Proceedings of the Pakistan Academy of Sciences 51 (4): 345-353 (2014)

ISSN: 0377 - 2969 (print), 2306 - 1448 (online)

Pakistan Academy of Sciences

Research Article

Solar Flares Data Analysis on application of Probability Distributions and Fractal Dimensions and a comparative analysis of North-South Hemispheric Solar Flares Data Behavior

Danish Hassan[1], Shaheen Abbas[1], M.R.K Ansari[2], and Bulbul Jan[3]

[1]Mathematical Sciences Research Centre Federal Urdu University of Arts, Sciences and Technology, Karachi, Pakistan
[2]Department of Mathematics Sir Syed Engineering University, Karachi, Pakistan
[3]Institute of Space and Planetary Astrophysics, University of Karachi, Pakistan

Abstract: Since Solar Flares are associated with the sunspots. Solar flares have different length, duration and peakness in different intervals as sunspots have. The Solar Flares index data starting from 1966 to 2008 (Monthly Flare Index Cycles "20, 21, 22, 23") in both North and South hemisphere separately along with total Solar Flares data. This paper compared the North, South and total hemispheric Solar Flares data in perspective of probability distributions such as Gamma, Log-Gamma and Chi-square which tested by Kolmogrove-Smirnov D-test and then their persistency by using fractal dimension. Further we estimate the Hurst exponent with the help of fractal dimension. Fractal dimension shows complex of the data while Hurst exponent represents smoothness of the data. All the tested probability distributions in this paper show persistency and positively correlated. The use of probability distributions allows to represents the uncertainties in the solar flares data. This study will be helpful to verify the long term smoothness of data trends and conclude the results will be present so forecast values will be more accurate.

Keywords: Solar Flares, Gamma Distribution, Log-Gamma Distribution, Chi-Squared Distribution, Fractal Dimension , Hurst Exponent

1. INTRODUCTION

The sun is releasing energy continuously the major types of it are namely solar flares (SF) and coronal mass ejection (CMEs) [15-6]. The explosive event that produces a sudden brightness on the sun is known as SF these flares produces a burst of energetic particles the energy released by the SF is equal to 10^{32} erg [14, 10]. The SF are normally classified by their brightness for example the X-class, M-class and C-class which are the basic types of SF. The X-class SF can produces the major storms on the earth, M-class SF can cause brief radio blackouts at polar latitudes and the C-class SF have no significant consequences [14, 7]. SF occurs near or inside the sunspots but it is not necessary that all the sunspots have the SF so there is a interaction between SF and the sunspots [11, 7].

In this paper we use FD method to observe the complexity of the SF data. Higher the complexity of the data lower will be the smoothness or wise versa. Fractal dimension expresses the space filling property of the data whereas, Hurst exponent represents the smoothness of the data.

The FD is defined by the following expression.

$$\mathrm{FD} = \frac{\text{no of small pieces}}{\text{magnification}} \tag{1.1}$$

FD and H are related as follows

$$F = 2 - H \tag{1.2}$$

FD and H represent the dynamical behavior of the time series data. Particularly, Hurst exponent compares the persistency, and anti persistency and Brownian nature of the time series data.

Received: April 2012; Accepted, September 2014
*Corresponding author: danish10ansari@gmail.com, shaheen838@yahoo.com, rashidka2000@gmail.com

Relation (1.2) in this study will be used to analyze the persistency of SF data.

2. MATERIAL AND METHODS

We collected the SF data from WDC (1966 to 2008). We applied most suitable probability distributions on each cycle of SF data as well as for total data. For this kind of estimations we used statistical software Easy Fit (EF).

The second part of this paper is based on the estimations of FD of each SF cycle with total SF data. We have used Fractalyse 2.4.1 software to obtain the FD. We also obtained the parameter H (Hurst exponent) with the help of FD and analyze the persistency of each cycle for North, South and total SF data.

2.1 Probability Distribution Approach

In most of the experimental analysis of any branch of science one often encounters the problems where the probability distributions are applicable. These probability distributions can be helpful in generating the random numbers [2, 5].

In the term of probability distribution the comparative study of SF is very helpful to observe the change and variation on the earth climate. There is a marked correlation between sunspots cycles and earth climate (rainfall and temperature) [4]. The study of probability distributions of random numbers for climatic parameters generally gives the basic knowledge about the physical processes governing in it.

