

# Time Series Models for Internet Traffic\*

*Sabyasachi Basu*  
Department of Statistical Science  
Southern Methodist University  
Dallas, TX 75275

*Amarnath Mukherjee*  
College of Computing  
Georgia Institute of Technology  
Atlanta, GA 30332

*Steve Klivansky*  
College of Computing  
Georgia Institute of Technology  
Atlanta, GA 30332

FIRST DRAFT: July 31, 1995  
LAST REVISED: February 29, 1996

**GIT-CC-95-27**

## Abstract

Data traffic sequences from two campus FDDI rings, an Ethernet, two entry/exit points of the NSFNET, and sub-sequences belonging to popular TCP port numbers on one of the FDDI rings indicate that appropriately differenced time-series generated from these traces can be modeled as Auto-Regressive-Moving-Average (ARMA) processes. The variates of the ARMA filter are, however, non-Gaussian.

A sequence of steps leading through (i) parameter estimation, (ii) generating the distribution of the variates, (iii) forecasting tail percentiles, and (iv) synthetic generation of non-negative integer sequences is presented. The data indicates that parameter estimates drift slowly with time and may need to be re-computed periodically for accurate forecasts. The forecasting algorithm has potential application in dynamic resource allocation. The synthetic traffic generation algorithm may be used in simulation studies of resource management algorithms.

---

\*Extended version of a paper to be presented at the IEEE Infocom Conference, San Francisco, 1996. The work of A. Mukherjee and S. M. Klivansky was supported by the National Science Foundation under grant NCR-9396299.

# 1 Introduction

## Overview

Properties of network traffic have been the subject of considerable recent interest, see for example [2, 3, 4, 6, 8, 9, 10]. The objective of this paper is to investigate parametric time-series models that can be used for forecasting aggregate data traffic. The primary result is an algorithm that predicts the  $u^{th}$  quantile of the distribution of data traffic given past history. We also report, to a limited extent, on algorithms for generating synthetic traffic based on the time-series models developed.

Let  $\{Y_1, Y_2, \dots, Y_t, \dots\}$  be the number of packets observed over adjacent time-intervals, each interval being of length  $T$ . The key observation on which the models are based is that appropriate differenced sequences  $\{W_t = Y_t - Y_{t-s}\}$ , ( $s \geq 1$ ), are stationary processes with short memory, and can as such be modeled as Auto-Regressive Moving-Average (ARMA) processes with orders  $(p, q)$ , where  $p$  and  $q$  are small integers. For most of the datasets we have studied,  $s$  was equal to one or two. For two datasets, its value was eight for one time interval ( $T$ ), indicating a non-random periodicity of length 8, but one at a smaller (and also larger) time interval. A complete list of datasets studied are listed in Tables 1 and 2. Spectral densities of the two datasets listed in Table 3, however, show quite a few non-random periodicities, and are an exception to the above rule.

By way of example, consider the autocorrelation function (ACF) of two of the datasets. First, in Figure 1(a), we see the ACF of a trace over the San Diego external interface to the NSFNET (ENSS), studied in [3], with time interval,  $T$ , set to 0.1 seconds. In Figure 1(b), the ACF of  $\{Y_t - Y_{t-1}\}$  is seen to have only a few significant auto-correlations left; these were found to be modeled by a non-Gaussian Moving Average Process of order  $q = 1$ .

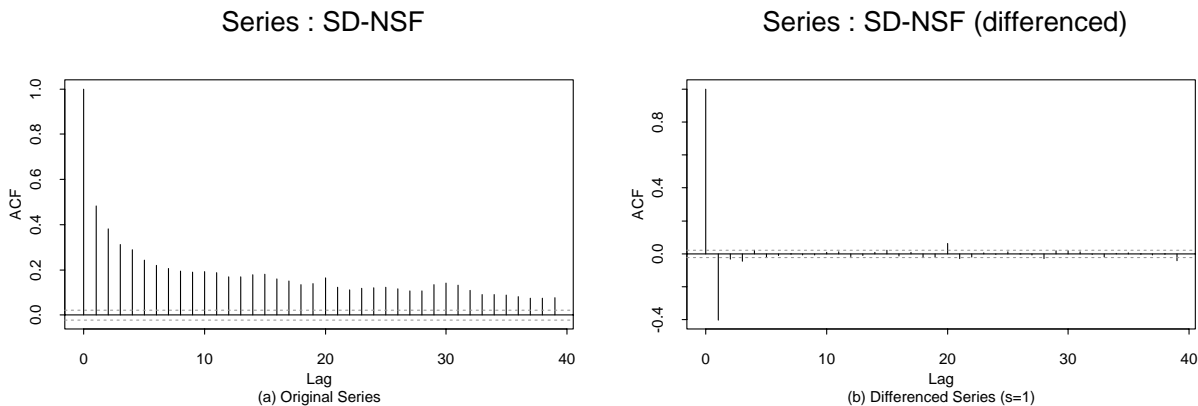


Figure 1: Autocorrelation function for the ENSS trace, SD-NSF.  $T = 0.1$  seconds; total length used = 15 minutes. (a) ACF of the sequence  $\{Y_t\}$ . (b) ACF of the sequence  $\{W_t = Y_t - Y_{t-1}\}$ .

In Figure 2(a), the ACF of the aggregate traffic, GT-I, over the Georgia Tech campus FDDI ring is shown. The corresponding ACF for  $\{Y_t - Y_{t-1}\}$ , shown in Figure 2(b), is seen to have oscillating positive and negative correlations into far lags. However,  $\{Y_t - Y_{t-2}\}$  (Figure 2(c) ) shows correlations into only a few lags, indicating that (i) there is a non-random periodicity of 2 in

this data, and that (ii) the resulting differenced process can be potentially modeled by a ARMA filter. Two models were found to fit this data equally well — an ARMA(2,2) model and an MA(6) model.

