An Experiment on the Hurst Exponent based on FARIMA

Chen Pu[1] , Li Ni[2] , Xu Jie[1] , Zhao Ting[1] , Liu Chen[3][4]

(1): State Grid Hubei electric power company, China
(2): PowerChina Hubei Electric Engineering Corporation, China
(3): School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, China
(4): Beijing Chuang Ming Technology Co., Ltd., China

Keywords: Long-range Dependence; FARIMA Model; Hurst Exponent

Abstract. Based on the theory of time series analysis for the definition of FARIMA model, we generate time seriesin accordance with the FARIMA model, and measure its Hurst exponent and fractal order by using R/S analysis method. The experimental results show that the Hurst exponent used to generate the series has difference from the Hurst exponent of the real sequence measured by using R/S analysis. The further analysis indicate that, the noise types and the estimation of negative time-series data series length both have effect on the deviation.

I. INTRODUCTION

Most of the time series exhibit erratic significant mutations in network traffic, in which the data is related in a long time scale, and this property is so-called long-range dependence (LRD)[5]. However, previous models, including Markov process, Poisson process, autoregressive (AR) model and moving average (ARMA) model and so on, can only describe the data of short-range dependence (SRD)[3]. Yet, self-similar models such as fractional gaussian noise (FGN) model and fractional Brownian motion (FBM) model have been developed to resolve this problem. Unfortunately, these models are incapable of giving a description of SRD, which is also a significant property of the network traffic. Fractional autoregressive integrated moving average (FARIMA) model is provided by this work to reveal the co-existence of both SRD and LRD in the network flow [3], and the model also has a wide range of applications in other domains.

Dr Kristoufek found that the Hurst exponent he got from the time series he previously generated had large deviation to the expected one which was used to generate the time series [1]. In this work, the times series will be generated with FARIMA model based on different noise sequences and different models for data of negative time axis. Then rescaled range analysis (R/S Analysis) will be employed to estimate the Hurst exponent of the time series generated by FARIMA. Then the impact of the different noise sequences and the data of negative time axis on the generated series will be summarized in this paper.

II. FARIMA MODEL

The FARIMA model can be viewed as a special case of ARIMA model [11].In ARMA(p,q,d),p is the autoregressive order, q is the moving average order and d is the differential order. In the ARIMA model, d can only be integer, but in FARIMA model, d is form -0.5 to 0.5. If $\{X_t\}$ is a long memory process, and$d \in (0,0.5)$. Such process$\{X_t\}$, denoted by the FARIMA model, is defined by equations:

$$\varphi(B)\Delta^d X_t = \theta(B)a_t$$

Where a_t is independent and identically distributed noise sequence, and

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \cdots - \varphi_p B^p$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

Where B is a backward shift operator:$BX_t = X_{t-1}$, $\varphi(B)$ and $\theta(B)$ has no common roots and $\varphi(B)$ has no roots in the closet unit disk. $\Delta = 1 - B$, and

$$\Delta^d = (1 - B)^d = \sum_{j=0}^{\infty} \binom{d}{j} (-B)^k$$

Where

$$\binom{d}{j} = \frac{\Gamma(d + 1)}{\Gamma(j + 1)\Gamma(d - j + 1)}$$

Γ is the gamma function.

According to the FARIMA model definition above, two data points lying far away in the time axis still correlates with each other due to the existence of the fractional difference order d, which reflects the long memory properties of sequences [2].

III. THE ALGORITHM OF GENERATING FARIMA MODEL

We directly get time series based on the definition of FARIMA (p,d, q) model.

According to FARIMA definition, we obtain

$$X_t = \theta(B)\varphi(B)^{-1}\Delta^{-d}a_t$$

Therefore we can firstly take the difference of $\{a_t\}$ with fractional difference order d and obtain a new noise sequence $\{z_t\}$:

$$z_t = \Delta^{-d}a_t$$

Finally, we can get $\{X_t\}$ by using the ARMA (p, q) model to drive $\{z_t\}$.

To start with, let's see how we can get $\{z_t\}$. Let

$$z_t = \Delta^{-d}a_t = \sum_{j=1}^{\infty} \pi_j z_{t-j} + a_t$$

Where

$$\pi_j = \binom{d}{j}(-1)^{j+1} = \frac{(-1)^{j+1}\Gamma(d + 1)}{\Gamma(j + 1)\Gamma(d - j + 1)}$$

$$\pi_{j+1} = \frac{(-1)^{j+2}\Gamma(d + 1)}{\Gamma(j + 2)\Gamma(d - j)}$$

According to the properties of the Gamma function $\Gamma(x + 1) = x\Gamma(x)$, we can obtain:

$$\pi_{j+1} = \frac{j - d}{j + 1}\pi_j$$

Where $\pi_1 = d$.