The KST [Kolmogorov-Smirnov *D*-test] for the check of deviation of normality to assessing the nature of the distribution with the real time data. This tests that whether the statistic

$$D = \max |F(x) - G(x)| \quad (2.1)$$

exceed a critical value in the K-S table or not. Where G(x) is sample cumulative distribution and F(x) is the predetermined cumulative distribution corresponding to a given sample of size n [9]. The normal distribution is well define distribution for continuous variables applied to symmetrically distributed data. Mathematically it can be define as follows:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad (-\infty < x < +\infty) \quad (2.2)$$

Where, σ and μ are the standard deviation and mean respectively of the sample.

The Log-Gamma Distribution (LGD) is define as it is the natural log of variable x which is Gamma distributed, the Log-Gamma Distribution has minimum value 1 when Gamma variable = 0. The LGD is given by

$$f(x) = \frac{[\ln(x-y+1)]^{\alpha-1}(x-y+1)^{-\left(\frac{1+\beta}{\beta}\right)}}{\beta^{\alpha}\Gamma(\alpha)} \quad (2.3\text{ a})$$

where $(1-\beta)^{-\alpha} + \gamma - 1$ is the mean of LGD and $x \geq \gamma$, $\alpha > 0$, $\beta > 0$ [9,2]. Γ and β are the parameters of the distribution. Similarly if the random variable Y follows gamma distribution with parameters α and β , then the likelihood of Y is expressed as,

$$g(y) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} Y^{\alpha-1} e^{-\beta y}, \quad (y \geq 0, \alpha > 0, \beta > 0) \quad (2.3\text{ b})$$

Where, α and β are the shape parameter and scale parameter respectively

$$\Gamma(\alpha) = \int_0^{\infty} t^{\alpha-1} e^{-t} dt \quad (2.3\text{ c})$$

$$E[Y] = \frac{\alpha}{\beta}, \quad \mathrm{Var}(Y) = \frac{\alpha}{\beta^2} \quad (2.3\text{ d})$$

zThe Chi-square Distribution (CSD) which can be approximated by the Normal Distribution (ND) and is given by the following.

$$f(x;n) = \frac{\left(\frac{x}{2}\right)^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{2\Gamma\left(\frac{n}{2}\right)} \quad (2.4)$$

Where the variable $x \geq 0$ in CSD and the positive integer represents the number of degrees of freedom [2]

2.2 Fractal Dimension

Box-counting or box dimension is one of the most widely used dimensions [3]. Its popularity is largely due to its relative ease of mathematical calculation and empirical estimation. The definition goes back at least to the 1930s and it has been variously termed Kolmogorov entropy, entropy dimension,

