# PREDICTION OF LONG-RANGE-DEPENDENT DISCRETE-TIME FRACTIONAL BROWNIAN MOTION PROCESS

*Lei Yao, Miloš Doroslovački*

Department of Electrical and Computer Engineering
The George Washington University
801 22nd St., NW
Washington, DC 20052

**ABSTRACT**

In this paper, we propose an approach to linear minimum-mean-square-error (MMSE) prediction of a discrete-time fractional Brownian motion (dt-fBm) traffic arrival process, a long range dependent traffic model that well represents the characteristics of observed Internet traces. Linear multi-step forecasts of the future values of the dt-fBm process and the corresponding prediction errors are first derived. We then proposed sliding window finite-memory predictors suitable for the practical implementation. Simulations using real-life traffic traces are performed to compare the proposed finite-memory dt-fBm predictors with fractional auto-regressive integrated moving average predictors and an empirical predictor. We find that the multi-scale sliding window dt-fBm predictor achieves best performance on forecasting the future traffic level.

## 1. INTRODUCTION

The seminal study of Leland, Taqqu, Willinger, and Wilson [1] pointed out the failure of the Poisson processes in modeling the traffic in modern data networks and the long-range dependent (and self-similar) nature of that traffic. Since then, long-range dependence and self-similarity have been reported in various types of data traffic: LAN [1,2], WAN [3,4], VBR video [5], SS7 control [6], WWW [7], and most recently Internet backbone [8]. In particular, Yao [8] reported the fitness of the fractional Brownian motion (fBm) to the Internet backbone traces.

It is generally believed that the long-range dependence is a bad news for network engineering since it introduces high variability into the traffic arrival process: periods of sustainable low and high traffic arrival rates. It could result in large delay variation and delay even at low average link utilization. However, it could also be good news for network engineering since the prediction-based traffic control can be implemented.

Tuan and Park [9] utilized the predicted traffic level to modulate the linear increase/exponential decrease rate adaptation used by TCP congestion control. Yao [8] proposed the prediction-based random early detection. They all reported encouraging results of improved TCP throughput and reduced average delay.

The optimal predictor for the long-range dependent (LRD) traffic is an essential component of the prediction-based traffic control. Tuan and Park [9] proposed a heuristic predictor based on the conditional expectation of the future traffic level. The optimal predictor for the fractional auto-regressive integrated moving average (fARIMA(*p*,*d*,*q*)) model has been well studied [10]. However, for fBm, only a continuous-time predictor has been proposed by Gripenberg and Norros [11], which utilizes double integrals and is typically difficult to implement. This paper proposes an asymptotic linear minimum-mean-square-error (MMSE) predictor for the discrete-time fBm (dt-fBm) model.

The rest of the paper is organized as follows: Section 2 describes the asymptotic MMSE dt-fBm predictor; Section 3 discusses finite-length dt-fBm predictors for practical implementation; Section 4 compares the relative performance of various LRD predictors; Section 5 finally makes the conclusions.

## 2. OPTIMAL DT-FBM PREDICTOR

We first define the discrete-time fBm model.

**Definition 2.1:** Let a random process $Z_{n\Delta}$, $n \in Z$, have the following properties:

1) $Z_{n\Delta}$ has stationary increments;
2) $Z_0=0$, and $E\,Z_{n\Delta}=0$ for all $n$;
3) $EZ_{n\Delta}^2 = c_\Delta (n\Delta)^{2H}$, where $c_\Delta > 0$ is a constant depending on the sampling period $\Delta$;
4) $Z_{n\Delta}$ is Gaussian, i.e. all its finite-dimensional marginal distributions are Gaussian.