Using the coefficient recursion formula above, the emergence of gamma function can be avoided in the algorithm, and we can calculate the new noise sequence $\{z_t\}$ after we have differenced at $\{a_t\}$ with fractional difference order d. Then we replace the original noise in ARMA (p, q) model with the new noise sequence $\{z_t\}$ and obtain $\{X_t\}$.

According to the above method about how to generate a time series donated by the FARIMA model, the values and length of the time series on the negative axis and the value of noise sequence will have a strong impact on the generated FARIMA data. For the following part, we will analyze how these factors influence the Hurst exponent of the sequence based on our computer simulation results.

IV. THE EXPERIMENT AND ANALYSIS

A. *Introduction to distribution functions*

This work involves four kinds of distributions: t-distribution, normal distribution, lognormal distribution and autoregressive process.

Normal distribution is a well-known random process. Standard normal distribution, N (0,1), is used in this paper to generate the target time series. t-distribution is similar to standard normal distribution to some extent, but it only needs one parameter while normal distribution needs two. t-distribution has greater similarity with normal distribution when the parameter which is called variance becomes larger. This paper employs the t-distribution whose variance is 1 and 25. In this work, lognormal distribution is derived from standard normal distribution. And p-level AR process $\{Y_t\}$ can be defined as following:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + e_t$$

or

$$\varphi(B)Y_t = e_t$$

e_t is a noise sequence, and the roots of the polynomial $\varphi(B)$ must be outside the unit circle.

B. *Experiment and methods*

Because FARIMA model is a long-range dependence model, so all the data is not only related to the data on the positive timeline, but also associated with the data of negative timeline according to the definitions above in the second section. The fact that data on negative timeline can influence the generated data is due to the fact that the actual data can't have an infinite number of non-zero value on the negative timeline. In addition, the length of the time series and the distribution of noise also have a great impact on the subject sequence. It can be concluded that all the above factors result in the discrepancy between the actual Hurst exponent and the anticipatory one.

So this work generates the data out of FARIMA (0, 0.25, 0) based on the above algorithm. The length of generated sequences is 20000, 15000, 10000 and 5000 respectively. The noise sequences follow standard normal distribution, t-distribution whose variance is 1 and 25 and lognormal distribution respectively. On the negative timeline, standard normal distribution, AR (8), AR (2) and all-zero are tested individually. The length of the noise sequence respectively take 100,500 and 1000.

After generating the data according the different combinations, R/S Analysis will be used to calculate the Hurst exponent firstly, then compare the Hurst exponent with the expected Hurst exponent to dig out the influence of different factors. The Hurst exponent is 0.75 in this paper, because R/S Analysis is more accurate in this regard [5]. Another reason is that large data volume ensure accuracy of the experiment [11].

The conclusions and analysis

This work first analyses the experimental result in which the length of data is 20000 and the length of data on negative timeline is 1000. The detail is listed in the table 1 table 2 and chart 1. When the noise follows t- distribution and the data on the negative timeline follow AR process, the Hurst exponent is 0.7490 which is most close the target one. The paper [18] mentioned that the data on the negative timeline can be forecasted by using the high level AR process. However, AR (8) process and AR (2) process are fundamentally the same for this case. So promoting the level of AR process have no influence on the discrepancy between the actual Hurst exponent and the expected one.

However, for the same noise distribution, AR process is more accurate than all-zero and normal distribution. And t(1) is most accurate and t(25) is less accurate than t(1) while normal distribution is the worst one. Of course, it can't be asserted that the generated Hurst exponent is closer to the target one when the variance of t-distribution becomes bigger.

Now the emphasis is on the situation where the noise distribution is lognormal distribution. The Hurst exponent is bigger than 0.74 while using lognormal distribution as the noise distribution

whichever the model on the negative timeline is. And the susceptibility of lognormal distribution is obviously smaller than the one of t-distribution referring to the chart and data from the simulation. Under other circumstances, the similar phenomenon appear too.

Then this work will have a look at the standard deviation of the generated Hurst exponent. Although it can't be applied to all, the variance becomes bigger as the generated Hurst exponent becoming closer to the expected one. It is uncovered that larger data volume makes a better result, but the result can be very different unexpectedly sometimes.