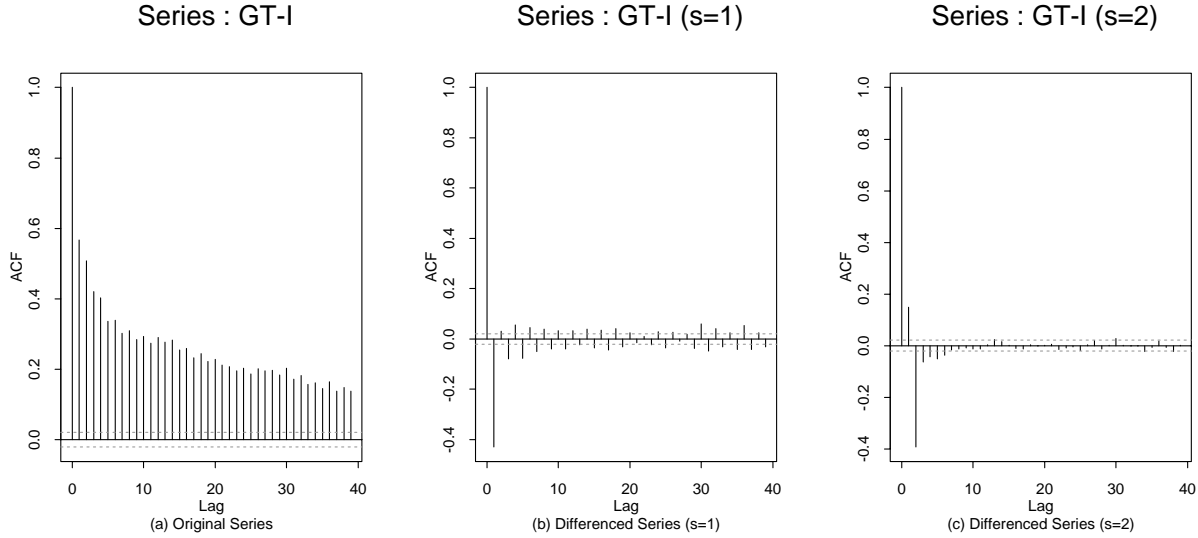


Figure 2: Autocorrelation function for the FDDI trace (GT-I).  $T = 0.1$  seconds; total length used = 15 minutes. (a) ACF of the sequence  $\{Y_t\}$ . (b) ACF of the sequence  $\{W_t = Y_t - Y_{t-1}\}$ . (c) ACF of the sequence  $\{W_t = Y_t - Y_{t-2}\}$ .

Differencing makes a non-stationary series stationary when there is a drift in the data. However, not all classes of non-stationarities admit a simple differencing strategy. The data listed in Table 3, for example, show quite a few non-random periodicities, and do not admit to this modeling strategy.

The Ethernet datasets, BC-I and BC-II, have also been shown to possess the properties of a second-order self-similar process — which is defined as a wide-sense stationary process (constant, time-independent, mean, variance and auto-covariance function) possessing long-memory (sum of the auto-correlations diverge), and further, the auto-correlation function is the same for all time scales over which the data is aggregated. These results have been shown to hold for Ethernet data [8, 13], and have subsequently led to insights into the burstiness properties of individual TCP connections that constitute aggregate traffic [6, 9] — based on processes that are known to be second-order self-similar. For instance, the distribution of connection lifetimes and number of packets transmitted are both heavy-tailed, the latter being close to a Pareto distribution:  $F(x) = 1 - x^{-\alpha}$ , with  $1 < \alpha < 2$ , (which leads to a finite mean and infinite variance; in practice the estimated variance can never be infinite for finite datasets, but they are large).

On the other hand, the models discussed in this paper are adequate and appropriate for the purposes of forecasting, and provide an alternative view of the data. However, no model is true reality, and its usefulness lies in ones ability to ask and obtain answers to questions of interest from the model. Alternative models and paradigms allow for answering different questions on the same real phenomenon.

In this specific case, the theoretical assumptions for the two models are different. Second-order self-similar models are defined for the family processes that are wide-sense stationary. The correlations for these processes extend into far lags, and by definition, the sum of their autocorrelations diverge. ARMA models for the differenced sequences, on the other hand, result in the original se-

Location	Trace Label	Start Time	Duration(sec)	#Packets
Campus FDDI Ring				
Georgia Tech	GT-I	5/16/94, 11:00	941.8	999955
Georgia Tech	GT-II	5/16/94, 13:00	4611.5	4999980
Georgia Tech	GT-III	5/16/94, 15:00	3127.5	3621155
Georgia Tech	GT-IV	5/16/94, 17:00	4357.7	4999922
San Diego Internal FDDI Ring	SD-FDDI	6/28-29/93, 14:00	3657.4	3736356
NSFNET External Interface (ENSS)				
San Diego	SD-NSF	3/23/93, 14:00	1040.6	2982851
NCSA	NCSA	3/25-29/93, 14:11	2043.0	3822950
Ethernet				
Bellcore-Aug89	BC-I	8/29/89, 11:25	3142.8	999999
Bellcore-Oct89	BC-II	10/5/89, 11:00	1759.6	999999

Table 1: Summary of traffic data studied. Data on the San Diego FDDI ring and the NSFNET ENSS's provided by Hans Werner Braun, San Diego Supercomputer Center. A detailed description of these datasets appear in [3]. Ethernet data provided by Will Leland, Bellcore. A detailed description of these datasets appear in [7, 8, 13]. Data on the Georgia Institute of Technology campus FDDI ring collected locally by the authors.

Trace	FTP-data	NNTP	SMTP	HUMAN	WWW
GT-I	69551	201995	49637	58745	6759
GT-II	390434	854032	195214	326875	97720
GT-III	198841	1060947	129531	230226	70124
GT-IV	521495	1006932	122681	253978	39360

Table 2: Total number of packets for popular TCP protocols over Georgia Institute of Technology campus FDDI backbone.

Location	Duration(sec)	#Packets
Cornell-ENSS	1982.2	3036800
Pittsburgh-ENSS	1525.2	2429182

Table 3: Traffic data that exhibit numerous periodicities. These two datasets point to open issues that remain to be studied. Data provided by Hans Werner Braun, San Diego Supercomputer Center.

quence to be non-stationary. In practice, whether the mean, the variance and the auto-covariance are all stationary or non-stationary is often difficult to determine. Both classes of models appear to be useful, and although, theoretically, stationarity and non-stationarity are exclusive of one another, we believe that they ought to be considered inclusively in modeling repertoire (based on the results obtained from the two classes on the same datasets).

In other related work,

- Claffy et al have studied flow-parameterization and traffic characteristics of the T1 NSFNET backbone [3, 4].
- Danzig et al [5] have developed a traffic library for TCP conversations based on empirical burst distributions they studied in [2].
- Klivansky et al have studied the impact of different factors on potential long-range dependence in aggregate TCP traffic over NSFNET core switches, see [6].
- Paxson has developed statistical distributions for TCP burst distributions, see [10].
- Paxson and Floyd have reported significant dependence among conversation arrivals for specific TCP protocols (e.g., for FTP-data and NNTP conversations), and independence among others (e.g., for TELNET and FTP-control conversations). They also showed that Danzig et al’s TCP traffic library produces similar variance-time plots as measured TCP traffic, see [9].
- Willinger et al have explained the the presence of self-similarity in traffic auto-correlation structure at multiple time-scales through a heavy-tailed distribution of source on-periods.

## Summary of Results

The following is a summary of results:

1. Traffic data listed in Tables 1 and 2, with the exception of the FTP-data traces, can be modeled as multiplicative ARIMA processes, i.e., as ARMA processes after appropriate differencing. Spectral density of data listed in Table 3 show quite a few non-random periodicities, and are as such beyond the scope of this paper. Therefore, it is important to recognize that while many datasets fit the multiplicative ARIMA models, there are exceptions as well.

The value of  $s$  in  $\{Y_t - Y_{t-s}\}$  is one for the the SDSC FDDI ring, the ENSS traces, and constituent TCP traces SMTP and WWW over GT FDDI ring. For the Bellcore traces,  $s$  is 1 for some time intervals (e.g.,  $T = 0.01\text{sec}$ ), but for  $T = 0.1\text{sec}$ ,  $s$  is 8 because underlying non-random periodicities appear at this time-scale.  $s$  is equal to two for the aggregate traffic over GT FDDI ring, and constituent TCP traces belonging to NNTP and HUMAN (aggregate of TELNET, RLOGIN and FTP-control) traffic.