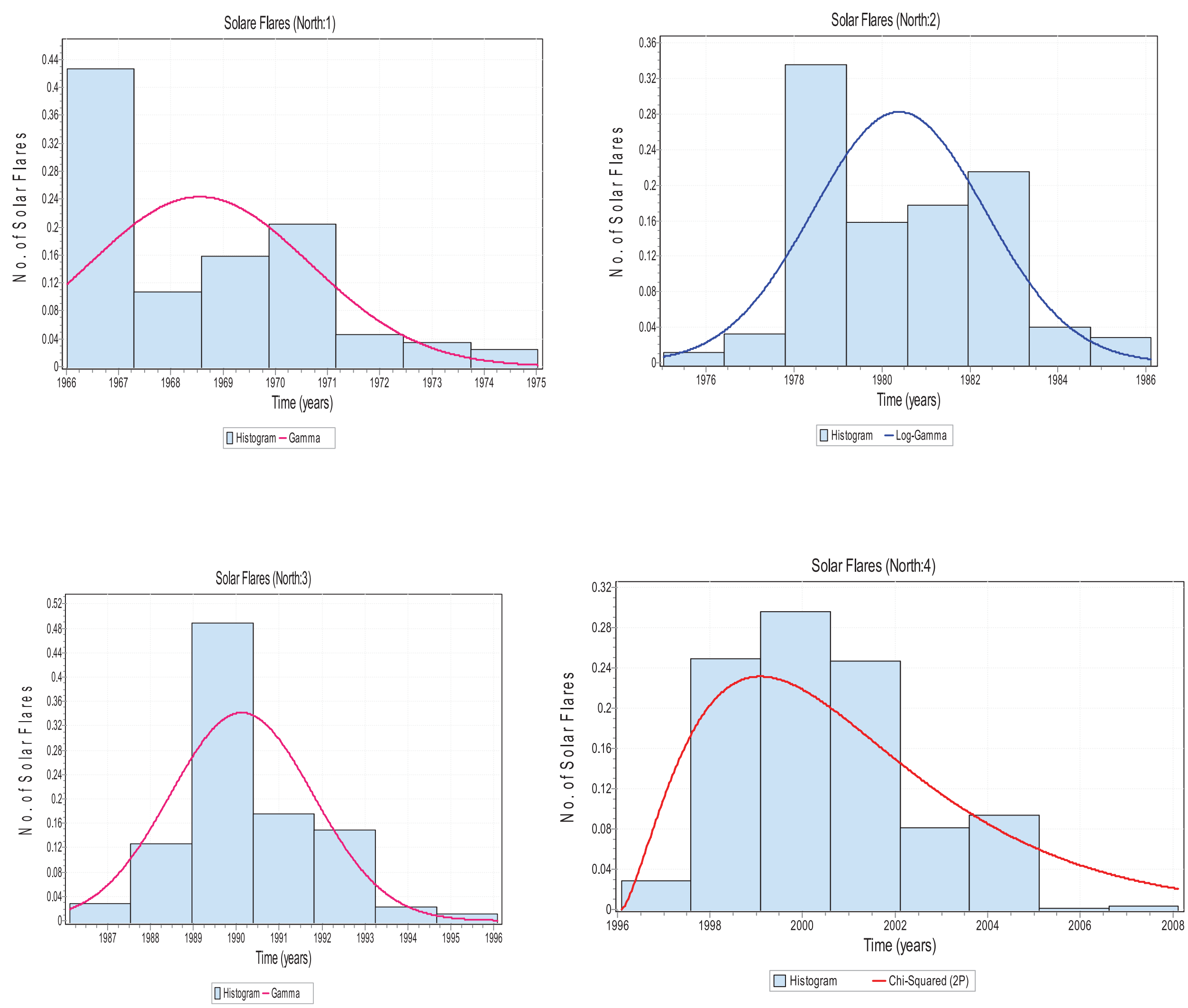

Fig. 1. Probability distributions of 4 cycles of the solar flares North-Hemisphere data.

capacity dimension (a term best avoided in view of potential theoretic associations), metric dimension, logarithmic density and information dimension. We shall always refer to box or box-counting dimension to avoid confusion [3]. This kind of analysis includes more insight into complexity and structure of the system [1]. The FD of self similar object with self similar pieces **r** scaled down by a factor **s** can be express mathematically as.

$$D = \frac{lnr}{lns} \tag{2.5}$$

Fractal dimensions are generally comparison of different numbers that are associated with fractals [8]. The importance of Fractal dimensions is because they can be defined in connection with real world data which can be measured approximately by experiments.

The clouds, trees, feathers, coastlines, neurons networks in the body, dust in the air, the clothes, and the distribution of frequencies, the colors emitted by the sun, and the wrinkled surface of the sea during a storm are attached with the fractal dimensions [1].

Consider the relationship between fractal dimension and Hurst exponent

H	FD	Correlation	Nature of Process
0.5	<1.5	Positive	Persistent
=0.5	=1.5	Zero	Brownian
<0.5	>1.5	Negative	Anti-Persistent