Finally, it can be concluded that all the generated Hurst exponent is smaller than the expected value because of limited data volume and limited length of the data on the negative timeline. And it is impossible to achieve ideal conditions for each data associated with infinite data before. So it is reasonable that the generated Hurst exponent is smaller than the expected one.

C. Analysis

Chart 1 and table 1 are the variance and mean of Hurst exponent in the experiment where the data volume is 20000 and the length of data on the negative timeline is 1000. ND stands for noise distribution and NT is negative timeline. Normal in the table stands for normal distribution, hence lognormal stands for lognormal distribution.

TABLE I.   THE VARIANCE OF THE HURST EXPONENT

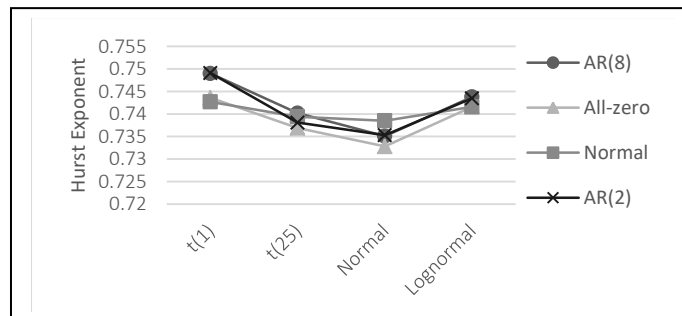| NB / NT	t(1)	t(25)	Normal	Lognormal
AR(8)	0.0267	0.0199	0.0156	0.0174
All-zero	0.0223	0.0199	0.0171	0.0200
Normal	0.0217	0.0171	0.0178	0.0168
AR(2)	0.0307	0.0178	0.0179	0.0155

Figure 1. The mean of Hurst exponents

The result listed in table 2 and 3 comes from the experiment where the length of the negative timeline is changed (100 and 500).

TABLE II.   THE MEAN OF THE HURST EXPONENT

NB / NT	t(1)	t(25)	Normal	Lognormal
100	0.7469	0.7376	0.7336	0.7378
500	0.7446	0.7373	0.7335	0.7432

TABLE III.   THE VARIANCE OF THE HURST EXPONENT

NB NT	t(1)	t(25)	Normal	Lognormal
100	0.0234	0.0179	0.0165	0.0197
500	0.0248	0.0162	0.0159	0.0159

It can be informed that the generated Hurst exponent is more different with the expected one because of the relation of the data on the positive timeline and the negative timeline, when the length of negative timeline change from 1000 to 100. And this idea is verified by the experiment. However, when the length of the negative timeline reach at 2000, the result's variance is bigger than the previous one, which reveals that the larger length of the negative timeline doesn't necessarily mean smaller deviation.

Then the length of the generated data is 15000 and 10000, and the length of the negative timeline takes 1000. It can be predicted from the earlier experiment that the error of the generated Hurst exponent will be larger. And the experiment have proved the prediction. Results of AR (8) are listed in table IV and V while the other results are not included in this paper.

TABLE IV.   THE MEAN OF THE HURST EXPONENT

NB NT	t(1)	t(25)	Normal	Lognormal
15000	0.7428	0.7383	0.7315	0.7402
10000	0.7383	0.7314	0.7279	0.7380

TABLE V.   THE VARIANCE OF THE HURST EXPONENT

NB NT	t(1)	t(25)	Normal	Lognormal
15000	0.0244	0.0205	0.0227	0.0185
10000	0.0302	0.0231	0.0248	0.0252

However, when it drops to 5000, the results become more random and less disciplinary and it is a little different from the 20000 case. The results coming from the experiment where the data volume is 5000.

As it can be seen from above tables, the error is bigger than that in the 20000 volume experiment. When it drops to 5000, the inequality, t (1) > t (25) > normal distribution, becomes false. And meanwhile the AR process which satisfies the data on the negative timeline is not the model that makes the error smallest. Totally, the error in this experiment is much larger than the error of the experiment where the data volume is 2000, because of the smaller correlation of the data which is caused by the smaller volume of the data. But the error of normal distribution situation is not very bad when it keeps around 0.74. And normal distribution lacks of sensitivity when the model on the negative timeline changes. So it is reasonable to choose normal distribution as the noise distribution when the volume of data is small.