The GT and Bellcore traces were available over different time periods. Comparison of models for them show that (i) the model forms that fit these traces across time are similar, and (ii) the corresponding optimal parameter estimates are not too far apart. This sample size is, however, too small to make generalizations. Model forms across networks were found to be different.

2. The white noise process that drives the Moving Average filter is, in most cases, non-Gaussian. However, if the parameter estimates and the model form are available, they can be computed

directly from the data. Since the distribution of these variates is not known a-priori, parameter estimates for the multiplicative ARIMA models are obtained through a non-linear least squares optimization. Typically, one assumes a known form for the distribution (usually from the exponential family of distributions) and computes estimates for the parameters that maximize the likelihood of these observations. In this paper, parameter estimates are obtained first, and subsequently, advances in computing and storage facilities are used to bear on the problem of computing the distribution of the variates that make up the moving average filter.

3. Quantiles of the residuals obtained in (2), in conjunction with the model form derived from the data are used for forecasting tail percentiles of future data values.
4. Preliminary investigation on generating traffic sequences from the time-series models developed shows that a simple discrete summation can lead to sequences that do not resemble the original traces. The problem here is similar to noise introduced in numerical integration — unless special action is taken, the noise can get amplified over time. Preliminary algorithms for noise-suppression and their pros-and-cons are presented. These algorithms also maintain the constraint that  $Y_t^l$ s are non-negative integers.

## Outline

Section 2 presents preliminary processing steps and identification of plausible model forms. Section 3 briefly describes Multiplicative Auto-Regressive Integrated Moving Average models. Section 4 describes the parameter estimation method and the models that appear to fit best. The residual distributions are also generated in this section. Section 5 discusses forecasting the  $u^{th}$  percentile given the known residual quantiles. Section 6 discusses synthetic traffic generation and Section 7 presents concluding remarks and directions of future research.

## 2 Preliminary Processing and Model Identification

The datasets studied have similarities and differences. The following heuristic tests provide some guidelines for (i) constraining the search space of models considered for a given dataset, and (ii) identifying non-random periodicities and their relationship with time intervals chosen.

In Section 1, we have seen that for SD-NSF and GT-I, the ACF of differenced sequences  $\{Y_t - Y_{t-s}\}$  damped out quickly with  $s = 1$  and  $s = 2$ , respectively. The Ethernet datasets, BC-I and BC-II, however, point to the presence of a non-random periodicity at  $s = 8$  exhibited at a time scale of  $T = 0.1\text{sec}$ . These disappear for many other time scales. Therefore, data sequences needs to be studied for a range of time intervals, at least at the preliminary stage.

Figure 3 shows the ACF of the trace BC-I for two successive 15 minute intervals from the start of the trace with  $T = 0.1\text{sec}$ . There appears to be a period of length approximately 8 time units, corresponding to  $0.8\text{sec}$ . This is confirmed with a cumulative periodogram plot (see [1, pp 321-324]):

$$C(f_j) = \frac{\sum_{i=1}^j I(f_i)}{n\sigma^2},$$

where  $n$  is the sample size, and  $I(f_i)$  is the amplitude at frequency  $f_i = i/n$ :

$$I(f_i) = \frac{2}{n} \left[ \left( \sum_{t=1}^n Y_t \cos 2\pi f_i t \right)^2 + \left( \sum_{t=1}^n Y_t \sin 2\pi f_i t \right)^2 \right].$$

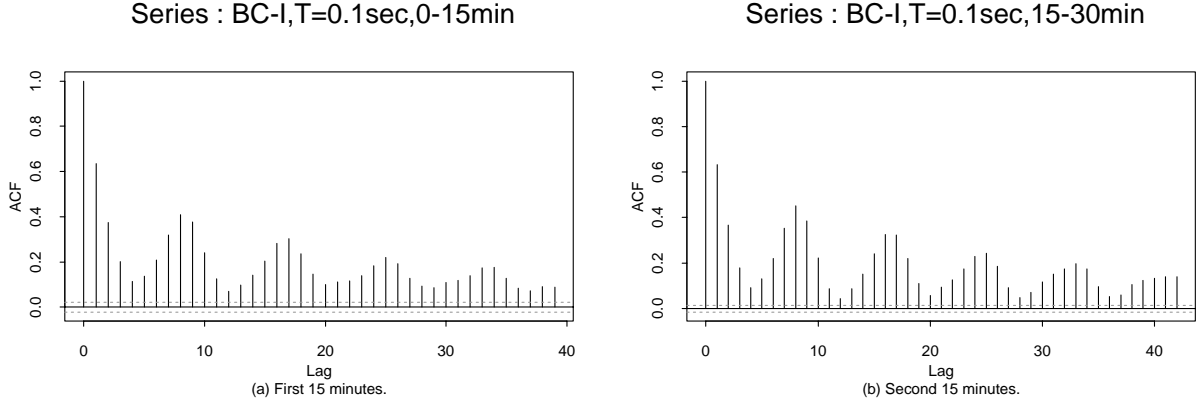


Figure 3: ACF of Trace BC-I, with  $T = 0.1\text{sec}$ . (a) First 15 minutes. (b) Second 15 minutes.

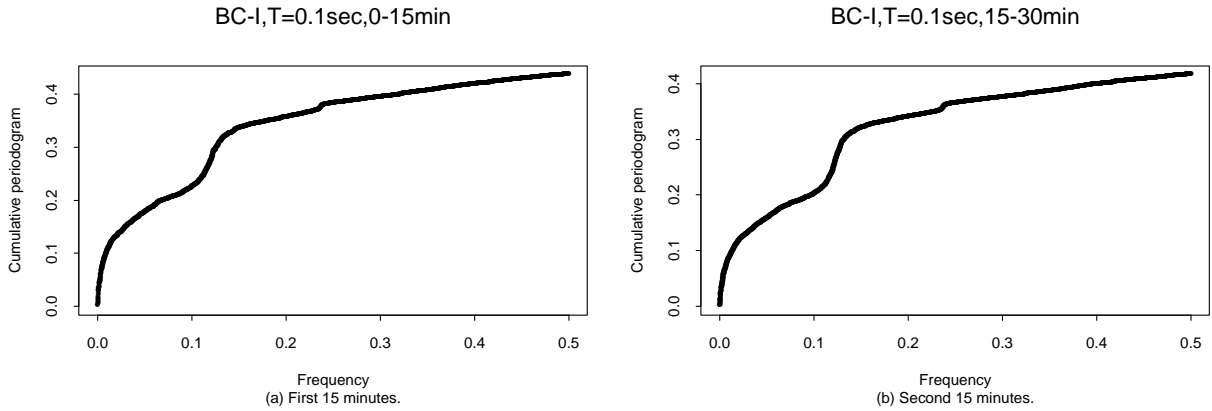


Figure 4: Cumulative periodogram of Trace BC-I, with  $T = 0.1\text{sec}$ . (a) First 15 minutes. (b) Second 15 minutes.

For a process with a non-random periodicity at frequency  $f_j$ ,  $C(f_j)$  will show a ‘bump’ at  $f_j$  — which we indeed see for Trace BC-I at  $f_j$  slightly greater than 0.1, the corresponding period being,  $1/f_j \approx 8$  (see Figure 4).

