International Seminar on Hydrology of Extremes 1-3 December, 1988
Roorkee (India)

DROUGHT ANALYSIS AND SYNTHETIC GENERATIONS

Mandakinee Majumdar
Asstt. Professor

Civil Engineering Department, Indian Institute of Technology
Kharagpur, W.B. (India)

SYNOPSIS

Synthetic data generation on two Indian rivers, Damodar and Cauveri was carried out using ARIMA (1,0,1) and fast fractional Gaussian noise model (ffGn). The quality of generated data regarding preserving the historical basic properties like mean, variance as well as long term properties like Hurst Coefficient (H) and run lengths of low flow generation as depicted by drought curve was investigated. The results for both type generation suggest that ffGn has a better capability of generating long low flow sequences than ARIMA model.

1.0 INTRODUCTION

Synthetic data generation has an increasingly important role to play at the design and planning stages as well as the operation stage of water resources projects. A vast number of short term and long term generation models are available for generation. Choice depends on the purpose and type of generation. In this study, the focus is on long term generation. For this type of generation Auto Regressive Integrated Moving Average models of order p, d, q [ARIMA (p,d,q)], fractional Gaussian noise models (fGn), Broken Line models (BL), etc. are found useful (1). These models are intended to preserve basic statistical properties like mean, variance, auto correlation structure, skewness etc. In addition their capability to reproduce Hurst coefficient (H) of the historical record is also of interest. Besides, their ability to incorporate the so called "Noah and Joseph" effects is also of concern. Mandlebrot and Wallis (2) have noted that markovian models fail to exhibit Noah and Joseph effects.

In recent years several researchers have carried out investigations focusing on drought analysis of streamflows. Srikanthan and Mcmahon (4) have developed an analytical procedure to determine recurrence interval of long hydrologic events, droughts among them. Zekai Sen (3) has suggested a method of carrying out statistical analysis of critical droughts. Zelenheric and Salvai (5) suggest a method of describing and analysing the process of streamflow drought.

In this paper investigation is carried out to explore capability of two long term generation models, namely, ARIMA (1,0,1) and fast fractional gaussian noise (ffGn) with regard to drought periods with specific application to two Indian rivers Damodar (Eastern India) and Caueri (South India).

2.0 METHODOLOGY

2.1 ARIMA (1,0,1) model can be expressed as

$$X_t = \delta + \phi X_{t-1} - \theta a_{t-1} + a_t \quad .. \quad (1)$$

Where X_t is flow at time t and a_t s are normally distributed $N(0,\sigma_a^2)$. Parameters can be estimated by standard procedures. Stationarity requires that $|\phi| < 1$ and $|\theta| < 1$.

2.2 ffGn generation involves sum of one high frequency term X_h and one low frequency component X_L.

$$X(t,H) = X_h(t,H) + X_L(t,H) \quad .. \quad (2)$$

$$X_L(t,H) = \sum_{n=1}^{N(T)} W_n X(t,r_n) \quad .. \quad (3)$$

where $X(t,r_n)$ is a lag one autoregressive model with zero mean, unit variance and covariance function

$$r_n = e(-B^{-n}).$$

N(T) is the number of Markov models, which depends on length of desired generation T and quality parameter Q in addition to parameter B.

$$N(T) = |\log(QT)/\log(B)| \quad .. \quad (4)$$

$$X_t(r,n) = r_n X(t-1, r_n) + (1 - r_n^2)^{\frac{1}{2}} a_t \quad .. \quad (5)$$

The weight W_n is given by

$$W_n^2 = \frac{[H(2H-1)B^{-1-H} - B^{-(1-H)}]}{T(3-2H)} B^{-2(1-H)n} \quad .. \quad (6)$$

H being the Hurst coefficient. Value of Q is usually taken as 4,5 and 6 whereas 2,3 and 4 are typical values of B.

High frequency term $X_h(t,H)$ has variance of,

$$\sigma_h^2 = \frac{1-B^{-(1-H)} H(2H-1)}{T(3-2H)} \quad .. \quad (7)$$

and lag one autocorrelation,

$$\rho_h(1) = 2^{2H-1} - 1 \sum_{n=1}^{N(T)} W_n(1-r_n) - \frac{B^{-(1-H)} H(2H-1)}{T(3-2H)} \quad .. \quad (8)$$

3.0 RESULTS

Results of the generation experiments carried out using above mentioned model for annual streamflows of rivers Damodar and Cauveri are presented below.

3.1 Generation by ARIMA on Damodar with sample size fixed (100) was done to see the effect of number of samples on the average properties of generation. 100, 1000 and 10000 realisations were generated. Table 1 shows the results. The mean and coefficient of variation are normalised by dividing the sample values by corresponding historical (observed) values. ($C_v = C_{vg}/C_{vo}$ and $\bar{X} = \bar{X}_g/\bar{X}_o$). The average maximum and minimum generated values are similarly shown as multiples of observed mean.

