Using q-distributions on the study of side inflows for Koman basin in the Drin River, Albania

Dode Prenga

***Abstract*. We consider herein the q-Gaussian distributions, fractal and multi fractal analysis on the study of time data series of inflows in the lake of Koman at Drin River, Albania. It is a typical hydro-electric system which encompasses natural hydrological events, specifics of terrain and human based activity at the time. We see that for direct records as from routine technical measurements, the distribution is difficult to be determined, or the regime belongs to the non-stable distributions. Using q-functions and complex systems methods, a good picture of statistical and dynamical aspect if found. Accordingly, the smoother series produced by averaging values over some few successive points do have stable distribution of values and are analyzed. Evaluating thereof the triplet Tsallis, specific dynamic properties and behavior are identified and discussed for each case. Next, a log-periodic-like perturbation over q-Gaussians appears for some distributions. It may report a complex effect of the stochastic processes interferences or self-organized behavior that will emerge if some specific conditions are met.**

***Index Terms*— q-Gaussians, hydrological, complex system, multi fractal spectrum, log periodic.**

I. Introduction

Complexity approach in the study of hydrological systems dates long ago, with fractal and multifractal theoretical analyses. Combined views were successfully applied since. Here we will consider a case study on a particular such a system in Albania, the Koman basin. It is an artificial lake build in Drin River, and gathers the waters from Albanian Alps and from Fierza lake discharges or working waters on the HPP.

The flows from rainfall and snowfall on the area are coupled with terrain specifics, imposing unpredictable modification on side and natural water inflow. Many statistical distributions for similar system as flood-frequency analysis have been investigated in hydrology. The most commonly applied have been Extreme Value and Generalized Extreme Value, Log Paerson, Lognormal etc., [1] and others one were found suitable. Stretched exponentials can be used [2] as well. In general, no theoretical arguments will favor a particular distribution [3]. Hence we use the very interesting class of functions called q-exponentials as derived from generalized entropic principles [4] by C.Tsallis. There are strong arguments to classify hydrological-like systems as complex ones [5],[6] hence their typical methods can be used as well. Complex systems embrace physical entities of interacting elements, particles or subparts, typically nonlinearly, and full of dynamics [11], therefore standard equilibrium statistics and thermodynamics is not applicable. Remember that in the thermodynamic limit, the statistical mechanics uses the principle of the maximum Boltzmann-Gibbs entropy defined as $S = -\int_{suport}^{\Omega} (p_x \ln p_x) dx$ to obtain the distribution of the observables. This principle says that the most probable distribution {p} will maximize the above entropy. Here the probability distribution p_x in normalized. If the only constraint is the mean an observable, the well-known Gibbs distribution is found. Next if the constraints on the system fixed the variance of the observable, the distribution is found Gaussian.

Manuscript received October 10, 2014.

II. DISTRIBUTIONS IN COMPLEX SYSTEMS

The complex systems are generally found in out-of-equilibrium states [7],[8], therefore, Gibbs-Boltzmann statistics mentioned above is not applicable. To renormalize this situation C.Tsallis proposed another form for the entropy to be maximized, and defined it by the formula $S_q^T = \frac{1}{q-1}\left[1-\int p^q dx\right]$ [8].

Here the parameter q is a distance indicator from the equilibrium, as if q→0, the Tsallis entropy will result exactly the Boltzmann-Gibbs entropy. By optimizing it under appropriate constraints, he obtained the distributions in the so called q-exponential form defined as $p(x) = \alpha\{1-\beta(1-q)(x-\mu)\}^{\frac{1}{1-q}}$ and q-Gaussian form $p(x) = \alpha\{1-\beta(1-q)(x-\mu)^2\}^{\frac{1}{1-q}}$ according to the fact of which quantity is known in this state [8], the mean or even variance. Again in the limit q→1the above function will produce the exponential and Gaussian forms respectively. In Tsallis's q-statistics there are many others q-parameters, but generally a triplet is significantly more important [7]. It consists to the sensitivity to the initial condition q-parameter (q_{sens}), the q-relaxation of observable's values parameter (q_{relax}) and the above introduced q-statistics parameter (q_{stat}) providing the relation $q_{sensitive} \le q_{statistics} \le q_{relaxation}$is hold for a stable distribution state. The sensitivity q-parameter is found using relationship $\frac{1}{1-q_{sensitive}} = \frac{1}{\alpha_{\min}} - \frac{1}{\alpha_{\max}}$ where $\alpha_{\min,\max}$ are the singularity point of the multi-fractal power spectrum function of the structure [7]. The relaxation q-parameter is calculated from the q-exponential fitted to the autocorrelation function of the time series. Q-distributions do have important advantage from the physical point of view, because q-parameters are closely related dynamical properties of the system. Specifically form the triplet we know the distance of actual state from the stationary state, the relaxation rate of the observables, their sensitivity to the initial condition and the rate of entropy production etc [7],[8]. Adding to that, for other complex systems, a log periodic term is reported to decorate the q-Gaussian [12]. In the original paper this decoration is reported as of the preferential attachment origin, which seems not to be the case here. Nevertheless, it will be fruitful to analyze its presence, because it is possible to derive from complex stochastic phenomena. The general form of this perturbation is proposed