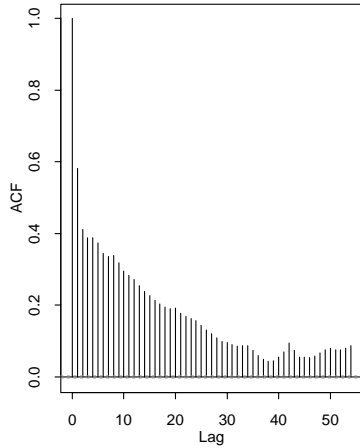
Next, we consider the effect of  $T$ . Plots 5(a)-(c) show the ACF’s of  $\{Y_t\}$  and  $\{Y_t - Y_{t-1}\}$ , and the cumulative periodogram of  $\{Y_t\}$  for  $T = 0.01\text{sec}$ . Corresponding figures for  $T = 0.8\text{sec}$  are plotted in Figures 5(d)-(f). We observe that (i) the ACF’s damp out for  $\{Y_t - Y_{t-1}\}$  beyond the first few lags, and (ii) the cumulative periodogram for  $\{Y_t\}$  has no abrupt bumps. They do have a sharp increasing trend at lower frequencies — which is to be expected from the slowly decaying ACF of  $\{Y_t\}$ . For  $T = 0.01\text{sec}$ , the periodicities seen for  $T = 0.1\text{sec}$  disappear — most likely because they get sufficiently randomized. For  $T = 0.8\text{sec}$ , successive 8 intervals are lumped together, making the periodicity equal to one.

Finally, Figure 6 shows the ACF of  $\{Y_t - Y_{t-8}\}$  for Trace BC-I for  $T = 0.1\text{sec}$ . The ACFs appear to show a short memory. However, there are still quite a few non-negligible auto-correlations, and these do not lend well to a parsimonious model fit — in our experiments, either the models considered did not converge, or if they did, had error residuals that were not completely uncorrelated.

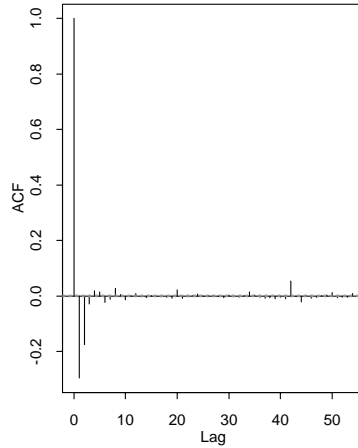
Series : BC-I, T=0.01sec, 0-15min

Series : differenced series, s=1

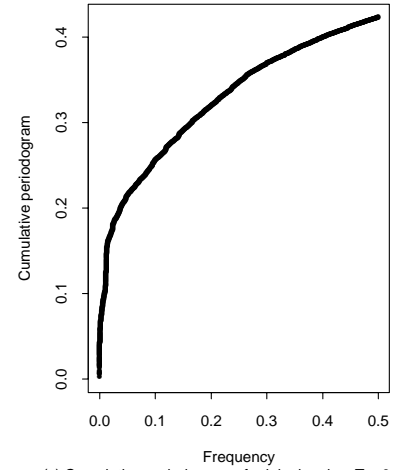
BC-I, T=0.01sec, 0-15min



(a) ACF of original series, T = 0.01sec



(b) ACF of differenced series, T = 0.01sec

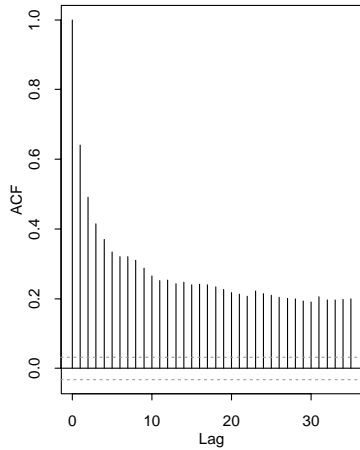


(c) Cumulative periodogram of original series, T = 0.01sec

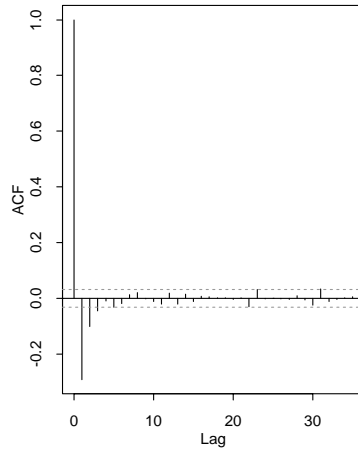
Series : BC-I, T=0.8sec, 0-15min

Series : differenced series, s=1

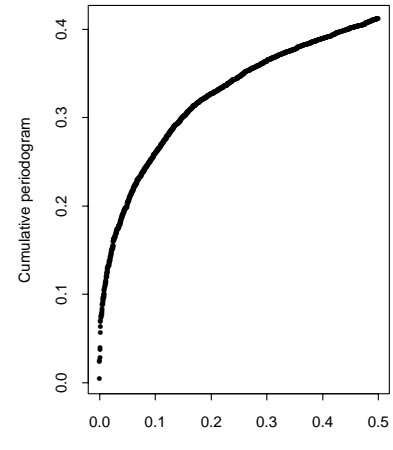
BC-I, T=0.8sec, 0-15min



(a) ACF of original series, T = 0.8sec



(b) ACF of differenced series, T = 0.8sec



(c) Cumulative periodogram of original series, T = 0.8sec

Figure 5: ACF and cumulative periodogram of Trace BC-I (first 15 minutes) for different values of  $T$ . (a) ACF of  $\{Y_t\}$ ,  $T = 0.01\text{sec}$ . (b) ACF of  $\{Y_t - Y_{t-1}\}$ ,  $T = 0.01\text{sec}$ . (c) Cumulative Periodogram of  $\{Y_t\}$ ,  $T = 0.01\text{sec}$ . (d) ACF of  $\{Y_t\}$ ,  $T = 0.8\text{sec}$ . (e) ACF of  $\{Y_t - Y_{t-1}\}$ ,  $T = 0.8\text{sec}$ . (f) Cumulative Periodogram of  $\{Y_t\}$ ,  $T = 0.8\text{sec}$ .

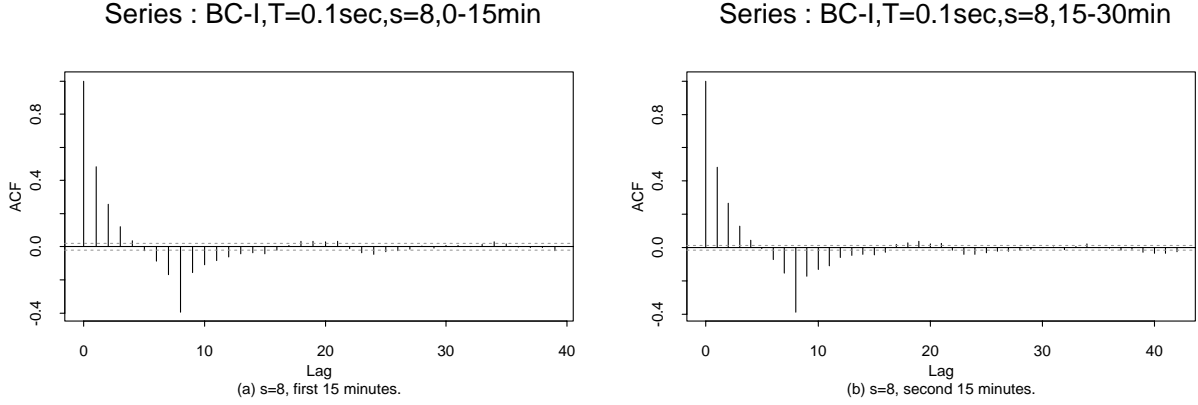


Figure 6: ACF of  $\{Y_t - Y_{t-8}\}$  for Trace BC-I, with  $T = 0.1\text{sec}$ . (a) First 15 minutes. (b) Second 15 minutes.