As for variance, all the results become larger, which implies that the data volume is an important property for estimating the Hurst exponent but normal distribution is an exception. The

figure 2 shows the generated Hurst exponent obeying AR (8) process when taking different length of the data and different noise distributions.
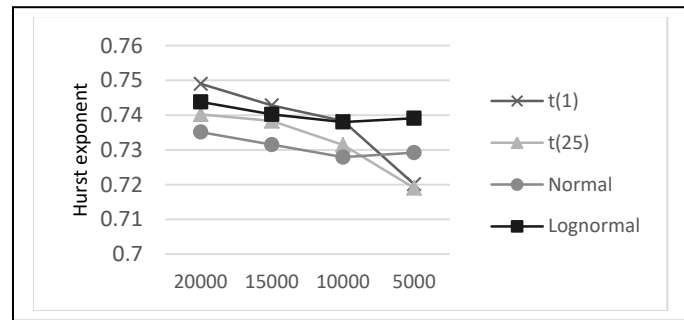

Figure 2. The mean of Hurst exponents

It can be concluded from the figure that when the volume is larger than 10000, the generated Hurst exponent is becoming lower as the data length becoming smaller. Obviously, t-distribution has a larger rate of decay than normal distribution and lognormal distribution. And the generated Hurst exponent changes a little when the noise distribution is lognormal distribution. So, like normal distribution, lognormal distribution is not sensitive to the changing of the data length.

As a matter of fact, lognormal distribution is not sensitive to the data length, the distribution of generated data, the noise length and the distribution of noise, compared with other distributions, and the generated Hurst exponent is around 0.74 which is much close to the expected Hurst exponent.

All the results in the above tables shows that lognormal distribution exhibits good properties. However, R/S Analysis is unsuitable for the situations when the data length is smaller than 5000. So this work doesn't research on the situation.

V. CONCLUSION

This work analyzes Ladislav Kristoufek's experiment and checks different factors that may influence the Hurst exponent of the time series generated by FARIMA model. It can be concluded that lognormal distribution is an ideal noise distribution for FARIMA model to generated the times series, which is of great help to make time series generated by FARIMA model be more similar to the actual network flow data. As FARIMA model is applied to predict the future network flow to help optimize the network, lognormal distribution can guarantee more accurate predictions.

REFERENCES

[1] Ladislav Kristoufek. How are rescaled range analyses affected by different memory and distributional properties? A Monte Carlo study[J].Elsevier,2012,391(17):4252-4260.

[2] Burnecki.K,Sikora,G. Estimation of FARIMA Parameters in the Case of Negative Memory and Stable Noise[J].Siganal Processingr,2013,61(11):2825-2835.

[3] Li Song,Pascal Bondon,Yang Cao,Qi Cheng,. A Time-Varying FARIMA Model for Internet Traffic[j].Siganal Processingr,2008,5:83-87.

[4] Xunyi R, Yu Y, Junfeng Z. Parameter estimation and application of time-varying FARIMA model[J]. International Journal of Advancements in Computing Technology, 2011, 3(3): 89-94.

[5] Yue J, Shang P, Dong K. Time-dependent Hurst exponent in traffic time series[C]Information Theory and Information Security (ICITIS), 2010 IEEE International Conference on. IEEE, 2010: 744-746..

[6] Hamed K H. Improved finite - sample Hurst exponent estimates using rescaled range analysis[J]. Water resources research, 2007, 43(4).

[7] Bassingthwaighte J B, Raymond G M. Evaluating rescaled range analysis for time series[J]. Annals of biomedical engineering, 1994, 22(4): 432-444.

[8] Ellis C. The mis-specification of the expected rescaled adjusted range[J]. Physica A: Statistical Mechanics and its Applications, 2006, 363(2): 469-476.

[9] Stolojescu C, Isar A. A comparison of some Hurst parameter estimators[C]Optimization of Electrical and Electronic Equipment (OPTIM), 2012 13th International Conference on. IEEE, 2012: 1152-1157.

[10] Wen Y, Zhu G. Prediction for non-gaussian self-similar traffic with neural network[C]Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on. IEEE, 2006, 1: 4224-4228.

[11] Yang S, Xiaomin T, Hao D, et al. Network predictive control based on FARIMA time-delay model[C]2012 24th Chinese Control and Decision Conference (CCDC). 2012: 3580-3585.

[12] Ellis C. The sampling properties of Hurst exponent estimates[J]. Physica A: Statistical Mechanics and its Applications, 2007, 375(1): 159-173.