TABLE 1 : ARIMA (1,0,1) Generation : Damodar

	No. of samples	Mean-$\bar{X}$	C_v	H	Avg Min	Avg Max	Max
	100	1.002	.965	.549	.0704	2.296	2.865
A.	1000	1.007	.974	.544	.0650	2.283	2.963
	10000	1.005	.973	.544	.0630	2.284	3.408
	Observed	1.000	1.000	.516	Min .3145	Max 2.765	
Sample length							
	100	1.002	.965	.549	.070	2.286	2.865
B.	1000	1.007	.970	.520	.008	2.602	2.898
	10000	1.005	.974	.515	-	2.765	3.772

A. Sample length = 100 B. No. of samples = 100

Average drought curves corresponding to A and B are shown in Table 2. Here they are normalised by a normal curve, which is a drought curve generated by annual flows with average observed value. As expected number of samples did not have much influence on the drought curve as long as sample length was same. However, as seen in case B longer the sample more severe the drought curve. The "extreme drought curve" may be considered to be made up of minimum values of the drought curve coordinates obtained by different realisations. Table 2 shows the values of extreme drought curves as well. It was observed that no single sample matched the extreme curve fully. In Fig.1 average drought curve and extreme drought curve are shown for the first case of Table 1, case A.

Generation on Cauveri resulted in almost identical conclusions. (Tables 3-4). In both cases the H property of historical record was not preserved except that in case of Damodar it was preserved when sample length was very large.

TABLE 2 : Drought Curves : Damodar

A.	5 Year	10	20	30	40	50	100
100	.614	.765	.871	.912	.938	.957	1.012
ED	.306	.510	.706	.738	.799	.822	0.897
1000	.600	.759	.861	.905	.932	.950	1.006
10000	.603	.756	.860	.822	.931	.951	1.005
B.							
100	.614	.715	.871	.912	.938	.957	1.012
1000	.480	.654	.780	.828	.863	.883	0.929
10000	.373	.575	.714	.776	.811	.834	0.889

ED - extreme drought curve

Although ARIMA models have the capacity of preserving a particular H value there is no explicit way to do it and hence we have no control over the value.

TABLE 3 : ARIMA (1,0,1) Generation : Cauveri

	NO. of samples	Mean-$\bar{X}$	C_v	H	Avg Min	Avg Max	max
	100	.996	.988	.646	.149	1.627	2.189
A.	1000	.999	.998	.600	.168	1.860	2.560
	10000	1.004	.989	.640	.164	1.655	2.602
	observed	1.000	1.000	.739 Min	.624 Max	2.792	
	Sample length						
	100	.996	.988	.646	.149	1.627	2.286
B.	1000	1.005	.985	.596	.186	1.836	2.177
	10000	1.000	.999	.578	-	2.024	2.530

TABLE 4 : Drought Curves : Cauveri

A.	5 Year	10	20	30	40	50	100
100	.635	.769	.866	.907	.933	.953	1.007
1000	.620	.755	.855	.899	.927	.947	1.001
10000	.619	.755	.855	.900	.927	.946	1.000
B.							
100	.635	.769	.866	.907	.933	.953	1.007
1000	.491	.659	.771	.821	.851	.872	0.921
10000	.376	.553	.693	.754	.791	.818	0.877

3.2 Investigation with ffGn was done for all nine combinations of parameters B and Q. In each case 100 samples of length 100 were generated. No significant difference was apparent in results from different combinations. Hence representative results are shown in Table 5 and 6 for Damodar and Cauveri.

In general average maximum value was lower than the observed maximum flow values. One striking feature was that almost in all cases the extreme drought curve turned out to be from the name realisation. This was not so in ARIMA generation. This clearly shows that longer length low flow extreme events are better depicted by ffGn model.

TABLE 5 : ffGn results - Basic Properties

	Mean	C_v	H	Avg Min	Avg Max	Max
Damodar	.999	1.008	.5568	.060	2.292	3.198
Caueri	.988	1.007	.7000	.223	1.857	2.167

TABLE 6 : Drought curves - ffGn

	5 Year	10	20	30	40	50	100
Damodar							
Average	.564	.725	.844	.887	.919	.937	.991
Extreme	.237	.478	.460	.565	.794	.813	.906
Caueri							
Average	.615	.728	.817	.863	.894	.919	.989
Extreme	.298	.466	.596	.646	.687	.698	.727

Effect of sample number was found virtually insignificant. However, effect of sample length had very strong influence on the drought curve.