Models with  $T = 0.01\text{sec}$ , however, led to quick convergence (e.g., 3 iterations for the above dataset with starting values quite far from the optimal solution), and the residuals were white noise.

In summary, the following heuristic steps appear to be useful in identifying the model form:

- plotting the ACF for different values of  $T$ ;
- studying the ACF of differenced sequences,  $\{Y_t - Y_{t-s}\}$  for different values of  $s$ ;
- if periodicity is suspected, (e.g.,  $s > 1$ ), looking for non-random periodicities with the help of a cumulative periodogram.

### 3 Background on Autoregressive Integrated Moving Average Models

The data are collected over time and we analyze the data using time-series analysis. The autoregressive moving average (ARMA) series models were popularized by Box and Jenkins[1] for purposes of modeling time series data. Letting  $W_t$  denote the observations at time  $t$  from some stationary process with zero mean, an ARMA model of order  $(p, q)$  for  $W_t$ , denoted as ARMA( $p, q$ ) has the form

$$W_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (1)$$

where  $a_t$  is white noise (uncorrelated) and  $\phi_j$ 's and  $\theta_j$ 's are real constants. Define a lag operator  $B$  as  $BW_t = W_{t-1}$ . The model in (1) can also be written using the operator  $B$  as  $\phi(B)W_t = \theta(B)a_t$  where  $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$  and  $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$ . The model is stationary if the roots of the  $\phi(B) = 0$  are greater than one in absolute value. These models are also extended to model non-stationary behavior in the data, and the models that have proved useful in practice are the Autoregressive Integrated Moving Average Models, ARIMA( $p, d, q$ ), with  $d$ , an additional differencing parameter such that

$$W_t = (1 - B)^d Y_t = (1 - B)^{d-1} (Y_t - Y_{t-1})$$

is stationary and follows an ARMA( $p, q$ ) model.

If the data have a seasonal component, multiplicative models of the form

$$\phi(B)\Phi(B^s)W_t = \theta(B)\Theta(B^s)a_t \quad (2)$$

are often used to model the data. The operators  $\Phi(B)$  and  $\Theta(B)$  are of order  $P$  and  $Q$ , respectively. Here  $W_t = (1 - B)^d(1 - B^s)^D$  and the model is described as an ARIMA( $p, d, q$ )  $\times$  ( $P, D, Q$ ) $_s$  model.

A special case of these models, the ARIMA(0,1,1), gives rise to the exponentially weighted moving average algorithm for forecasting the expected value  $\hat{Y}_{t+1}$  given  $Y_t, Y_{t-1}, \dots$ . This model form has had many applications; for instance, in networking, it is used to keep track of round-trip delay estimates in the TCP protocol.

The ARIMA models are linear time series models. Another class of models which are useful are the non-linear class of time series models ([11]). These models might be more appropriate if the variability in the data varies with time and/or the mean level. Besides, for different segments the model might be different and non-linear models are more appropriate than the linear models. However, in this work we will not discuss nonlinear models and we will use the class of multiplicative models to study the datasets.

## 4 Nonlinear Estimation of Model Parameters

The step after identifying a particular ARIMA model from the class of multiplicative models given in (2) is to estimate the parameters of the model, namely,  $\phi = (\phi_1, \dots, \phi_p)'$ ,  $\Phi = (\Phi_1, \dots, \Phi_P)'$ ,  $\theta = (\theta_1, \dots, \theta_q)'$ , and  $\Theta = (\Theta_1, \dots, \Theta_Q)'$ . There are basically two methods available for estimating the parameters in (2), the least squares method and the maximum likelihood estimate.

To use the method of the maximum likelihood, the parametric distribution of the noise  $a_t$  needs to be specified. Usually, the Gaussian distribution is assumed. In the case of Internet traffic data, the assumption of Gaussianity does not lend itself, so we used the least squares method instead. Under this method, one chooses the values of the parameters which minimize the sum of squares of the squared residuals. Hence the estimators of the parameters,  $\phi$ ,  $\Phi$ ,  $\theta$  and  $\Theta$  are the ones which minimizes

$$S(\phi, \Phi, \theta, \Theta) = \sum_{t=1}^n a_t^2. \quad (3)$$

Note that (3) is nonlinear in the parameters if the model has a moving average component. The other difficulty that arises in estimating the parameters is that the starting values for the data,  $W_0, W_{-1}, \dots, W_{1-(p+sP)}$  and that for the errors  $a_0, a_{-1}, \dots, a_{1-(q+sQ)}$  are needed. The second problem can be addressed in two different ways — (i) by assuming that the initial values of the data are known and that the errors prior to this time are all zero, one minimizes the sum of squares starting at time  $p + sP + 1$ ; or (ii) by using the method of forecasting the errors in the past and then using these values to calculate the errors in the time 1 through  $n$ . Since the datasets are four or more orders of magnitude longer than the missing values, the results of the two methods can be expected to be similar. To analyze the traffic data, we used the subroutine G13AFF of the Numerical Algorithms Group (NAG) subroutine, which uses the forecasting approach. These subroutines can be interfaced with Fortran or C programs to estimate the parameters of ARIMA models.

A result of these computations is that, for the optimal parameter set  $(\phi^*, \Phi^*, \theta^*, \Theta^*)$  thus obtained, one obtains the corresponding residuals,  $\{a_t\}$ , and their empirical distribution (specifically, tail percentiles) can be directly used in forecasting tail percentiles of the data as discussed in Section

5. Since the cost of under-forecast is high (potentially dropped packets), forecasting tail percentiles for generally distributed residuals may have applications.

The model forms fit to the datasets appear in the Appendix. These were obtained using the following steps.

- (a) Determining a set of potential values for  $s$  with  $T$  fixed. The objective was to ensure that  $W_t = Y_t - Y_{t-s}$  had only a small number of significant auto-correlations.
- (b) Based on the ACF and the partial correlation function of  $W_t$ , identifying a set of ARMA( $p, q$ ) model forms. See [1, Ch. 6].
- (c) Obtaining parameter estimates for the alternatives selected by minimizing (3).
- (d) Checking the ACF of  $\{a_t\}$  for each alternative to ensure that it is white noise (i.e., uncorrelated). In practice, small correlations may remain for a few lags, even for the best fits available for a dataset. See Figure 7 for some examples.

If  $\{a_t\}$  were uncorrelated and Gaussian with sample size  $n$ , the estimated auto-correlation at lag  $k$ , denoted  $r_k$ , may be assumed to have a standard error bounded above by  $n^{-1/2}$  for larger lags, see [1, Ch. 8]), and if the model is appropriate,

$$Q = n \sum_{k=1}^K r_k^2$$

is approximately  $\chi^2(K - p - q)$  distributed for  $K$  sufficiently large. If, however, the model is inappropriate,  $Q$  is inflated. Therefore, the value of  $Q$  may be compared against quantiles of  $\chi^2$  distribution to determine model adequacy.