This relationship is summarized by [14] which

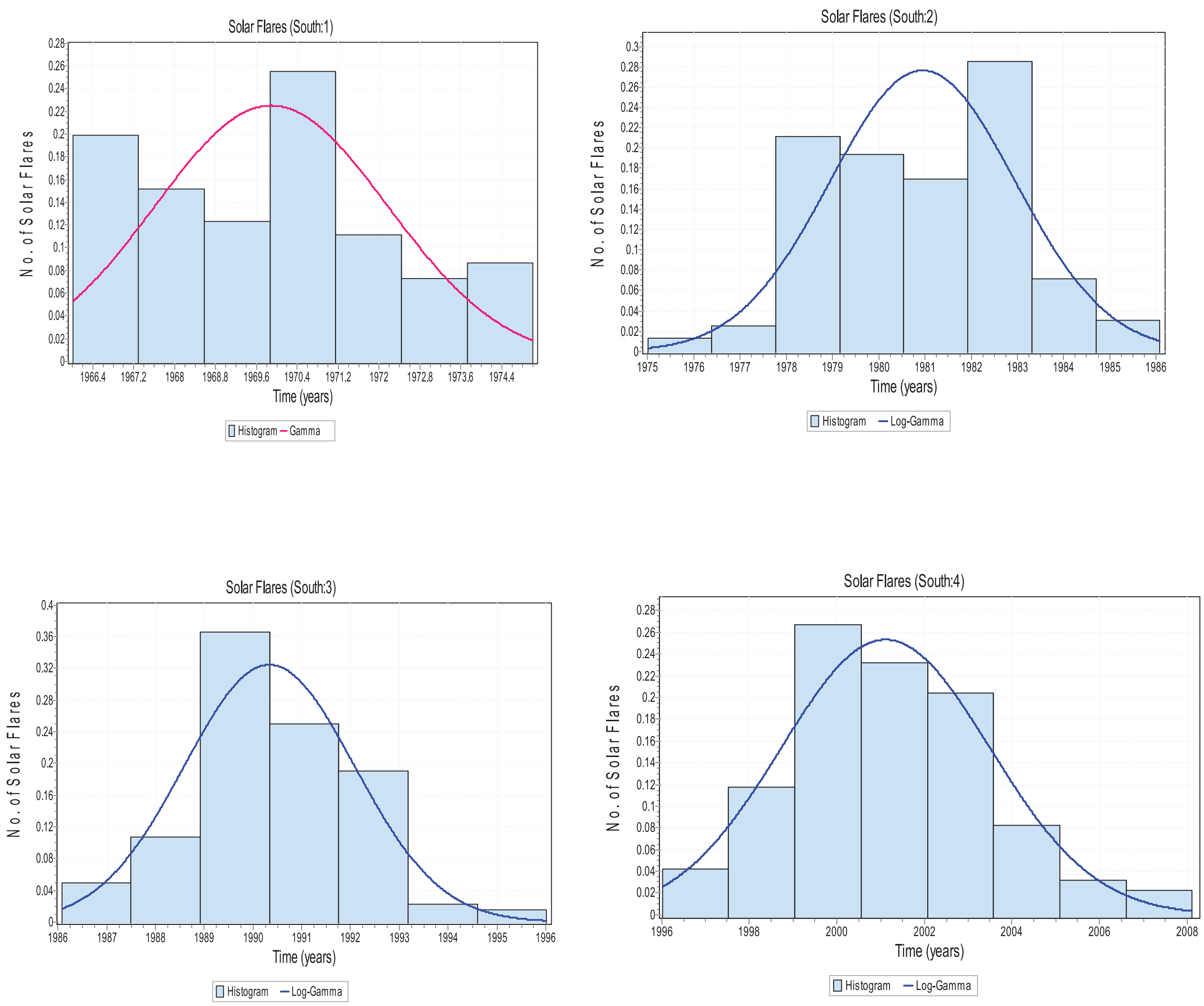

Fig. 2. Probability distributions of 4 cycles of the solar flares South-Hemispheric data.

represent correlation as well as nature of the data.

3. RESULTS AND DISCUSSION

In the first section we have obtained the most suitable probability distributions for SF North-South hemispheric data sets along with the total SF data from 1966 to 2008. The SF data of all kind is maintained by the world data center. The results show that the probability distribution for solar flares cycles 1 and 2 of North-South hemispheric SF data follows the GD and LGD respectively (see figure:1 and 2). This is also same for the total solar flares cycles 1 and 2 (depicted by Tables: 1 and 2). The results represent that the fluctuation of solar flares activity are almost similar for N-S hemispheric and total SF data particularly for Cycle 1 and 2. The variations of probability distributions in solar flares cycles 3 and 4 represent that there is fluctuation in solar activity cycles for North and South hemisphere (depicted by Tables: 1 and 2 also figures:1, 2 and 3). This variation of solar flares activity cycles corresponds to the change in climate in two hemispheres. Solar flares cycles 3 and 4 follows the GD and CSD respectively of North hemisphere (see figures:1 and Table:1). While for South hemisphere the results show GD for cycles 3 and 4 as depicted by the table: 2 (see also figure:2). This represents no change in the probability distribution for South hemisphere SF data. But the change in North hemisphere shows a prolonged