Then the process $Z_{n\Delta}$, $n \in Z$, is a discrete-time fBm process with time-scale $\Delta$. ■

It is easy to prove that $Z_{n\Delta}$, $n \in Z$, is second-order self-similar with the Hurst parameter $H$. $c_\Delta$ accounts to the fact that the self-similarity in real Internet traffic may vary with the timescale and therefore the variance coefficient as defined by Norros [12] for fBm model may also be different at different timescales. For example, it has been observed in [8] that at high traffic arrival rate, there is also self-similarity at small timescales (<100ms), but the correlation structure of small timescales is different from that of large timescales (>100ms).

The classical prediction results for Gaussian process can not be applied directly to dt-fBm because it does not satisfy the following assumptions: $Z_{n\Delta} = \sum_{i=0}^{\infty} b_i \varepsilon_{n-i}$ and $\sum_{i=0}^{\infty} a_i Z_{(n-i)\Delta} = \varepsilon_n$ with absolutely summable $\{b_i\}$, where $a_i$ and $b_i$ are constants, and $\varepsilon_i$ are i.i.d Gaussian random variables. However, dt-fBm has an important property as given in Theorem 2.1 [8], on which the optimal dt-fBm predictor can be constructed.

**Theorem 2.1**: If $Z_{i\Delta}$ is a dt-fBm process,

$$M_{m\Delta} \overset{def}{=} \begin{cases} 0, & m = 1 \\ \sum_{i=1}^{m-1} a_{m\Delta,i} \cdot [Z_{i\Delta} - Z_{(i-1)\Delta}], & m = 2,3,\ldots \end{cases} \quad (2.1)$$

where $a_{m\Delta,i} \overset{def}{=} \int_{(i-\frac{1}{2})\Delta}^{(i+\frac{1}{2})\Delta} u^{\frac{1}{2}-H} (m\Delta - u)^{\frac{1}{2}-H} du$, is a martingale having *independent increments* when $m \to \infty$. ■

The weighted sum of Gaussian random variables, $X_m = \sum_{i=1}^{m} b_i \varepsilon_i$ has the same distribution as $M_{m\Delta}$, where $\varepsilon_i$ are i.i.d Gaussian random variables with mean *0* and variance $\Delta$ and $b_i = \sqrt{c_1 \sqrt{(i\Delta)^{2-2H} - [(i-1)\Delta]^{2-2H}} / \Delta}$. Although $\{b_i\}$ are not absolutely summable, $\{ b_i^a \}$, $a > 4/(2H-1)$, are summable.

We define the optimal dt-fBm predictor as the one that minimizes the *h*-step prediction error $E\{(Z_{(n+h)\Delta} - \hat{Z}_{(n+h)\Delta})^2 \mid \xi_n\}$, where $\xi_n \overset{def}{=} (Z_{n\Delta}, \cdots Z_0)$ and $h \geq 1$. According to the definition of $M_{n\Delta}$, $n \geq 2$, the following relationships exist:

$$M_{n\Delta} - M_{(n-1)\Delta} = a_{n\Delta,n-1}\left(Z_{(n-1)\Delta} - Z_{(n-2)\Delta}\right) + \sum_{i=1}^{n-2} \left(a_{n\Delta,i} - a_{(n-1)\Delta,i}\right)\left(Z_{i\Delta} - Z_{(i-1)\Delta}\right) \quad (2.2)$$

$$Z_{n\Delta} - Z_{(n-1)\Delta} \overset{def}{=} \sum_{i=2}^{n+1} b_{n+1,i} \left(M_{i\Delta} - M_{(i-1)\Delta}\right) \quad (2.3)$$

Let

$$A_\Delta \overset{def}{=} \begin{pmatrix} a_{2\Delta,1}, & 0 & \cdots & 0 \\ a_{3\Delta,1} - a_{2\Delta,1}, & a_{3\Delta,2}, & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ a_{n\Delta,1} - a_{(n-1)\Delta,1}, & a_{n\Delta,2} - a_{(n-1)\Delta,2}, & \cdots & a_{n\Delta,n-1} \end{pmatrix}$$

and

$$B_\Delta \overset{def}{=} \begin{pmatrix} b_{2,2}, & 0 & \cdots & 0 \\ b_{3,2}, & b_{3,3}, & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ b_{n,2}, & b_{n,3}, & \cdots & b_{n,n} \end{pmatrix}.$$