If the Gaussian assumption for  $\{a_t\}$  is not justified, (as was usually the case for the data), the above test for model adequacy does not apply. We have used plotting and visualization of the ACF of  $\{a_t\}$  to determine if any of its  $r_k$ 's are large, and if not, we have accepted the model form as adequate. While this is a heuristic approach, it was able to filter out non-acceptable models. Figure 7 shows example ACFs of  $\{a_t\}$  for models that we accepted.

If residual correlations do remain, they often suggest potential changes in the model form for further experimentation. For example, an extra moving-average component may be added to an ARMA model for each  $r_k$ , if the number of such correlations are small. Steps (c) and (d) are then repeated with the new model form. This approach is, however, not recommended if the number of  $r_k$ 's is large, because the optimization algorithm may not converge rapidly, or not at all, if the number of model parameters become large.

- (e) Among the models with approximately white-noise  $\{a_t\}$ 's, the one with smallest error variance, or equivalently, the smallest value for the right hand side of (3), is selected as the model for the data.

Experiments with plausible models showed that for most datasets, there existed models with  $\{a_t\}$ 's close to white noise (no significant  $r_k$ 's). However, for a few models, there did remain some small auto-correlations in the residuals. In Figure 7, the left column shows the ACFs of the  $\{a_t\}$ 's, and the right column shows their density functions. The series name and the seasonal ARIMA model form are shown at the top of each figure. The numbers represent ( $pdq - PDQ - s$ ). Corresponding parameter estimates are given in the Appendix. A long tail in the density function was not an uncommon occurrence. Aggregate FTP-data traffic was quite bursty and appropriate

ARIMA models were not found. The models reported in the appendix are for cases when we believed  $\{a_t\}$  were uncorrelated, or nearly so.

Based on the models in the Appendix, we observe the following.

- Seasonal ARIMA models fit a fairly large class of Internet traffic. Their exact forms and parameter estimates differ across networks (and across TCP protocols).
- For traffic over the same network, the degree of differencing,  $s$ , remained the same for all models fit to a given data-set for fixed  $T$ . (See the GT, BC and TCP protocol models.) The sample size is, however, too small to make generalizations. It is possible that non-random periodicities are due to some deterministic events, e.g., chron events of some subset of applications using the network, or some timing property of the network itself.
- The alternate model forms have a similar structure for a given network. For example, all GT aggregate traces are either MA(6) or ARMA(2, 2). Further, their parameter estimates are not too far from each other over different time-periods. Here too, the spatial sample-size is small, and generalizations are not appropriate.
- For GT traces, NNTP models have approximately the same form as aggregate traffic. Most likely, this is because it dominates the volume of traffic seen in each trace (see Tables 1 and 2).

Also, while aggregate FTP-data does not lend itself to an accurate multiplicative ARIMA model, the aggregate GT backbone FDDI traces which contain FTP-data packets do lend themselves to such models. Most likely, this is because the volume of FTP-data packets is small (5.49% – 10.42% in the four data-sets) relative to the volume of aggregate traffic.

For new traffic traces where seasonal traffic models may apply, one would need to investigate the model form, obtain parameter estimates, and depending on what it is to be used for, either (i) store selected percentiles of the error distribution for forecasting tail percentiles (Section 5) or (ii) store the entire empirical distribution of  $\{a_t\}$  for use in traffic generators for simulation (Section 6). If the model form at a multiplexor is known (e.g., does not change appreciably with time), parameter estimates may be recomputed on-line (if desired) for forecasting applications with previous data seen at the multiplexor.

The minimization in (3) will take some compute time, so re-fitting, if implemented online (e.g., in forecasting applications), should be spaced out appropriately. The number of iterations needed by the non-linear optimizer was usually fairly small when the model form was appropriate. For example, with initial values of all parameters set to 0.1, which usually was far from their optimal values, the number of iterations needed was between 3 and 12 in most cases. It usually increased with increased number of parameters to be fit. Since parameter shifts are likely to be small with time, the actual number of iterations for a re-fitting operation starting with the previous optimal parameter-set as initial values may in fact be quite small.

## 5 Forecasting

Apart from estimating the parameters of the seasonal ARMA models given in (2), predicting the behavior of the network traffic for future time intervals can be useful for monitoring and allocating buffers and/or bandwidth for incoming traffic. Traditional forecasting techniques predict the mean of the process by minimizing some cost function, usually the mean squared error. However, this is inappropriate in this situation because if the predicted value is less than the observed value, then

