THE USE OF PARAMETRIC DISTRIBUTIONS AND EMPIRIC DECOMPOSITION TO ANALYSE STOCHASTIC SYSTEMS AND NONLINEAR SIGNALS

Sandër Kovaçi[1], Erarda Vuka[2], Dodë Prenga[3]

[1] Department of Mathematical Engineering, Polytechnic University of Tirana

[2]Department of Informatics, Faculty of Natural Sciences, University of Tirana.

[3]Department of Physics, Faculty of Natural Sciences, University of Tirana

e-mail: s_kovaci@yahoo.com

Abstract: In this research we consider combined techniques on some signal analysis using empiric mode decomposition and parametric distributions analysis. The IMF mode whose amplitudes exhibit a stable q-distribution is considered as physical and the dominant processes governing it are identified by the analysis of the nature of the distribution. Next, a specifically de-noised and mostly unique process-type signal is reconstructed. The procedure is applied directly in real hydrological data series from our environments whereof some practical result has been concluded.

Key words: *empiric mode decomposition, signal analysis, q - distribution.*

INTRODUCTION

Empirical Mode decomposition technique.

The stochastic processes analysis stand in the basis of the study of many real systems. Empirical Mode Decomposition introduced by Huang, is an alternative analysis aside Fourier or wavelet techniques. It assume that a nonlinear signal could be written in the very simple form

$$f(t) = \sum_{i=1}^{N} IM_i(t) + r_N(t) \qquad (1)$$

where the IM-s are the intrinsic modes and *r* are residuals as detailed in [1] and improved versions presented in [2],[3]. It has been successfully applied in nonlinear signal analysis as in [4], alternative of signal analysis as in [5] or as de-noising alternative reported in [6] etc. but there is no theoretical proof for EMD decomposition[3]. The Huang algorithm consist in the following: Compute lower and upper envelopes from interpolations between extremes; subtract mean envelope from signal; iterate until mean envelope = 0 and number of extreme = number of (zero-crossings) ± 1; subtract the obtained *Intrinsic Mode Function* (IMF) from signal and iterate on residual until a stopping criteria is met. Not all IMF are physical and therefore their concrete interpretation is of great importance. Adding to that the Hilbert transformation is applied for the IMF [1],[2] resulting in the instantaneous frequency (IF) series, hence there will be twice IMF series to be considered. To challenge problems related to the empiric basis, white noise-added algorithm (EEMD) or adaptive noise-added (CEEMDAN) as given in [7] as [8] were introduced. Those approaches consider the fact that when dealing with a concrete set of data one must consider it as a realization and therefore it behave as a measurement with error. Assuming that the error has normal distribution, one prose to add white noise so the cancel successively. Elsewhere, higher order statistics has been used in de-noising process [8]. Based on this idea and following the idea posed recently in [9], here we use non Gaussian noise in some real data analysis with EMD technique. Without a proof we expect that depending on specifics of the concrete time series the nature of the noise will be selected carefully and in this context this could not be generalization. It is worth to brief in the following some elements of parametric distribution introduced by Tsallis, multifractal, and self-organization behavior etc.

Parametrical Distribution.

Following the bringing into play for q-exponentials $\exp_q(y) = \left[1 + (1-q)y\right]^{\frac{1}{1-q}}$ as distribution for correlated systems introduced in 1998 in [10], a special such case $y=x^2$ called q-Gaussian has been found the attractor of the distribution for the sum of correlated variables and a q-Central Limit Theorem is formulated in [11]. This last has the form

$$P_{q-Gaussian}(x-<x>) = \alpha\left[1-(1-q)\beta(x-<x>)^2\right]^{\frac{1}{1-q}} \qquad (2)$$

It result that the q-parameter herein report the stability of distribution as seen from the formula $\alpha_{Levy} \sim \frac{3-q}{q-1} \Rightarrow 1 < q \le 3$. The stable distribution will have definite variance given by the relation

$$\left[\sigma_q = \frac{1}{(5-3q)\beta}\begin{cases} \ge 0 \\ < \infty \end{cases}\right] \Rightarrow 1 < q < \frac{5}{3} \qquad (3)$$

but Tsallis et al [10],[11] have shown that the distribution is defined for $q \in [1,3)$. Q-Gaussians belong to heavy tails parametric distribution and are able to interpret many properties in real systems. As it recover the classical Gaussian in the limit q=1, its use has been extended in several applications even in formal absence of the proof.