$$f(X)\sim e_q(-X)\left\{w_0+w_1\cos\left[\frac{2\pi}{\ln(1+\alpha)}\ln\left(1+\frac{X}{Ta}\right)\right]\right\}$$

where X is an energy-like function or variable. Similar complex systems do exhibit discrete scale of invariance behavior [9] therefore testing our data in this sense will shed some light on this aspect. Multi-fractals analysis has been used to learn more the about affine structures for wide class of phenomena in physics, geophysics etc. [10], and it natural to be considered here, but we will extend our view in this aspect even because the calculation of the triplet Tsallis needs the calculation of the multi-fractal spectrum. We follow our previous works [13],[14], to add some more light on traditional procedures of measurement and analyses for such systems.

III. DISTRIBUTIONS AND FRACTAL STRUCTURE OF TIME DATA SERIES FOR SIDE INFLOWS IN KOMAN BASIN

The data used herein are taken from the regular records in the framework of dike and water management. They are the only set of such type of existing data for those systems. Hence, no methods and procedures of concrete measurement are provided so far. Under this limitation, we used them in the original form, that is daily averaged, hourly averaged etc, and in an elaborated form, which consist of only extending the averaging over some successive points. An empiric distribution could be approached to many functions, and there are many different statistical tests to discriminate between them. From physical point of view q-distributions are of a particular interest because the additional information they provide. To adjust the performance of the q-distribution evaluation, we've considered other distribution too. We see that according to the goodness of fit, the lognormal approach is ranked the firsts between many others expected distributions, except the q-Gaussian which is considered separately. We use it to improve our q-Gaussian estimate during the histogram optimization procedure in two ways. Firstly, by comparing the results of standard tests as Kolmogorov-Smirnov, Lilliefor, Anderson-Darling, Jarque-Bera. Secondly, to fix a necessary level of goodness of fit to be kept. So, for daily inflows we obtain a distribution that is fitted very well with a q-Gaussian. Usually it is the lognormal that is nearest possible underlying distribution among many others we tested at the preliminary stage herein, and see that q-Gaussian is better fitted (Fig.1).

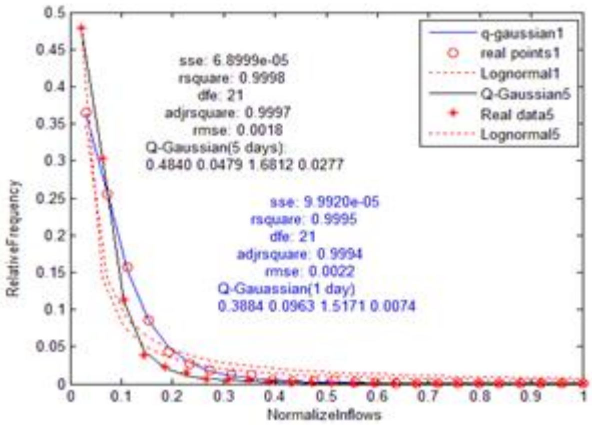

Fig1. Distribution of side inflows (Koman): by circles, 5 days averaging, by asterisks, 1 day averaged