Since $B_\Delta \cdot A_\Delta \cdot \vec{Z} = \vec{Z}$, where $\vec{Z} = (Z_\Delta - Z_0, \cdots, Z_{(n-1)\Delta} - Z_{(n-2)\Delta})^T$, $\{b_{n,i}\}$ are entries of $A_\Delta^{-1}$. Since $a_{n\Delta,n-1} > 0$, we have $b_{n,n} > 0$. Also since $A_\Delta$ is a triangular matrix, $A_\Delta^{-1}$ exists for $n < \infty$.

Thus, the optimal dt-fBm predictor can be obtained by directly computing $E\{Z_{(n+h)\Delta} \mid \xi_n\}$ using (2.3) and Theorem 2.1. The results are given in Theorem 2.2 [8].

**Theorem 2.2:** The *h-step* ($h \geq 1$) asymptotic MMSE predictor of dt-fBm can be represented as

$$\hat{Z}_{(n+h)\Delta} = \sum_{i=2}^{n+1} \left( \sum_{j=1}^{h} b_{n+j+1,i} \right) \left(M_{i\Delta} - M_{(i-1)\Delta}\right) + Z_{n\Delta} \text{ as } n \to \infty . \quad (2.4)$$

The corresponding MSE is

$$E\left\{ \left(Z_{(n+h)\Delta} - \hat{Z}_{(n+h)\Delta}\right)^2 \middle| \xi_n \right\} = \sum_{i=n+2}^{n+h+1} \left( \sum_{j=i-n-1}^{h} b_{n+j+1,i} \right)^2 c_1 \Delta^{2-2H} \left[ i^{2-2H} - (i-1)^{2-2H} \right]. \quad (2.5)$$

In particular, the *one step* dt-fBm predictor and its MSE are respectively,

$$\hat{Z}_{(n+1)\Delta} = \sum_{i=2}^{n+1} b_{n+2,i} \left(M_{i\Delta} - M_{(i-1)\Delta}\right) + Z_{n\Delta} \text{, as } n \to \infty , \quad (2.6)$$

$$E\left\{ \left(Z_{(n+1)\Delta} - \hat{Z}_{(n+1)\Delta}\right)^2 \middle| \xi_n \right\} = c_1 \left(b_{n+2,n+2}\right)^2 \left[(n+2)^{2-2H} - (n+1)^{2-2H}\right] \Delta^{2-2H} . \quad (2.7)$$

■

For each $h \in Z_+$, $\{b_{n+h,i}\}$ can be obtained by solving the following equations

$$A \cdot \begin{pmatrix} b_{n+h+1,n+h+1} \\ b_{n+h+1,n+h+1} - b_{n+h+1,n+h} \\ \vdots \\ b_{n+h+1,3} - b_{n+h+1,2} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (2.8)$$

where

$$A = \begin{pmatrix} a_{(n+h+1)\Delta,n+h}, & 0 & \cdots & 0 \\ a_{(n+h+1)\Delta,n+h-1}, & -a_{(n+h)\Delta,n+h-1}, & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ a_{(n+h+1)\Delta,1}, & -a_{(n+h)\Delta,1}, & \cdots & -a_{2\Delta,1} \end{pmatrix}.$$

Figure 1 shows the normalized error variance, $Var(Z_{(n+1)\Delta} - \hat{Z}_{(n+1)\Delta}) / E\{(Z_{(n+1)\Delta} - Z_{n\Delta})^2\}$, of the *one step* predictor as a function of *H*. The performance of the dt-fBm predictor is close to that of the continuous time fBm predictor derived by Gripenberg and Norros [11].