Fractal and critical behaviour.

The multifractal behavior is quite common in many real data series. Detailed presentation about those structures and their occurrence is given in [13]. Generally the fractal properties are shown in the relationship around a point $y(x+a)-y(a)=a^{h(x)}$ where h is the singularity exponent and a is the scaling parameter, or by using affine structure $y(x)\to k^{H}y(kx)$ where H is the Hurst exponent [13]. The structure is studied using Dentrended Fluctuation Analysis by the calculation of multifractal power spectrum according to the relations

$$f(\alpha)=q\alpha-\tau(q) \text{ where } \quad \alpha=\frac{\partial\tau}{\partial q};\tau(q)=qh(q)-1 \qquad (4)$$

where $h(q)$ is the generalized Hurst exponent (Holder) and q is the order of fluctuation. The verification of chaotic multi-scale and multi-fractal dynamics underlying time series observations is found of the high theoretical interest because the coupling of q-statistics with fractal dynamics as detailed in [14]. Therefore we will incorporate such aspect in our algorithm. A specific situation of the scaling is the case of discrete scale of invariance (DSI) according to the relation $y(\lambda x)=\lambda^{\alpha}y(x)$ for some discrete value of λ. The time behaviour in this case is found log periodic function [15]

$$I=I_0+a*(x-x_c)^m+bx-x_c{}^m\cos[\omega\log(x-x_c)+\varphi]+\ldots \qquad (5)$$

We can expect that in such case the dynamics on the data series could affect the analysis therefore we propose to consider this case separately.

METHOD AND ALGORITHM

The herein strategy of the EMD analysis is proposed to consider harmonization of the elements. Hence, we start our algorithm with the classification of some relevant properties of the signal according to the stability of the distribution, multifractal properties and critical behavior. Initially we read the stability of the distribution in $q_{stationary}$ parameter form the fit of empiric distribution (data frequencies by bins) using an ad-hoc script based on Nonlinear Least Squares method and an optimization of bins width e routine as described in [17] following appropriate reference application. Next, we recognize the dynamical rate by identification of Tsallis $q_{sensitive}$ calculated by the formula given in [11]

$$q_{sens}=\left(\left[1-\frac{\lambda q_{sen}}{\lambda_1}+\frac{\lambda q_{sen}}{\lambda_1}e^{(1-q_{sen})\lambda_1 t}\right]^{\frac{1}{1-qe_{sns}}}\right)=\left[\left(\frac{1}{\alpha_{\min}}-\frac{1}{\alpha_{\max}}\right)^{-1}-1\right].$$

Here a_{min} and a_{max} are the zero point of the multi fractal power spectrum f(a)=0. If q_{sens}~1 the process is assumed as highly dynamic and therefore the application of white noise is expected to work better. We estimate the multifractal structure by running a script based on Multi Fractal Detrended Fluctuation Analysis (MFDFA) [16] and use it to classify in the algorithm the remaining procedure in EMD routine. If q>2 which consist to q-variance indefinite we propose to fit q-lognormal proposed in [11] and applied for our system in [17] using the formula

$$p_{q_1,q_2}(x) = a\frac{1}{x^{q_2}}\left[1 - b(1-q_2)\left(\left(\frac{x^{1-q_1}-1}{1-q_1}\right) - m\right)^2\right]^{\frac{1}{1-q_2}} .$$

