = 2 - 2H$, and which also supports the form $D_t^\gamma(1-\psi(t)) = \lambda^{-\sigma}\psi(t)$, where $\lambda^{-\sigma}$ is a proportionality constant which generally depends on the spatial behavior of the phenomenon.

$$D_t\psi(t) = -\lambda^{-\sigma}D_t^{1-\gamma}\psi(t) \qquad (2)$$

One solution is expressed by $\psi(t) = E_{1,\gamma}\left(-\lambda^{-\sigma}t^\gamma\right)$, which corresponds to the Mittag-Leffler function whose parameter is $\gamma = 0.99972$ and displayed in equation (3), with $(\sigma = \gamma, \lambda = 10)$, [8]. The Mittag-Leffler function is normalized to $t = 0$ by the value 1 and is decreasing convex; so the limit is zero, with zero slope. Then we define the complementary characteristic function by function 1-Mittag-Leffler, we now have a growth curve with saturation is normalized to zero 0, which is concave increasing and the limit is 1, with zero slope. Its graph is filed in Figure 10 and there also represent again the ogive points in class centers, [8].

$$E_{1,\gamma}\left(-(t/\lambda)^\gamma\right) = \sum_{j=0}^{\infty}\frac{\left(-(t/\lambda)^\gamma\right)^j}{\Gamma(1+\gamma\cdot j)} \qquad (3)$$

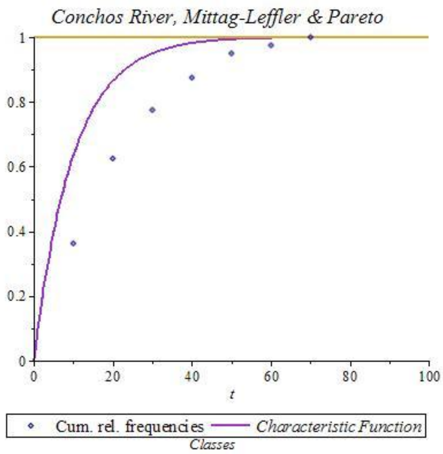

Figure 10. Conchos River. Complementary Characteristic function, Mittag-Leffler,

Phenomenon description as fBm has a variance that scale according to $\gamma = 2 - 2H$, and its correlation function. When it is estimated for large lags, in approximate way, corresponds to the generatrix function of a binomial, with cascade shape as power of the success probability, but the power of failure probability without heft, see equation (4). The power of probability success depends linearly on Hurst exponent, [6].

$$\tau(s) = {}_1F_1(0.99973\,,\, 4.7055\,, -s) \qquad (4)$$

Regarding the first option of the above, we recall briefly that we are dealing with sets like those with scattered and irregular variations. For the multifractal construction, we cover the entire set with mesh cubes with h resolution. We define the partition function as the total sum of the system available macro-configurations. Sum runs over all those cubes that intersect the support of microscopic measurement. That sum behaves as a power of the h resolution, (see (5)), and the power that represents the partition function is known as the structure function; in our case is the Kummer function $\tau(s)$ and is shown in Fig. 11.

$$Z_h(s) \propto h^{-\tau(s)}, \qquad \frac{Z_h(s)}{Z_h(1)^s} = h^{-\tau(s)} \qquad (5)$$

Moreover, the $N_h(\alpha)$ traits also obey a power law and grow when the h resolution approaches zero,

$$N_h(\alpha) \approx h^{-f(\alpha)} \qquad (6)$$

But Legendre transform of the multifractal spectrum $f(\alpha)$ is the structure function $\tau(s)$. And, the maximum is reached in α, when the concave curve $f(\bar{\alpha})$ is above and as far as possible of the straight $s\bar{\alpha}$, with slope s and argument $\bar{\alpha}$, then

$$\tau(s) = \sup_{\bar{\alpha} \geq 0}\{f(\bar{\alpha}) - s\bar{\alpha}\}$$
$$\downarrow$$
$$\tau(s) = f(\alpha(s)) - s\alpha(s) \qquad (7)$$

Conversely, in the convex dual, when the Legendre inverse transform exists, in this way, the structure function is recovered from the multifractal spectrum, but now its argument is the s scale as dual variable to singularity α.

Therefore, it's possible to assign to Conchos River a structure function of Kummer type, as that shown in Fig. 11.