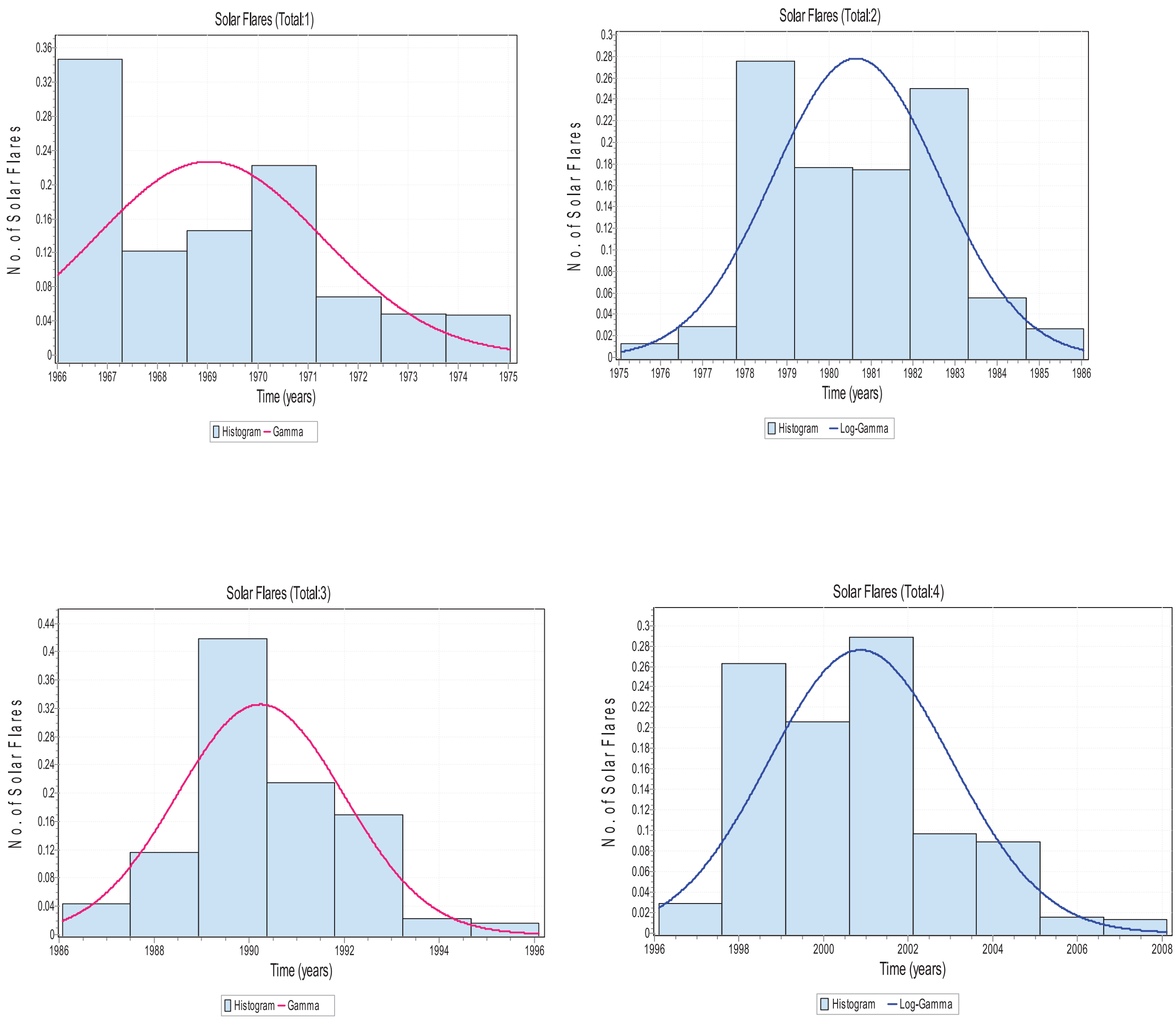

Fig. 3. Probability distributions of 4 cycles of the solar flares Total-Hemispheric data.

more smooth solar flares activity. The total SF data cycle 3 and 4 shows GD and LGD respectively (Table: 3 and figure: 3 depicts the distributions). This kind of variation in total solar activity indicates increase in cycle length but not as much as in North hemisphere. The probability distributions for complete SF data of both hemispheres represents CSD while for complete total SF data it is GD (as depicted by Tables: 1, 2 & 3). Same probability distribution on N-S hemisphere shows that total solar flares activity on two hemispheres have same ascending and descending phase and time while positive tail means prolonged length of cycles. These variation can also be observed with the help of mean and standard deviation of each cycle which shows a fluctuation between the hemispheres as depicted by tables:1, 2 and 3.

In the second section we estimated the FD for each solar flares cycle and observe complexity for both the N-S hemispheres along the total SF data. We also compare the complexity for each hemisphere as depicted in table: 4 depicts the FD of each cycle. The FD for cycle wise data indicates that solar flares data on North hemisphere is more or less complex (smoother) as compared to the South hemisphere. This means that SF data on North hemisphere is more persistent than the South hemispheric SF data and also than the total SF data sets. Table: 4 depicts the basic information and comparison between the fractal behavior of each hemispheric data.