If this last is fitted better, the multiplicative property is more characteristic and therefore we read the distance from the stationary state in parameter q_2. Having estimated q value we simulate q-Gaussian noise with zero q-mean and variance equal to a percentile of the variance in original series. The process is repeated many times and the averaged noise is evaluated and is added to the original by directly inserting the result in the EEMD algorithm given in [3]. To compare we analyze the EMD of the original signal and when a classical Gaussian noise is added. Finally, we test the critical behavior of the signal by testing the log periodic presence according to original function proposed in [12] and the relevance of it following to the discussion in [17] with this new form

$$P(t) = a + b(t-t_c)^m + c(t-t_c)^m\cos\left(\omega\left[\frac{(t-t_c)^{1-q}-1}{1-q}\right]^{1-q} + \varphi\right) + d(t-t_c)^m\cos\left(2\omega\left[\frac{(t-t_c)^{1-q}-1}{1-q}\right]^{1-q} + \varphi\right)$$

where t_c is the critical time. An ad-hoc calculation technique based in genetic algorithm is used in this step. Again we expect that in this case, the q- Gaussian noise could complicate the efficiency of the analysis because of criticality but analytically this is difficult to be proofed due to the empirical basis of the EMD. The next stage is the estimation of the distribution in IMF itself following the idea [9] for the identification of the measurable signals. Finally the trend of the last IMF is identified to get known the underlying trend of the overall average signal. All those steps were included in standard EMD, EEMD or VEMD algorithms.

DATA ANALYSIS AND RESULTS

Stochastic signal with multifractal structure and low dynamic rate. Hydrological data from Fierza basin, Albania

We consider as case study the daily and hourly side inflows in Fierza following our previous works [18] where we've found that the distribution are unstable or variance undefined, characterized by a low dynamical rate hence appropriate as specific for our consideration. In the analysis of the Fierza's daily or hourly inflows, we obtain that the signals behave mostly as being drawn from a q-Gaussian distribution. There is a weak significance of log-periodic behavior because q=1.12>>1 in the q-log periodic fit and smooth multifractal spectrum (Fig1). Daily inflows consist in an unstable distribution process but with definite q-variance as q_{stac}=1.9>5/3 and with low rate dynamics as read from q_{sens}=0.1483. As mentioned above this could be a good candidate where the nature of the noise could affect outcome. Hereof, using standard EMD we evaluate 11 IMF and evaluated the stationary parameter form the q-Gaussians fitted q= **1.0067 1.0002 1.1802** 2.2017 2.2618 1.002 1.0663 1.0000 1.0000 2.9665. It result that the IMF 2-7 represent processes that do have distribution even nonstable. Accordingly, the local (instantaneous) frequency conjectured to those signal could be classified as measurable quantity (Fig2). After calculating it and adding a Gaussian noise we obtain the stationary parameter form q-Gaussians fitted to its distribution is found that frequency of IMF-1,7,8 are not measurable because the q parameter is out of the range]1,3[of meaningful distribution as read from the q_{GN} vector [3.0 2.2571 2.0974 1.0696 1.0000 1.1674 3.0000 3.0000 2.1054]. In the case of the q-Gaussian noise added we see that only one IMF has not identified average frequency (q-3) as seen from the vector q_{qGN}= [2.3924 2.4764 1.9776 1.1350 1.0000 1.0001 1.3423 3.0000]. In the simulation of the noise we set q-1.6, because the q read from the distribution of the amplitude of the original series has the value out of definite range for q-variance (q-1.9>5/5) . We observe that in this case the number of IMF is greater (here n=12). Even it is hard to generalise for similar cases, we obtain that in this case the use of the q-Gaussian noise has improved the EMD analyses and result (Fig3).