From the q_{stat} value found in the first test of bin size (Scott) one observe that the overall state is far from stationary and variance is undefined. Really from the value q=1.68~5/3 and within the error measure, the variance might be infinite in the statistical sense but accounting for the truncated support of the observables we use the term undefined in both cases, that is when the variance is infinite ($5/3<q_{stat}<2$) or undefined ($2<q_{stat}<3$). Consequently the estimation of the optimal bin size during the discretization step is questionable. Spanning around the bin size found, we see that the optimal choice is near to the mean of the bins width found using Freedman-Diaconics and Scot rules. Here the R^2 indicator is much higher than the one of the lognormal. In this case the triplet Tsallis is obtained [-0.0056 1.6812 2.7716]. Reading it, we do not claim the system to be non-sensitive to the initial condition as expected from the $q_{sensitive}$~0, because it is estimated only formally and include high uncertainty, practically being incredible. In the other side, the statistical parameter q_{stat} is evaluated with relevant accuracy, therefore we conclude that this system is far from the stationary state, and perhaps the variance is undetermined as $q_{stat}>5/3$. For a short stop here, we can say that those series are not appropriate to deal with common statistical analysis or forecasting. This aspect is reinforced from the irregular and self-cut power spectrum curve (Fig. 2)

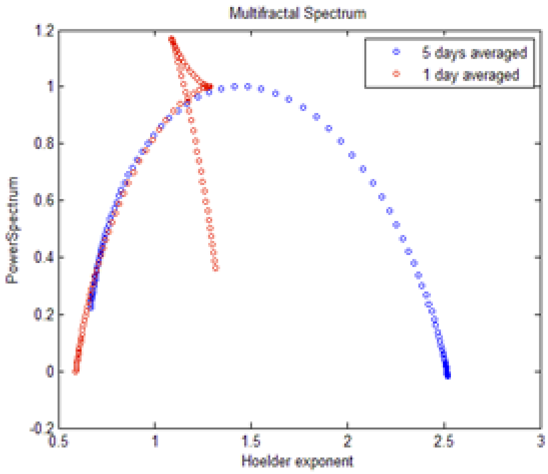

Fig2. Multi-fractal power spectrum. Red markers, original series (1 day). Blue markers, 5 days averaged.

that make impossible the correct evaluation of the q-sensitive parameter. We next consider a child series produced from the original ones by averaging over some few successive points $x_j = \frac{1}{k}\sum_{j}^{j+k} x_i$. In this case, when selecting the

averaging over 5 points, we see that power spectrum is regular, concave, near to a parabolic shape, letting us to calculate the q_{sens} with good accuracy. Hence the triplet now is obtained at [0.3691 1.5171 1.6558]. Moreover, now we reached a *"physical state"* as the relation $q_{sensitive} \le q_{statistics} \le q_{relaxation}$ is clearly fulfilled. Hence the system is moderately sensitive to the initial condition (and so is its entropy production) as q_{sens}<<1. Accordingly this can be viewed as near to minimal entropy production, hence a (locally) more stationary state between many out-of equilibrium states is found. From the value of q_{stat}>1we read immediately that the state is far from the stationary but with variance and mean definite because q_{stat}<5/3. Therefore, this series can be analyzed under standard statistical tools. This is to say that only for this series it is possible to make estimation, evaluation and prediction with known certainty. Next we considered the functional nature of the "perturbation" term over q-Gaussian by fitting the report r_{qg-e} of the values of the q-Gaussian fitted and real distribution. We do compare the goodness of fit of normal q-Gaussians and of the modified q-Gaussian with a log periodic term as mentioned above. One observe that this approach does not work for the original series, but again in the case of the 5-days averaged series there are traces of a log periodic term on the report r_{qg-e}. We've tested this last behavior using as variable *X=inflows* and *X=square (inflows)*, to account for the energy-like observable in *f(X)* above, according to the original paper notification [12]. The log periodic term seems to be present for both choices, but it's difficult to confirm the best because of the high volatile values of r_{qg-e} (Fig.3).