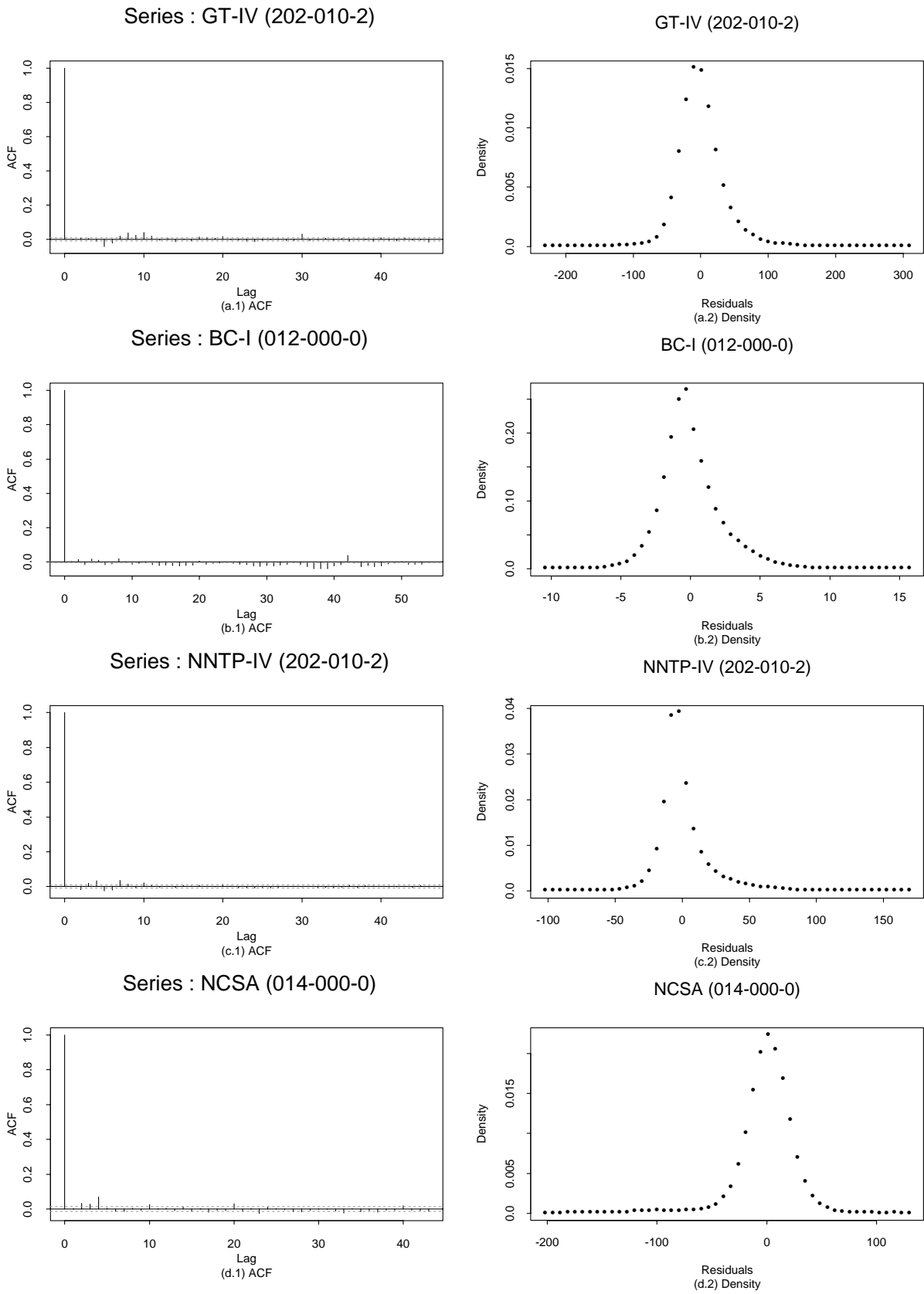


Figure 7: Residual auto-correlation and density functions: (a) GT-IV, (b) BC-I, (c) NNTP-IV and (d) NCSA.

some packets will be dropped. To alleviate this problem we propose, in this section, a prediction algorithm which can be used to predict the  $u$ th quantile of the observed values.

Let  $Y_t^{(u)}(r)$  denote the predictor of the  $u$ th percentile of  $Y_{t+r}$  based on the infinite past  $Y_t, Y_{t-1}, \dots$ . Recall that  $W_t = (1-B)^d(1-B^s)^D Y_t$  and  $W_t$  follows an ARMA model. Let  $W_t(r)$  denote the minimum mean square error forecast function of  $W_{t+r}$ . Expressions for  $W_t(r)$  are available in any standard text book on time series analysis and is given by

$$W_t(r) = \sum_{i=1}^{\infty} \pi_i W_{t+r-i},$$

where  $\pi_i \in \mathcal{R}$ , ( $i = 1, 2, \dots$ ), and are the weights given to  $W_{t+r-i}$ , ( $i = 1, 2, \dots$ ). For their relationship with the ARMA model parameters, see [1, Ch. 5]. Another way of writing the forecast function is in terms of the both the past  $p + Ps$  values of  $W_t$  and the past  $q + Qs$  noise terms as follows.

$$W_t(r) = \sum_{i=1}^p \phi_i W_t(r-i) - \sum_{j=1}^q \theta_j [a_{t+r-j}],$$

where  $[a_{t+r-j}]$  is  $a_{t+r-j}$  if  $r-j \leq 0$ , and is zero otherwise. Since data-sets do not have an infinite past, we use the second equation to obtain the forecasts. To use this, we need the last  $q$  noise terms. These can be obtained recursively by assuming that the initial  $a$ 's are zero and calculating one step ahead forecasts starting from  $Y_{p+1}$  and calculating  $a_j$  as  $Y_j - Y_{j-1}(1)$  for  $j = p+1, \dots, t$ .

For prediction purposes, we will assume that the parameters of the ARMA model in (2) are known. In reality, we will estimate the parameters using the methods described in the previous section and then use these estimated parameters in the forecasting algorithm.

**Lemma 1** *Let  $\xi_u$  be the 100th percentile of the distribution of the noise  $a$ . Then the  $u^{\text{th}}$  percentile of  $Y_{t+1}$  given an infinite past is given by*

$$Y_{t+1}^u = Y_{t+1-s} + W_t(1) + \xi_u. \quad (4)$$

*Proof:* Define  $z = Y_{t+1} - Y_{t+1-s} - W_t(1)$ . Note that the distribution of  $z$  does not depend on the past values of  $Y$ . Hence,

$$\int_{-\infty}^{Y_{t+1}^u} f(Y_{t+1}|Y_t, Y_{t-1}, \dots) dY_{t+1} = \int_{-\infty}^{\xi_u} f(z|Y_t, Y_{t-1}, \dots) dz = \int_{-\infty}^{\xi_u} f(z) dz = u. \blacksquare$$

To forecast the  $100u^{\text{th}}$  percentile of the distribution of  $Y_{t+1}$ , we use the following algorithm:

- Estimate the parameters of the ARMA model by using the nonlinear least squares algorithm described in the previous section.
- Calculate the residuals from the fitted model and obtain the  $100u^{\text{th}}$  percentile of the distribution of the residuals. Call it  $\xi_u$ .
- Obtain the minimum mean square error forecast for  $Y_{t+1}$  based on the ARIMA model. Call it  $Y_t(1)$ .
- The forecast for the  $100u^{\text{th}}$  percentile for  $Y_{t+1}$  is  $Y_t(1) + \xi_u$ .

Data Period	% Under-forecasts				
	$u = 0.5$	$u = 0.8$	$u = 0.9$	$u = 0.95$	$u = 0.99$
1 to 9000	41.7	19.1	11.3	5.3	1.8
9001 to 18000	40.8	13.3	5.1	1.9	0.1
18001 to 27000	43.7	22.8	13.6	8.0	1.6
27001 to 36000	45.0	18.4	9.4	5.0	0.6

Table 4: Percent of times the forecasting procedure under-forecasted using the one step ahead forecast for the SMTP-III trace data.  $u$  is the quantile used to forecast as given in Lemma 1. The results are based on 1000 out-of-sample forecasts.

The choice of  $\xi_u$  depends on how frequently we are willing to have less buffers/bandwidth available than is necessary. This will depend on the application in which it is used. If the cost of losing packets in transmission is high then  $u$  should be very close to one. On the other hand if the cost is not very high then  $u$  can be smaller. One problem of choosing  $u$  close to one is that it will result in under-use of capacity.

As an illustration of the method of forecasting, and how the forecast compares with true values, consider the SMTP-III trace. From the Appendix, the model that fits this data-set the best (with  $T = 0.1\text{sec}$ ) is

$$\nabla Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2},$$

where  $\nabla Y_t = Y_t - Y_{t-1}$ , i.e.,  $s = 1$ . We look at 15 minute segments of the data separately, starting at time 1. With  $T = 0.1\text{sec}$ , there are 9000 data points for each 15 minute segment. For each segment, the model was re-fitted because it was observed that the parameter estimates drifted slowly with time. These updated parameters were then used in forecasting. The results, therefore, are slightly better than what one would obtain without re-fitting. Different quantile values were used for the forecast, and results for three of these, 0.5, 0.9, and 0.99 quantile forecasts, are shown in Table 4. Figure 8 shows the difference between one step ahead forecasts and true values for a segment of the second 15 minutes. To keep it readable, 100 successive one-step-ahead forecasts are shown. The performance of the procedure varies from one time interval to the next. However, the tabulated values are approximately what one would expect from Lemma 1.