Mixed Biophysical with complex multifractal structure-ECGexample

In another application of our proposal we consider a highly volatile and irregular time series as ECG in arrhythmia. Theoretically there will be at least 5 stage on ECG output therefore those series contain a soup of processes each of them affecting one amplitude behaviour. Standard techniques in those cases apply typical filters to make it readable for physicians, and here we are not dealing with them. We consider those series as a specific item to learn the outcome of our propose by assumption that a common signal could consist more than one process each of one having the same importance. Herein, using a MDFA calculation, we observe that the multifratcal structure is not smooth (Fig4). Adding to this fact estimate the rate of the dynamics there is slight log-periodic behaviour on the trend in the sense of DSI presence. First, we analyzed the signal using noise free EMD, and we observe that the number of the IMF is relatively high, and the amplitudes of the IMF have not stable distribution. By the evaluation of the correlation of original signal with the sum of IMF, we obtain a high value, near to 1. Using standard Gaussian noise as usually in EEMD we observe a slightly lower correlation, other remaining unchanged. Next, we add a q-Gaussian noise to the original signal and perform the EMD analysis. In this case we have missing q-reference from q-Gaussian distribution. We obtain that the correlation of the sum of IMF with original signal in this case is lower than if white noise is added as in EEMD (Fig5). Moreover, in this complex signal we see that if there will be improvement in signal (IMF) distribution, this is not the case of corresponding instantaneous frequency in the sense that if it happen, it is occasionally and not a rule as seen on the table for $q_{stationary}$ and r-squared for each signal in tab (fig5). In those applications we estimate that Gaussian noise could be best solution if we prefer to filter noses or perhaps other techniques could be more effective as varational empirical mode decomposition that is not included in this survey.

CONLUSIONS

Study of the stochastic nonlinear signals will be improved if we use empirical mode decompositions method assisted with specific noises produced using q-Gaussian distribution and considering multifactal structure or other fractal properties. A careful estimation of those last will help in the qualification of the nature of the noise. In our example we obtained that for signals with regular multifractal structure, the injection a q-Gaussian noise improve the EMD analysis, but more effort should be made in the mathematical arguments and in the choice of q-parameter. On the other side, if the signal consists in a mixed set of signals or if multifactal power spectrum's not smooth, this modification does not improve the analysis or can worse some aspects.

References

[1] Patrick Flandrin, Fellow, IEEE,. Gabriel Rilling and Paulo Goncalves. Empirical Mode Decomposition as a Filter Bank IEEE Signal Processing Letters, 2003

[2] Norden Eh Huang, Samuel S. Shen Hilbert-Huang Transform and Its Applications. World Scientific, 2005 - Decomposition (Mathematics)

[3]Gabriel Rilling and Patrick Flandrin, Fellow, IEEE. One or Two Frequencies? The Empirical Mode Decomposition Answers. IEEE Transactions on Signal Processing. Vol. 56, No. 1, January 2008

[4] Agnieszka Kitlas Golińska1, Poincar´e Plots in Analysis of Selected Biomedical Signals Studies in Logic, Grammar and Rhetoric 35 (48) 2013

[5] Guy P. Nason and Rainer von Sachs. Wavelets in time series analysis. Phil. Trans. R. Soc. Lond. A (1999)

[6] Yi-Ming Fang, Hai-Lin Feng, Jian Li and Guang-Hui Li Stress Wave Signal Denoising Using Ensemble Empirical Mode Decomposition and an Instantaneous Half Period Model. Sensors **2011**, 11, 7554-7567

[7] Marcelo A. Colominas, Gaston Schlotthauer, Maria E. Torres, Patrick flandrin. Noise-assisted emd methods in action. Advances in adaptive data analysis. Vol. 4, no. 4 (2012) 1250025

[8] George Tsolis and Thomas D. Xenos Signal Denoising Using Empirical Mode Decomposition and Higher Order Statistics International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 4, No. 2, June, 2011 91

[9] ErardaVuka, SanderKovaci, DodePrenga. Empiric Mode Decomposition Algorithm and analysis as alternative study of stochastic nonlinear process. ICREA, Shkoder 23 October 2015.

[10] Tsallis C., 2009, Non-additive entropy and non-extensive statistical mechanics - An overview after 20 years, Brazilian Journal of Physics 39, 2A, 337.

[11] S. Umarov, C. Tsallis and S. Steinberg, On a q-central limit theorem consistent with no extensive statistical mechanics, Milan J. Math. 76, 307 (2008).

[12] D. Sornette. Critical market crashes. Physics Reports 378 (2003) 1–98.