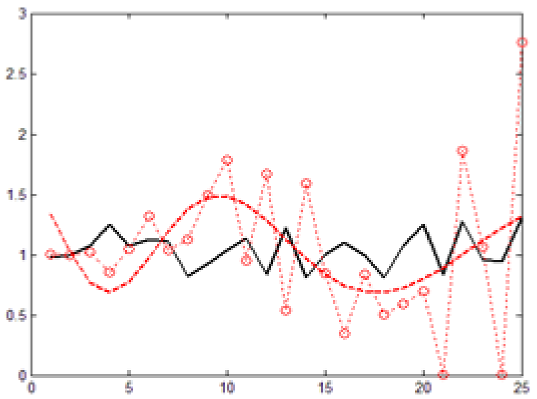

Fig3. Perturbation of q-Gaussian. possible log-periodic shape: With markers, the 5 days averaging, By red line, inflows, by black line, squared inflows

The data recorded in shorter time, respectively the side inflows averaged each hour, are expected to give other important information for system behavior. Those data series consist on the records for the first semester of 2011 when the hydrological system in Drin cascade has been rich in dynamics. The series of data for the hourly averaged inflows is to spurious so no clear distribution is observed. Local fluctuation might drive the system away from a stable process. The multi-fractal spectrum is nor smooth or concave. Using smoother data series as mentioned above, in the case of the averaging inflows for 3 hours, we observe that the distribution is somewhat stable but the goodness of fit for q-Gaussian is low, near to 0.75, therefore not appropriate. Modified q-Gaussian is found better fitted but again fairly, at R^2~0.79. Larger averaging time interval improve slowly the fit (Fig.4)

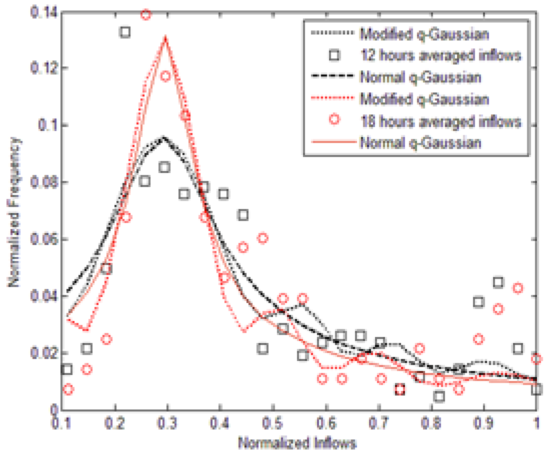

Fig.4. Pseudo-distributions of total inflows. Red color, 12 hours averaged, black color, 18 hours averages.

but decreasing number of available points limit further smoothing. For full water inflows, it seems to be a reasonably behavior accounting for the juxtaposition of hydrological events with human activities on working regime in Fierza lake. Interestingly the log periodic decoration of the q-Gaussian become apparently important, but nevertheless, the primary fit remains not good enough for quantitative analysis. We archive it for now as an interesting "signal" for further consideration. Another interesting data set comes from the series of records for side inflows of waters. Here the fit is found very good characterized by R-squared~0.998, and the triplet is found [0.2626* 1.2580 3.4777] providing that q_{sens}* is badly estimated as multi-fractal spectrum is a complicated shape.

From q_{stat} ~1.258 (<5/3) we see that system is in a variance known state. The relaxation rate is found far away from exponential form as q_{relax}>>1. For 3 hours averaged side inflows, the triplet is found at [-1.0476 1.2356 2.2099]. The overall state is distanced from the equilibrium but the variance and the mean are defined and hence, this series is appropriate for standard statistical and dynamical physics analysis. Up now, the normal q-Gaussians were found better fitted than the modified one, so decoration has not been analyzed. For 6 hours averaged inflows the triplet is obtained at [-0.3520* 1.2624 2.2821], again with q_{sens} not precisely estimated. For 12 hours averaged side inflows we obtain the Tsallis triplet [-0.0259 1.3777 2.6623] that is the state is more "exited" than the one of 3 hour-averaged data. The multi-fractal spectrum function is concave (Fig 5)

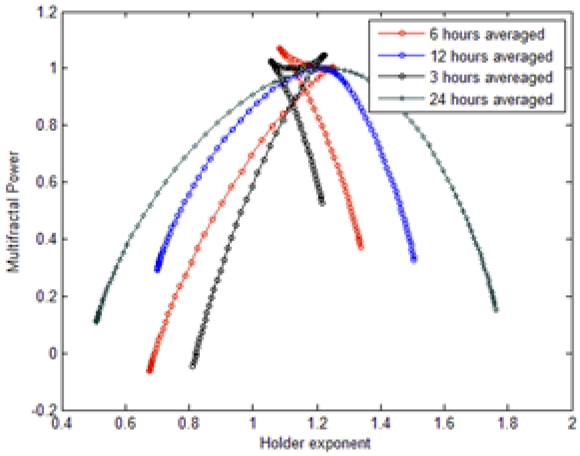

Fig 5. Multi-fractal spectrum for different series as by averaging time.