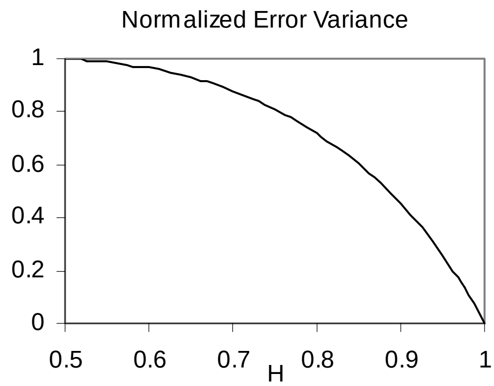


Figure 1: Normalized Error Variance vs. *H* (*n*=10000 and Δ=0.01)

## 3. FINITE-LENGTH DT-FBM PREDICTORS

The ideal dt-fBm predictor uses all the previous history to make prediction. However, in reality storage capacity is finite. It is impossible to keep on tracking the infinite history. This section discusses several methods for implementing finite length dt-fBm predictor.

The finite length prediction system can be created using sliding window (*SW*). In addition, we can use *N* latest increments at the same time scale, i.e. Δ, or alternatively at multiple time scales i.e. $\Delta_j \ (= m_j\Delta, m_j \in Z_+)$. We call the former, *single-scale (SS) scheme*, and the latter, *multi-scale (MS) scheme.*

Let $n_j$ ($1 \le j \le M$) denote the number of stored increments at the *j*-th time scale, where *M* is the total number of time scales and $\sum_{j=1}^{M} n_j = N$. The *MS* scheme keeps consecutive increments at several different time scales, and typically uses larger time scale to track older data, which allows it to track much longer history of the time series than the *SS* scheme with the same number of stored increments. The motivation for using larger time scale to track older data is the slow decaying of the coefficients $\{b_{n+h+1,j}\}$ at smaller *j* (See Figure 2 for an example of $b_{n+2,j}$ as a function of *j*). We assume $m_{j-1} > m_j$ and $m_N = 1$.

Let $d_{MS}$ and $d_{SS}$ denote the overall duration covered by *MS* and *SS* schemes, respectively. We have $d_{MS} = \sum_{j=1}^{M} n_j m_j \Delta$ and $d_{SS} = N\Delta$.

The *single-scale sliding window* (*SSSW*) predictor keeps on tracking the latest *w* increments at the same time scale, where *w* is the window size. The *h-step SSSW* predictor can be obtained from (2.4) with *n* set to *w*. With the varying time scale for different increments, the *multi-scale sliding window* (*MSSW*) predictor is more complex. We describe its details in the rest of the section.

Let $N_p = \sum_{j=1}^{p} n_j$, $t_p = \sum_{j=1}^{p} n_j \Delta_j$, where $1 \le p \le M$, $N_0 = 0$, $t_0 = 0$, and

$$d_i = \begin{cases} t_{p-1} + (i - N_{p-1})\Delta_p, & N_{p-1} < i \le N_p \\ t_M + (i - N)\Delta, & N < i \end{cases}. \quad (3.1)$$

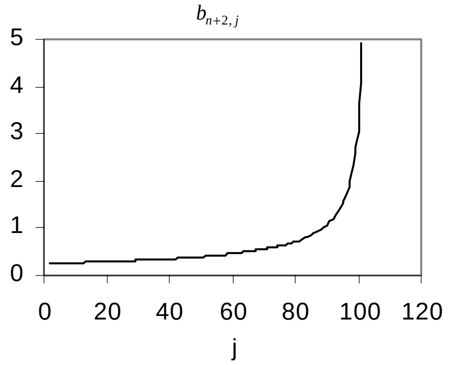


Figure 2: $b_{n+2,j}$ vs. *j* (n=100)

We have

$$M_i = \sum_{l=1}^{i-1} a_{d_i,l} X_l \text{, for } 1 \le i \le N, \quad (3.2)$$

where $X_l$ is the *l*-th increment at time-scale *q* for $N_{q-1} < l \le N_q$, and

$$a_{d_i,l} = \int_{d_l - \frac{\Delta_q}{2}}^{d_l + \frac{\Delta_q}{2}} u^{\frac{1}{2}-H} (d_i - u)^{\frac{1}{2}-H} du. \quad (3.3)$$