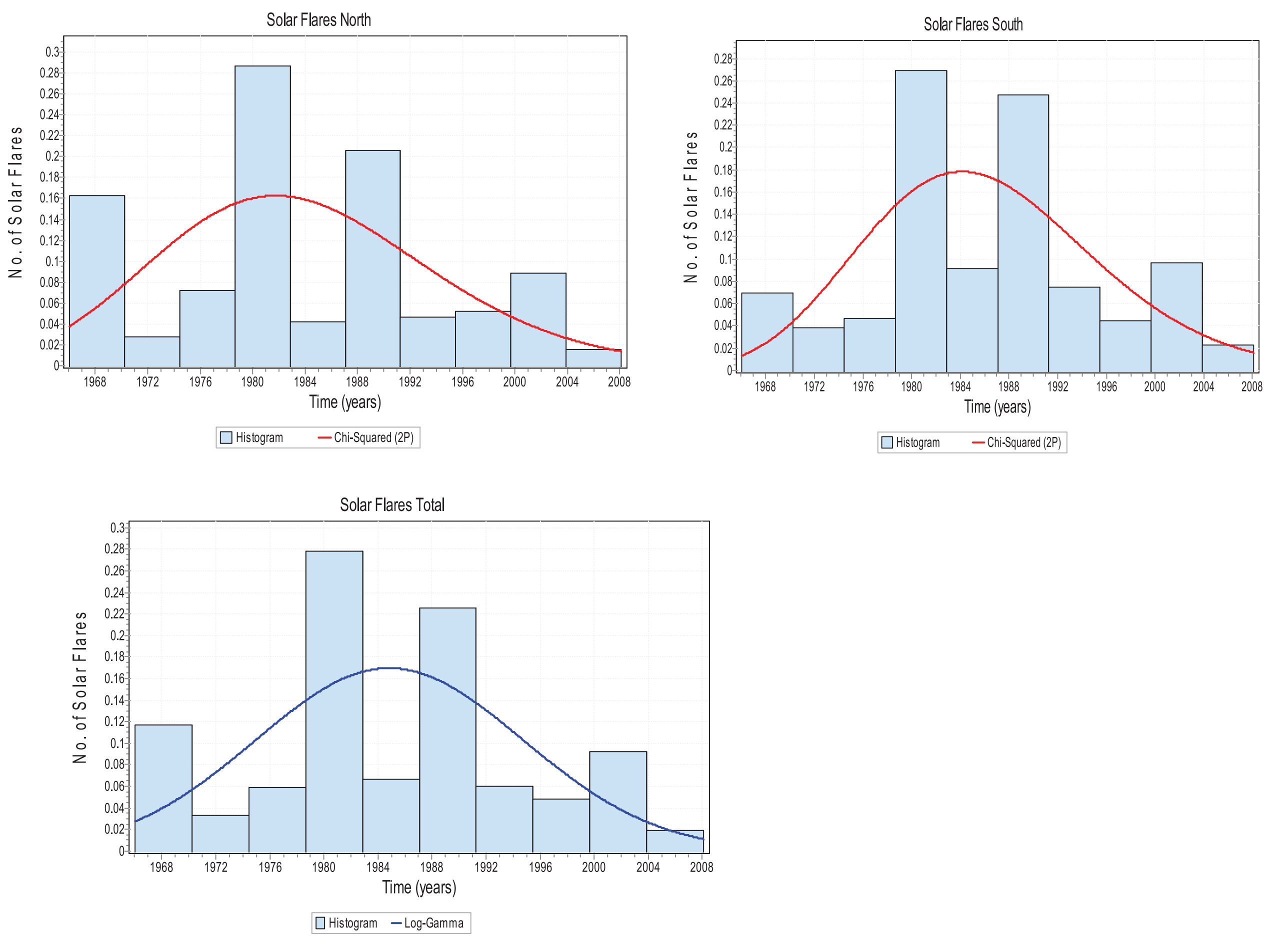

Fig. 4. Probability distributions of North, South and Total solar flares data.