A comparison of drought curves resulting from ARIMA and ffGn generation for a comparable case (100 samples with sample size 100) is shown in Fig. 1. Similar results were obtained for all cases investigated. It is clear from the graph that there is a clear trend to reach the "normal" line implying that at least upto 100 year length, the worst drought periods are of smaller lengths. Here any flows leading to below normal curve are considered to be low flow periods. This shows that both type of models are unsuitable for real long lengths of low flows (i.e. above 50 years). But the models are useful for simulating low flow periods of up to fifty years of which 20 years are real low flows.

4. CONCLUSION

It has been found that both ARIMA and ffGn are unsuitable for generating very high flows. However, they are adequate for low flow generation. In general ffGn has somewhat better capability to generate long run lengths of low flows as is demonstrated by generation experiments on Domodar and Caueri.

REFRENCES

1. Bras, Rafael L. and Rodriquez-Iturbe, I.(1984)'Random Functions in Hydrology.' Addison-Wesley Publ. Co.

2. Mandlebrot, B.B.; J.R. Vallis and Noah, Joseph (1968) ''Operational Hydrology.' Water Resources Research 4(5) pp.909-918.

3. Sen, Zekai (1980)'Statistical Analysis of Hydrologic Critical Droughts.' Journal of Hydraulics Div., ASCE, Vol.106, HY1, PP. 99-115.

4. Srikanthan, R. and T.A. Mcmahon (1986). 'Recurrence Interval of Long Hydrologic Events'. J. of ASCE, Hydraulic Engineering, Vol. 112, NO. 6, PP. 518-538.

5. Zelenhasic, E. and Atila Salvai(1987) 'A method of Streamflow Drought Analysis.' Water Resources Research Vol. 23(1), PP. 158-168.

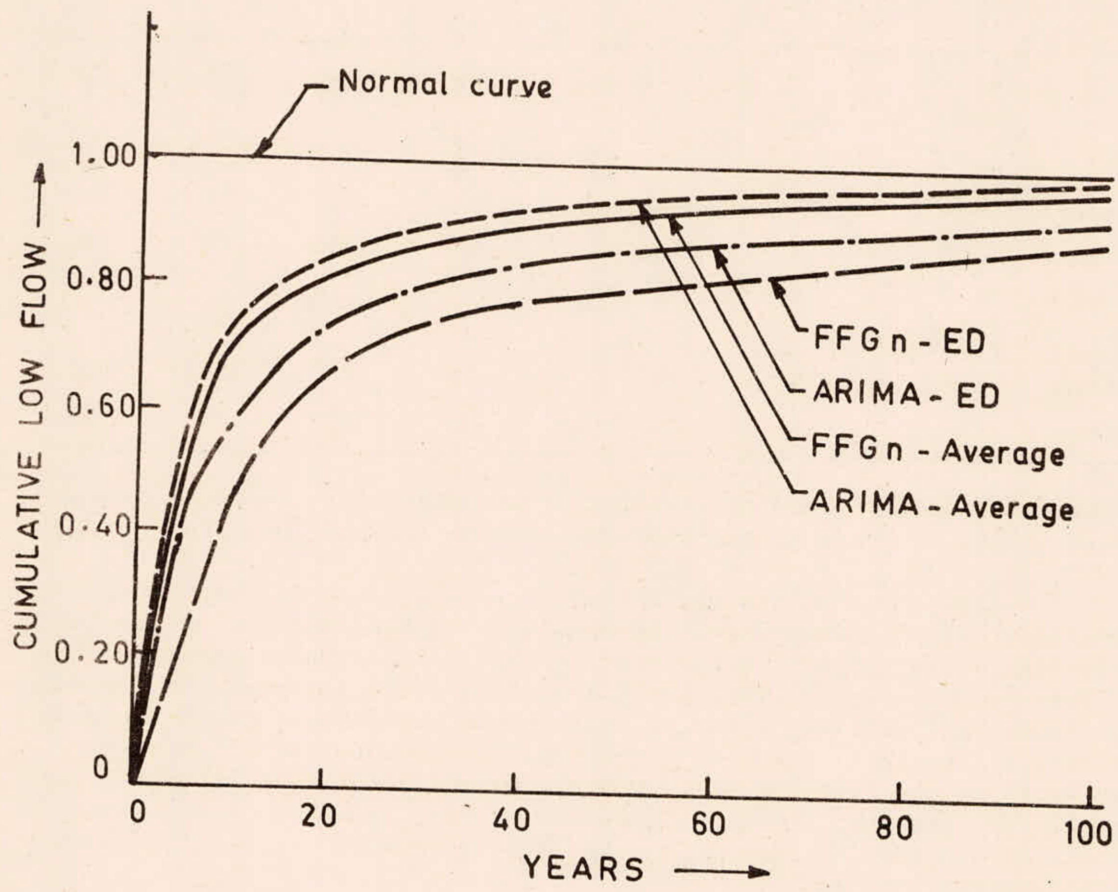

FIG. 1 COMPARISON OF DROUGHT CURVES BY TWO MODELS FOR DAMODAR