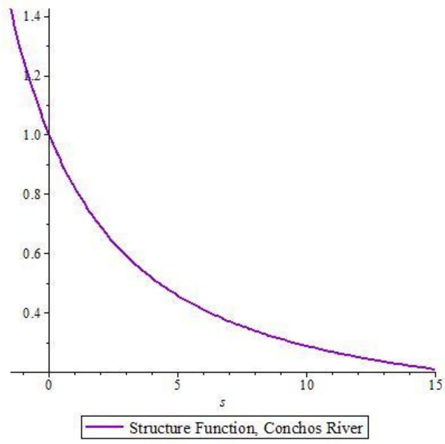

Figure 11. Kummer's function structure of Conchos River.

The most important features are: is decreasing convex, normalized to zero by the value 1, and when the scale variable increases, the function value decreases, and its decrement is progressively smaller until it eventually becomes zero. Or, the structure function is decreasing with decrements that are progressively decreasing. This decreasing is concomitant with progressive smallness of the traits of multifractal process.

Dually, a spectrum both punctual as continuous or implicit could also be assigned to Conchos River. Now it's about a concave function, such that when the singularity decreases, the function value also decreases and eventually reach zero. Again this behavior is concomitant with the multifractal process, where traits are progressively smaller, and reflects a cascade process from the vertex, as the singularity decrease, spectrum value decrease too. Here, however, the decrement of spectrum value is not given by the value of the scale equal zero, but from $s = 3$, approximately. Spectrum is shown in Fig. 12.

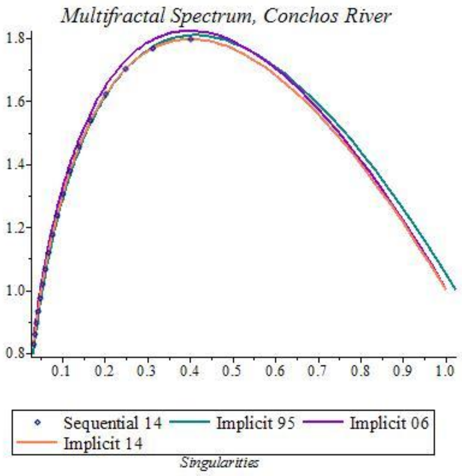

Figure 12. Spectrum evolution of Conchos River.

Finally, we show the evolution of the spectrum showing results with spacing approximately a decade: 1995, 2006 and 2014. In Figure 12, great stability in the spectrum is observed, despite the great variability of the data that feed it. In particular, the stability is still higher for sufficiently large scales, which is interpreted as the distant past or high frequency; so we would argue that their own time evolution is reflected throughout each curve of the spectrum.

III. CONCLUSIONS

Runoff variations, as illustrated in the case of Conchos River, have the characteristic of being scattered and irregular. Regularity within the irregularity is constructed by means a multifractal process. To do this, from the data, a probability density that structured it is found. Hurst exponent is approximated based on one of its shape parameters. Whereby a fractional differential equation is created, whose order varies with the Hurst exponent, and whose solution represents a Mittag-Leffler function, which shows an intermediate decreasing between a fractal and an exponential tail. It also describes decay phenomena that represent long-range correlations, such that the higher the Hurst exponent slower the fall, and the tail of the density distribution become longer. Further, the differential equation represents a saturation phenomenon, where the fractional derivative of the complement, is proportional to the function itself. This saturation phenomenon is reflected roughly through the assumption of Pareto efficiency: 80% of the consequences stem from 20% of the causes.

REFERENCES

[1] J. R. Mercado, La ecuación de Navier-Stokes y multifractales, Revista de Matemática: Teoría y aplicaciones, 15(1), 2008, 49-72.
[2] J. P. Briggs and F. D. Peat, *A través del maravilloso espejo del universo*, (Gedisa edit. Barcelona, 312p. 1996).
[3] H. Sturges, The choice of a class-interval, J. Amer. Statist. Assoc., 21, 1926, 65–66.
[4] J. R. Mercado, Las ecuaciones de Reynolds y la relación de clausura, Revista de Matemática: Teoría y Aplicaciones, 16(1), 2009, 104–125, CIMPA – UCR ISSN: 1409-2433.
[5] H. Hurst, The long-term storage capacity of reservoirs, *Transactions of the American Society of Civil Engineer*, 1951, 116 - 195.
[6] R. L. Bras and I. Rodríguez-Iturbe, Random functions and hydrology, Chapter 5, Long-term Persistence in Hydrologic Modeling, (Dover Pub., New York, 1993), 561p.
[7] F. J. Molz, H. H. Liu and J. Szulga, Fractional Brownian motion and fractional Gaussian noise in subsurface hydrology: A review, presentation of fundamental properties, and extensions, Water resources research, Vol. 33, No. 10, 1997, 2273-2286.
[8] R. Metzler and J. Klafter, The random walk's guide to anomalous diffusion: A fractional dynamics approach, Phys. Rep. 339, 2000, 1-77.