We have also analyzed the probability distributions of the SF data as used above. This also shows persistency in the perspective of FD as depicted by table: 4. It also confirms that persistency of any data points can be analyzed in the perspective of probability distribution. So the technique will be helpful to study the physical nature of any data. The results of FD for North, South and Total SF data indicates that North hemisphere and total SF data are more persistent than South hemisphere SF data. This is because of the shifting of solar flares from North to South and South to North. The North and total SF data have fluctuation in probability distribution in different cycles although South hemisphere SF data have the same probability distribution for cycles 2, 3 and 4 (table: 5). The Hurst exponent for each cycle are also obtained with the help of FD to observe the smoothness and persistency more accurately. Table: 4 depicts H for smoothness.

4. CONCLUSIONS AND OUTLOOK

The first section of this paper contains the probability distributions of 4 cycles of solar flares data. We have compared the probability distributions of North and South hemisphere along with the total solar flares data starting from 1966 to 2008. Tables: 1 and 2 depict the related distribution. The SF cycles have different durations in accordance with sunspots data. The shortest duration of SF data is 9 years and longest is 12 years approximately. Tables: 1 and 2 depict the length of the two durations for N and S hemispheres. The length of cycles 1, 2, 3 and 4 for North, South and Total SF data are same. Cycles 1,

Table 1. North-solar flares.

Cycle	Duration	years	mean	St.dev	Distribution	Statistic(KST)	Parameters
1	1966.01-1975.04	9	2.5659	2.289	Gamma	0.17799	α=8.6763E^{+5} β=0.00227
2	1975.05-1986.06	11	4.1737	4.7093	Log-Gamma	0.12278	α=5.9028E^{+7} β=1.2860E^{-7}
3	1986.07-1996.1	9	3.2277	3.6984	Gamma	0.16789	α=1.4309E^{+6} β=0.00139
4	1996.11-2008.12	12	1.5692	2.3141	Chi-Squared (2P)	0.15405	ν=5 γ=1996.1
1- 4	1966.01-2008.12	42	2.8887	3.5753	Chi-Squared (2P)	0.09801	ν=55 γ=1928.8

Table 2. South-solar flares.

Cycle	Duration	years	mean	St.dev	Distribution	Statistic (KST)	Parameters
1	1966.01-1975.01	9	1.3986	1.1179	Gamma	0.13215	α=7.4762E^{+5} β=0.00263
2	1975.02-1986.08	11	4.1737	4.7093	Log-Gamma	0.10387	α=5.6661E^{+7} β=1.3398E^{-7}
3	1986.09-1996.02	9	3.2277	3.6984	Log-Gamma	0.13005	α=7.5050E^{+7} β=1.0121E^{-7}
4	1996.03-2008.12	12	1.5692	2.3141	Log-Gamma	0.1094	α=4.0743E^{+7} β=1.8657E^{-7}
1- 4	1966.01-2008.12	42	2.7348	3.2472	Chi-Squared (2P)	0.0829	ν=46 γ=1940.1

Table 3. Total-solar flares.

Cycle	Duration	years	mean	St.dev	Distribution	Statistic (KST)	Parameters
1	1966.01-1975.01	9	3.9061	2.6764	Gamma	0.14652	α=7.5396E^{+5} β=0.00261
2	1975.02-1986.08	11	4.1737	4.7093	Log-Gamma	0.09863	α=5.7980E^{+7} β=1.3093E^{-7}
3	1986.09-1996.02	9	3.2277	3.6984	Gamma	0.14317	α=1.2847E^{+6} β=0.00155
4	1996.03-2008.12	12	1.5692	2.3141	Log-Gamma	0.13373	α=4.9148E^{+7} β=1.5466E^{-7}
1- 4	1966.01-2008.12	42	5.6232	5.904	Log-Gamma	0.09822	α=2.3181E^{+6} β=3.2757E^{-6}

2, 3 and 4 have the respective length as 9, 11, 9 and 12 years. Cycle 1 and 2 for North, South and Total SF data follows GD and LGD respectively while cycle 3 of North and Total SF data follows the GD and for South hemisphere SF data shows LGD. The results for cycle 4 show LGD for South and Total SF data and for North hemisphere it follow CSD. Tables: 1 and 2 depict the related information. The probability distribution for complete cycle of North, South hemispheric SF data is CSD and for Total SF data it follows LGD. All the results of probability distribution are tested with the help of Kolmogorov-Smirnov *D*-test. Tables: 1, 2 and 3 depict the statistical analysis. In the second section of this paper we analyzed the complexity of each cycle of SF data and then compared them by

Table 4. FD and H of solar flares data.