so all three q-parameters are known with good accuracy, and moreover the basic important relation $q_{sensitive} \leq q_{statistics} \leq q_{relaxation}$ is fulfilled. Therefore this choice belongs to a normal physical system. But as the q_{sens} become more accurate enlarging the averaging time interval, the relaxation parameter q_{relax} is found more uncertain. We obtain that around 3-4 hours of averaging time the state is more stationary among many non-stationary ones, and optimally physical. Next we observe that q_{stat} jumps over 5/3 for averaging time more than 15 hours. For an averaging time 24 hours, the modified q-Gaussian is better fitted to the real distribution, but q_{stat} in this case is in the upper limit of stable distributions (Fig.6).

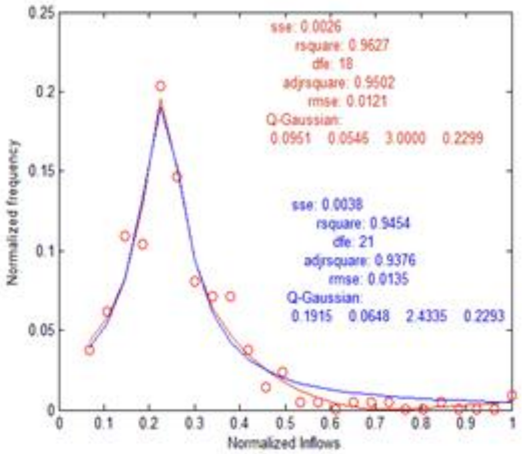

Fig 6. Distributions for 24 hours averaged inflows: Blue line, q-Gaussian, red line, the modified q-Gaussian.

Normal q-Gaussian showed the triplet [0.5380 2.4335 2.6987]. Here the physical relationship among q-parameters is fulfilled and the accuracy of q_{relax} is admissible, even not high. Estimation of q_{stat} showed moderate uncertainty (20% with 95% confidence) whereas q_{sens} is determined with higher certainty providing the multi fractal spectrum is regular concave and hence singularities extrapolated very well. In this case we can say that the sensitivity from initial condition on this system is moderate, and so is the entropy production, that is the rate of the movement toward a more (nearby) stationary state. The variance now is indefinite therefore it is not possible to use those data for standard statistics that routinely uses the variance. But the decoration of the q-Gaussian fitted to the empiric distribution qualitatively shows a log periodic shape (Fig 7). Using as energetic-like variable the marginal inflows $X=Inflows-\mu$ where μ is found from the q-Gaussian fit, we observe that the log periodic over its square value (X^2) is better fitted with the rapport as showed by the black line on the graph (Fig 7).

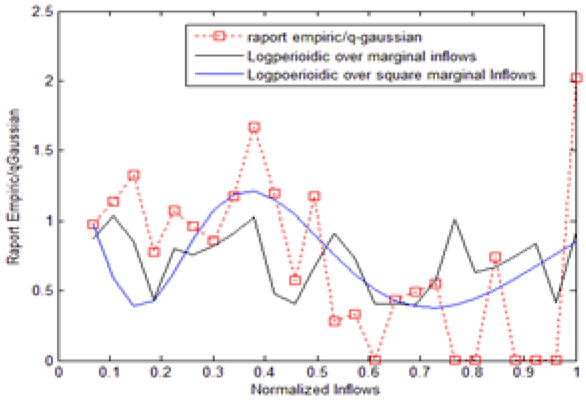

Fig. 7. The Log periodic decoration of the q-Gaussian fitted

It is quite easy to distinguish the missing quantitative accordance hereby, but a merged effect of complex stochastic processes could be argued to be present, with a similar effect as the self-organized regimes by preferential attachment mechanism claimed on for other systems [12]. We are not trying to explore arguments on the support of similar assumption, but at this point we can argue that signals of log periodic decoration over a q-Gaussian are likely to be found hereof.