Then the *h-step* predictor becomes

$$\hat{Z}_{d_N + h\Delta} = \sum_{i=2}^{N+1} \left( \sum_{j=1}^{h} b_{N+j+1,i} \right) (M_i - M_{i-1}) + Z_{d_N}. \quad (3.4)$$

For each *j* ($j \ge 1$), $b_{N+j+1,i}$ can be obtained by solving the following equations:

$$A_{MS} \cdot \begin{pmatrix} b_{N+j+1,N+j+1} \\ b_{N+j+1,N+j+1} - b_{N+j+1,N+j} \\ \vdots \\ b_{N+j+1,3} - b_{N+j+1,2} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (3.5)$$

where

$$A_{MS} = \begin{pmatrix} a_{d_{N+j+1},N+j}, & 0 & \cdots & 0 \\ a_{d_{N+j+1},N+j-1}, & -a_{d_{N+j},N+j-1}, & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ a_{d_{N+j+1},1}, & -a_{d_{N+j},1}, & \cdots & -a_{d_2,1} \end{pmatrix}.$$

## 4. SIMULATION RESULTS

The performance of the fARIMA, dt-fBm and empirical predictors for LRD traffic is tested on several collected Internet traces. The traces were collected on an Internet backbone link (OC-48) using a monitoring device attached by an optical splitter. Packet timestamps have the granularity of 12.5ns. Detailed descriptions of the traces can be found in [8].

The duration of the traces ranges from 511 to 1352 seconds. We compute the time series of aggregate traffic within non-overlapped 10ms time intervals. The resulted time series are further divided into smaller time series each with a length of 2000. For each smaller time series, the multi-step forecasts up to *h*=8 steps given by each predictor are evaluated.

The empirical (*EMP*) predictor is implemented according to [9]. The *Type-I* (*FARIMAT1*) and *Type-II* (*FARIMAT2*) fARIMA predictors are implemented according to [10]. *FARIMA1* uses a truncation length of 100 while the optimal truncation length in *FARIMAT2* is determined by the Akaike information criterion). *FARIMAT2* is constructed based on fARIMA(1,d,1) where *d* is estimated using the Whittle's estimator.

For dt-fBm predictors, the following configurations are used in simulations:

- The *SSSW* predictor uses a window size of 100;
- The *MSSW* predictor uses three time scales (4Δ,2Δ,Δ) and the corresponding number of increments are (25,25,50).

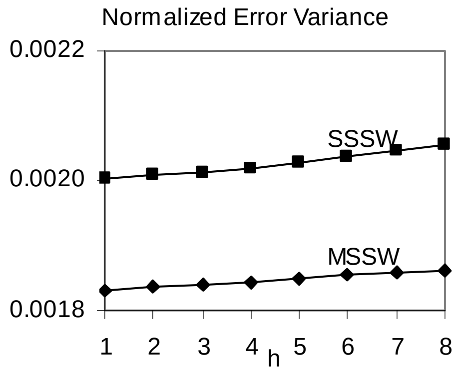


Figure 3: Normalized Error Variance of *SSSW* and *MSSW* Predictors

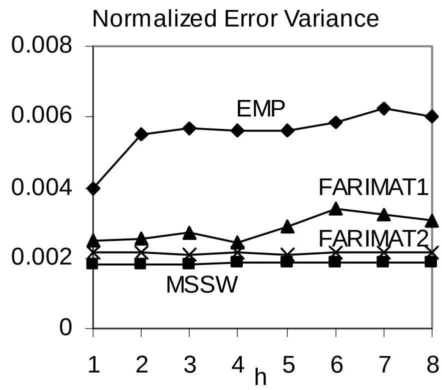


Figure 4: Normalized Error Variance of *EMP*, *FARIMAT1*, *FARIMAT2* and *MSSW* Predictors

Note that the optimal selection of the multiple time scales and the number of increments at each time scale needs further research.