Cycle	FD (North)	H (North)	FD (South)	H (South)	FD (Total)	H (Total)
1	1.271	0.729	1.463	0.537	1.394	0.606
2	1.072	0.928	1.282	0.718	1.047	0.953
3	1.121	0.879	1.154	0.846	1.048	0.952
4	1.122	0.878	1.176	0.824	1.112	0.888
1- 4	1.134	0.866	1.241	0.759	1.15	0.85

Table -5. Comparison of distribution and FD.

Cycle	FD (North)	FD (South)	FD (Total)	Distribution North	Distribution South	Distribution Total
1	1.271	1.463	1.394	Gamma	Gamma	Gamma
2	1.072	1.282	1.047	Log-Gamma	Log-Gamma	Log-Gamma
3	1.121	1.154	1.048	Gamma	Log-Gamma	Gamma
4	1.122	1.176	1.112	Chi-Squared (2P)	Log-Gamma	Log-Gamma
1- 4	1.134	1.241	1.15	Chi-Squared (2P)	Chi-Squared (2P)	Log-Gamma

estimating FD using FRACTALYSE 2.4 software. Hurst exponents also computed through FD which is depicted by table: 4. Results show that FD for North and Total SF data are less than those for South hemisphere SF data. This means that SF data for North and Total hemisphere are more persistent than South hemispheric SF data as depicted by table: 4 for FD and H. It also confirms the fact that if FD increases then H decreases. This kind of variation appears because of the difference in the ascending and descending phases of SF activity of each cycle. The cycle will prolong if the difference between the ascending and descending phase of SF activity has greater means and the tail prolongs.

In the end a relation between probability distribution and persistency is provided in table: 5 depicts. All probability distributions show persistency. The mean-tail analysis confirmed the FD-H analysis.

5. ACKNOWLEDGEMENTS

We are very thankful to the World Data Center (WDC) for providing the North, South and Total Solar Flares data. The contents of this paper form part of the first author's doctoral thesis.

6. REFERENCES

1. Michael F. Barnsley: Fractals everywhere. *Academic Press*, ISBN 978-0-12-079062-3, pp. I-XII, 1-394 (1988).
2. Walck. Handbook on statistical distribution for experimentalists. Internal Report SUF-PFY/96-01, Stockholm, p. 96-102 (2000).
3. Falconer, K., *Fractal Geometry Mathematical Foundations and Applications*. Wiley & Sons. (2007).
4. Hussain, M. Arif, Shaheen Abbas, M. Rashid Kamal Ansari, & Asma Zaffar. Study of Arabian Seawater Temperature Fluctuations. *Journal of Basic & Applied Sciences* 8: 557-560 (2012).
5. Jilani, Saifuddin Ahmed, & M. Ayub Khan Yousf Zai. Variability of Solar Flare Duration and Its Effects on Ozone Concentration at Pakistan Air Space, *Proceedings of the Pakistan Academy of Sciences* 48 (1): 51–55, (2013)
6. Kilcik, A., C.N.K. Anderson, J.P. Rozelot, H. Ye, G. Sugihara, & A. Ozguc. Nonlinear prediction of solar cycle 24. *The Astrophysical Journal* 693(2), 1173 (2009).
7. Li, K.J., H.F. Liang, H.S. Yun, & X.M. Gu. Statistical behavior of sunspot groups on the solar disk. *Solar Physics*, *205*(2), 361-370,(2002).
8. Kumaraswamy, Krishna. *Fractal Dimension for Data Mining.* Center for Automated Learning and Discovery School of Computer Science Carnegie Mellon University. 5000 Forbes Avenue, Pittsburgh,

PA 15213.

9. Hussain, M.A. *Mathematical Aspects of The Impact of Urban Greenhouse Gas Emissions on Global Warming,* Ph.D. Thesis, Federal Urdu University of Arts, Science and Technology, Karachi (2006).
10. Yousafzai, M.A.K., *A Quantitative Study of Effects of Ozone Layer Depletion on Marine Organisms with Reference to Coastal Regions of Pakistan.* Ph.D. Thsis. University of Karachi, Pakistan (2003).
11. Weiss, N.O. Sunspot structure and dynamics. *Space Science Reviews* 124(1-4):13-22 (2006).
12. Rabello-Soares, M. Cristina. Solar-cycle variation of sound speed near the solar surface. *The Astrophysical Journal* 745(2) 184 (2012).
13. Sugihara, G., & M May, R. Applications of fractals in ecology. *Trends in Ecology & Evolution* 5(3), 79-86 (1990).
14. Thomas, John H., & Nigel Oscar Weiss. *Sunspots and Starspots*. Cambridge University Press, Vol. 46. (2008).