Questionable discussion may rise from here and we are not dealing with them for once leaving them for another consideration. We've found elsewhere that the series of time data records for side inflows for the lake of Koman in this period do exhibit log periodic behavior on its trend [13] hence, traces of self-organization behavior with discrete scale invariance are present. Accordingly, another exhibition of such behavior can be considered as an indicator of hierarchic structural origination in particular condition.

IV. CONCLUSIONS

The q-statistics based on the Tsallis distributions was successfully applied to improve the statistical and dynamics analysis for a particular hydrological system. So, the distributions of the daily and hourly averaged inflows for the lake of Koman, at Drin River, Albania have been better and fully described with normal and decorated q-Gaussians. Daily and hourly data series are found in a far from stationary and usually in variance undefined state, and so the standard statistical analysis may leads to fake conclusions. Introducing child series produced by data averaged over few successive data point, more stationary states in clear physical condition were found and studied. Such series obtained from the averaged side inflows for 5 or more days and 3-4 hours were found more physical in the sense that the distribution are stable, with definite variance and physical relationship between q-parameters of the Tsallis triplets. In some cases we obtain a log periodic decoration over q-Gaussians, which could be function of the corrected inflow or squared one as energy-like parameter. This behavior could be related to the complex stochastic processes in the presence of local fluctuations and possible hierarchical level of organization in specific conditions. Evaluating the rate of entropy production in specific choices called "physical state" of the system, will help researcher to find more on the dynamics of the system. In the technical point of view, we propose the complex system approach when dealing with complicated systems as hydro-electrical environments. Hereof, it is recommended to use the *appropriate physical state date series* for management purposes.

REFERENCES

[1] Pilon, P. J. and Harvey, K. D.: Consolidated frequency analysis, Reference manual, Environment Canada, Ottawa, Canada, 1994.
[2] Didier Sornette. *Probability Distributions in Complex Systems*. Encyclopedia of Complexity and Systems Science. 2009, p 7009-7024
[3]. Benson, M. A.: Uniform flood frequency estimation methods for federal agencies, Water Resour. Res., 4(5), 891–908, 1968.
[4] C. Tsallis, *Possible generalization of Boltzmann-Gibbs statistics*, J. Stat. Phys. 52, 479-487 (1988)
[5] R. Donner, S. Barbosa J. Kurths and N. Marwan. *Understanding the Earth as a Complex System – recent advances in data analysis and modeling in Earth sciences*. Eur. Phys. J. Special Topics 174, 1–9 (2009)
[6] B. Sivakumar and V. P. Singh. *Hydrologic system complexity and nonlinear dynamic concepts for a catchment classification framework*. Hydrol. Earth Syst. Sci., 16, 4119–4131, 2012
[7]Constantino Tsallis *Nonadditive entropy and nonextensive statistical mechanics - An overview after 20 years* Brazilian Journal of Physics, vol. 39, no. 2A, August, 2009 337
[8] S. Umarov, C. Tsallis and S. Steinberg, *On a q-central limit theorem consistent with nonextensive statistical mechanics,* Milan J. Math. 76, 307 (2008).
[9] D. Sornette, "*Predictability of catastrophic events: material rupture, earthquakes, turbulence, financial crashes and human birth*", Proceedings of the National Academy of Sciences USA, V99 SUPP1:2522-2529 (2002)
[10] Wei-Xing Zhou, and Didier Sornette. *Discrete scale invariance in fractals and multifractal measures*. arXiv:cond-mat/0408600v1
[11] Jarosław Kwapieńa Stanisław Drożdża *Physical approach to complex systems*, Physics Reports 515 (2012) 115–226
[12]. G. Wilk, Z.W Lodarczyk. *Tsallis distribution with complex nonextensivity parameter q*. arXiv:1403.3263v2 [cond-mat.stat-mech], 18 Mar 2014.
[13] Dode Prenga, Margarita Ifti. *Complex System Approach to the Study of Albanian Drin River Inflows*. Submited to Hydrological Sciences Journal, September 2014.
[14] Dode Prenga, Teodor Kota. *Study of aerosol time series data using Tsallis statistics and fractal analyzes* . International Journal of Geology, Agriculture and Environmental Sciences Volume 02, Issue 3, May-June 2014. p 34-37