[13] Hiroki Suyari Mathematical structure derived from Tsallis entropy. Department of Information and Image Sciences, Chiba University, 263-8522, Japan

[14] Mike West. Kernel density estimation and marginalization consistency.. Biometrika (1991), 78, 2, pp. 421-5

[15] Jan W. Kantelhardt Fractal and Multifractal Time Series. arXiv:0804.0747v1 [physics.data-an] 4 Apr 2008

[16] Espen . F. Ihlen. Introduction to multifractal detrended fluctuation analysis in Matlab. Frontiers in Physiology. Fractal Physiology June 2012.Volume 3

[17] D. Prenga, M. Ifti, J.Fejzaj. Complexity Methods Used in the Study of Some Real Systems with Weak Characteristic Properties. Balkan Union of Physics Conference. Istanbul, 2015

[18] Sander Kovaci, Dode Prenga, Margarita Ifti. Specifics of the distribution and dynamics for side inflows in the Drin River basin, Albania. International Journal of Engineering and Technical Research. 2321-0869, Volume-03, Issue-01, March 2013

Appendix. Figures

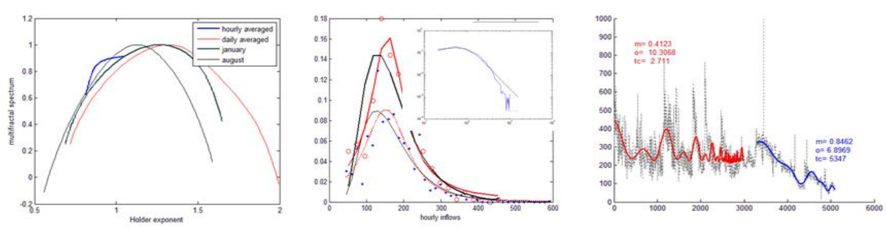


Fig.1. Multifactal spectrum for hourly inflows and q-distributions (midle picture) for different averaging time series of inflows. Red line correspond to the 3 hour averaged. In corner, q_1q_2-lognormal shows the fit for hourly (original serie). Right picture shows mixed weak log-periodic trend.

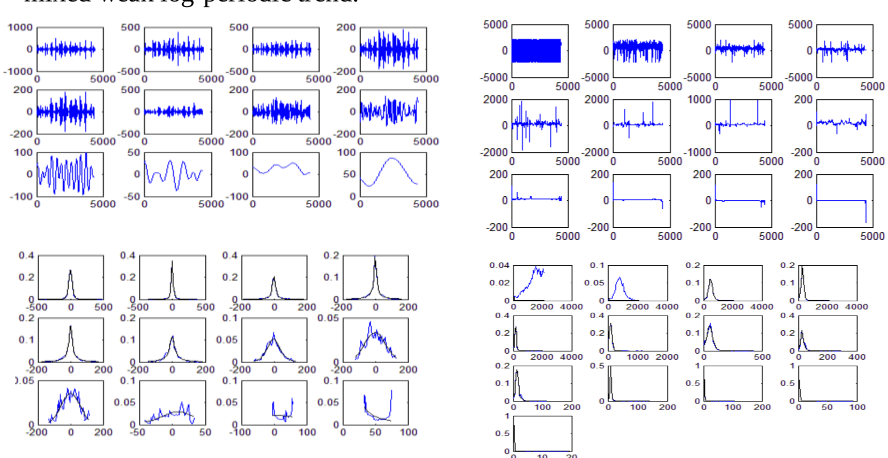
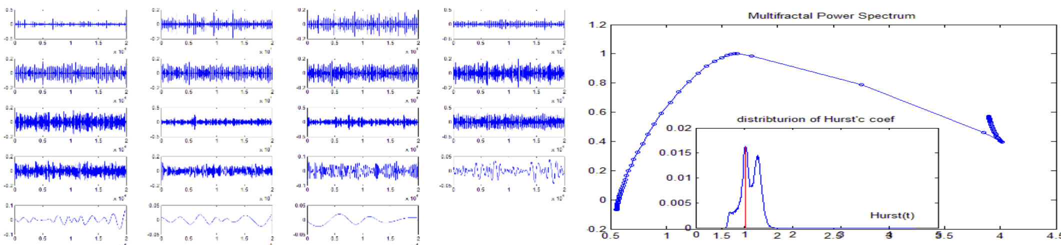
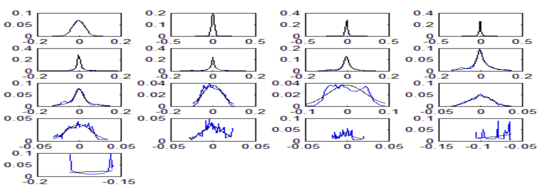