Figure 3 shows a typical example of the normalized error variance of the two dt-fBm predictors with respect to the increment of the traffic arrival process $E\{(A_{(n+1)\Delta} - \hat{A}_{(n+1)\Delta})^2\} / E\{(A_{(n+1)\Delta} - A_{n\Delta})^2\}$, in which the tested time series has a Hurst parameter *H=0.868*. It turns out that the *MSSW* predictor has better performance than the *SSSW* predictor for multi-step forecasts. Figure 4 shows the normalized error variance of the *EMP*, *FARIMAT1*, *FARIMAT2* and *MSSW* predictors for the same time series as used in Figure 4. As shown in the figure, the *MSSW* predictor also has better performance than other classical LRD predictors.

## 5. CONCLUSIONS

This paper derives an asymptotically optimal multi-step predictor for a special model of long range dependent traffic, discrete-time fractional Brownian motion (dt-fBm). The ideal predictor has infinite length. For the purpose of practical implementation, two types of finite-length dt-fBm predictor, i.e. the single-scale sliding window (*SSSW*) predictor and the multi-scale sliding window (*MSSW*) predictor, have been proposed as approximations to the optimal predictor. Simulations have been performed to compare the relative performance of different types of dt-fBm predictor and other predictors, i.e., empirical (*EMP*), *Type I* fARIMA (*FARIMAT1*) and *Type II* fARIMA (*FARIMAT2*) predictors. Simulation results show that *MSSW* is not only the better dt-fBm predictor but also more accurate than all other classical LRD predictors.

## 11. REFERENCES

[1] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of Ethernet Traffic," *Proc. ACM SIGCOMM'93*, pp.183-193, 1993.

[2] W. Willinger, M.S. Taqqu, R. Sherman and D.V. Wilson, "Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level," *IEEE/ACM Trans. on Networking*, Vol. 5, No. 1, Feb. 1997.

[3] V. Paxson and S. Floyd, "Wide-area traffic: the failure of Poisson modeling," *Proc. SIGCOMM'94*, Sept. 1994.

[4] J. Cao, W. S. Cleveland, D. Lin, D. X. Sun, "Internet Traffic Tends toward Poisson and Independent as the Load Increases", *Nonlinear Estimation and Classification*, C. Holmes, D. Denison, M. Hansen, B. Yu, B. Mallick - editors, Springer, 2002.

[5] J. Beran, R. Sherman, and W. Willinger, "Long-range dependence in VBR video traffic," *IEEE Trans. on Comm*, Vol. 43, No. 2/3/4, 1566-79, 1995.

[6] D. Duffy, A. McIntosh, M. Rosenstein, and W. Willinger, "Statistical analysis of CCSN/SS7 traffic data from working CCS subnetworks," *IEEE JSAC*, Vol. 12, 544-551, Apr. 1994.

[7] M. E. Crovella and A. Bestavros, "Self-similarity in World Wide Web traffic: evidence and possible causes," *IEEE/ACM Trans. On Networking*, Vol. 5, No. 6, pp.835-846, Dec. 1997.

[8] L. Yao, "Queuing Analysis and Control of Long Range Dependent Traffic – Applications to Internet Traffic Engineering," Technical Report, Department of Electrical and Computer Engineering, the George Washington University, Oct. 2002.

[9] T. Tuan and K. Park, "Congestion control for self-similar network traffic," *Performance Evaluation*, 36:359-386,1999.

[10] R. J. Bhansali and P. S. Kokoszka, "Prediction of long-memory time series: an overview," *Estadistica* 2001.

[11] G. Gripenberg and I. Norros, "On the prediction of fractional Brownian motion," *Technical Report*, 1995.

[12] I. Norros, "On the use of fractional Brownian motion in the Theory of Connectionless Networks," *IEEE JSAC*, Vol. 13, No. 6, Aug. 1995.