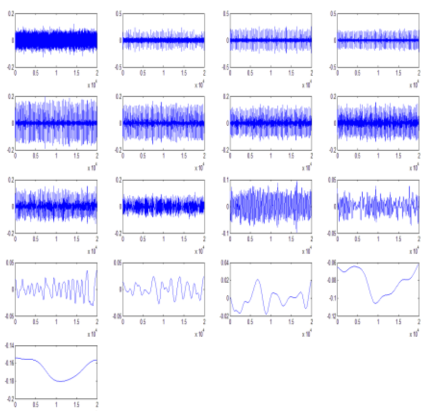
Fig2. IMF and q-distribution. Left, original signal. Right, instantaneous frequencies. Q-Gaussian assisted EMD analysis





Q	R2
1.4686	0.9992
1.0000	0.9927
2.0927	0.9961
2.1630	0.9950
2.4883	0.9943
2.9999	0.9953
2.9999	0.9678
3.0000	0.9552
2.9096	0.9398
1.5659	0.9870
1.5452	0.9763
3.0000	0.9297
1.0005	0.9743
1.0000	0.9441
1.0000	0.9471
1.8866	0.9337
1.3254	0.8608
1.0009	0.5309
2.1507	0.0506

Fig.4. Noise free ECG signal and its IMF amplitude distribution (left pannel). In the right the multifractal power spectrum graph. In the table, the q-Gaussian parameter.



IMF (w.noise)		Freq.W.noise		IMF.Q-noise		Freq. q-noise	
Q	R2	Q	R2	Q	R2	Q	R2
1.0025	0.9986	1.0000	0.9998	1.0016	0.9987	1.4686	0.9992
1.0722	0.9987	1.0395	0.9984	1.1377	0.9995	1.0000	0.9927
1.2973	0.9977	1.9731	0.9978	1.1460	0.9973	2.0927	0.9961
1.6854	0.9966	1.8949	0.9976	1.5870	0.9957	2.1630	0.9950
1.9790	0.9959	2.0528	0.9959	1.9752	0.9946	2.4883	0.9943
2.3787	0.9961	2.4442	0.9975	2.4549	0.9959	2.9999	0.9953
2.1581	0.9869	2.4443	0.9923	2.3229	0.9849	2.9999	0.9678
2.4860	0.9771	2.3153	0.9846	2.5061	0.9608	3.0000	0.9552
2.1482	0.9505	2.2895	0.9633	2.4192	0.9399	2.9096	0.9398
1.0000	0.8695	2.5025	0.9387	1.0000	0.6741	1.5659	0.9870
1.0000	0.7706	1.0000	0.5079	1.0000	0.4256	1.5452	0.9763
1.0441	0.9475	1.0000	0.4811	1.3068	0.8959	3.0000	0.9297
1.0001	0.8493	1.0000	0.9325	1.0000	0.7447	1.0005	0.9743
1.6273	0.5558	1.0000	0.7930	1.3510	0.8037	1.0000	0.9441
2.9948	0.3372	1.3944	0.8724	2.9999	0.6828	1.0000	0.9471
2.0288	0.0116	1.0001	0.4658	2.2509	0.0223	1.8866	0.9337
1.8217	-0.0413	1.9992	0.0359	2.0158	-0.0426	1.3254	0.8608
						1.0009	0.5309
						2.1507	0.0506

Fig.5. IMF in q-noise injected for ECG signal (left) and the table of q-parameters from q-Gaussians and R- squared.