From Figure 8, the higher the value of  $u$ , less often the forecasting procedure under-forecasts. It is important to compare this forecasting algorithm with the traditional least-mean-square-error forecast. The latter would predict the mean estimated  $\hat{Y}_{t+1}$ . For differenced datasets, the  $a$ 's are approximately symmetric around zero, so the mean forecast will be close to the  $u = 0.5$  forecast. From Lemma 1, it is expected that for a given value of  $u$ , about  $100u\%$  forecasts will be above the observed series, and  $100(1 - u)\%$  forecasts will be below the observed series. The under-forecasts will, therefore, be significant if a least-mean-square error forecast were used. This situation is mitigated somewhat by assuming a Gaussian distribution for the  $a$ 's, and adding  $G_u \sigma$  to the mean, where  $G_u$  is the  $u^{\text{th}}$  quantile of a standard Normal distribution, and  $\sigma$  is the standard error of the  $a$ 's. While this leads to a  $100u\%$  forecast under Gaussian assumptions, it does not give the true  $100u\%$  forecast if Gaussian assumption for the  $a$ 's is not appropriate. By deriving the quantiles directly from the distribution of  $\{a_t\}$ , the tail-forecast algorithm trades off known model acceptance tests (as discussed in Section 4) for the empirical quantile needed for accurate prediction.

It is instructive to see how this forecast compares with other more intuitive forecasting methods such as (i) using the last value, (ii) using average of the last  $n$  values, and (iii) using an exponential

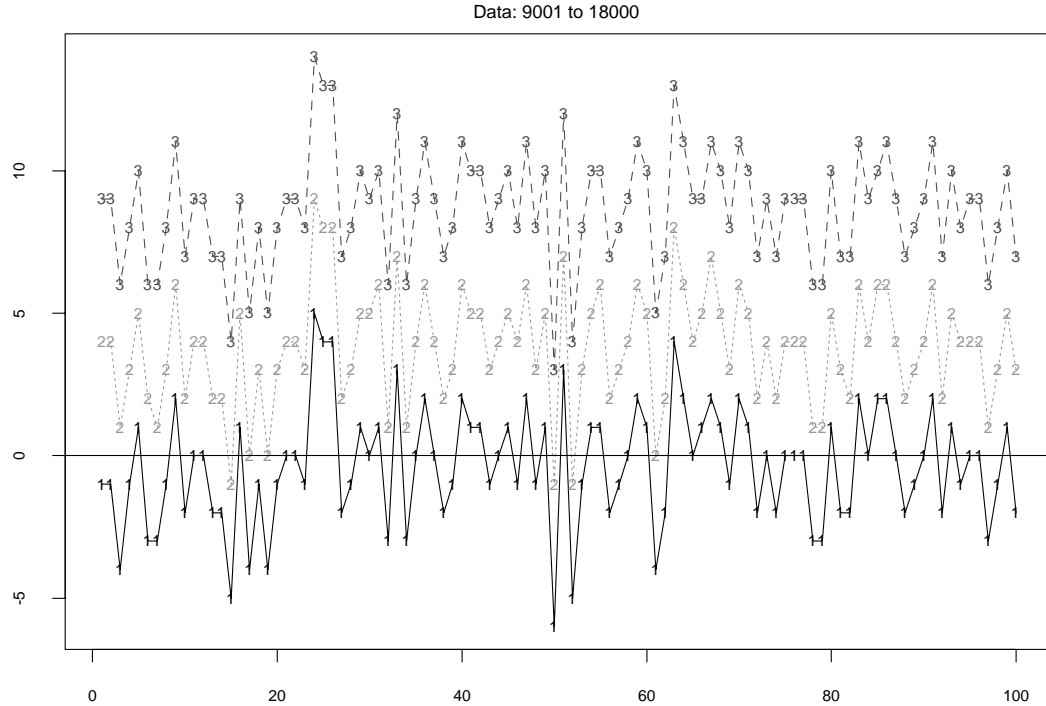


Figure 8: Differences between one step ahead forecasts and observed values for a section of the SMTP-III trace (second fifteen minutes). For clarity, the first 100 differences between the forecast and the corresponding true value are plotted. The horizontal line corresponding to  $y = 0$  represents the case when the forecast equals the true value. Values below the line are under-forecasts (true value exceeds forecasted value for a given  $u$ ). These would cause resource over-commitments, e.g., potential buffer overflows. Values above the line are over-forecasts. These would result in lower resource utilization. The application using the forecasting algorithm would need to decide on its own application-specific costs for these cases. The topmost line (labeled 3) corresponds to  $u = 0.99$ , the one in the middle (labeled 2) corresponds to  $u = 0.9$ , and the lowest line (labeled 1) corresponds to  $u = 0.5$ . A typical least mean-square error prediction is expected to be close to the  $u = 0.5$  line because the mean and median of  $\{a_t\}$  are approximately equal, and close to zero, for most differenced data-set models. In networking applications, where buffer overflows are expensive,  $100u\%$  forecasts are more appropriate. The forecasts in this figure used the  $u^{\text{th}}$  quantile of the residuals of fit, so it traded off known model acceptance tests (see Section 4) for the empirical quantile. Neither method is ‘better’ than the other; it is a trade-off.

smoothing algorithm.

- (i) A one-step-ahead forecast that uses the last value seen as its forecast, i.e.,  $\hat{Y}_{t+1} = Y_t$ , at time  $t$ , has the following underlying model:

$$\begin{aligned} Y_t &= Y_{t-1} + a_t \\ \Rightarrow Y_t - Y_{t-1} &= a_t \\ \Rightarrow W_t &= a_t \end{aligned}$$

where  $\{a_t\}$  is white noise. Since the data explicitly indicates the presence of at least a few significant auto-correlations for  $W_t$  for all traffic data-sets (e.g., see Figure 1(b)), this model will be less than adequate. Also, in some cases,  $s$  may be greater than one in  $Y_t - Y_{t-s}$ . A more accurate model would therefore exploit the correlation structure of the time-series.

- (ii) A one-step-ahead forecast that uses the average of last  $n$  values has the following implicit model-form:

$$Y_t = \frac{1}{n}(Y_{t-1} + \dots + Y_{t-n}) + a_t,$$

or equivalently,

$$Y_t - \phi_1 Y_{t-1} - \dots - \phi_n Y_{t-n} = a_t$$

with  $\phi_i = 1/n$  for all  $i$  — which is a restricted form of an AR( $n$ ) model. While the averaging operation is an attractive idea because of its simple and fast implementation, traffic data-sets studied do not support such a model. Forecast results are, therefore, unlikely to be accurate.

- (iii) A forecast that uses exponential smoothing assumes an ARIMA(0, 1, 1) model for the underlying correlation structure:

$$Y_t - Y_{t-1} = a_t - \theta_1 a_{t-1}$$

From the Appendix, SD-NSF admits this model. The SMTP datasets admit an ARIMA(0, 1, 2) model, with  $\theta_2$  small.

Therefore, empirical evidence supports the exponential smoothing algorithm for some of the traffic datasets.

The recommended approach for accurate forecasting would be to develop the appropriate model for a network site based on the data observed. This will explicitly take into account not only the ARMA( $p, q$ ) form for  $W_t$ , but also the non-random periodicity,  $s$ . Given that models for different sites are different, it is inappropriate (and inaccurate) to use a single model for all cases.

As seen from the Appendix, the model forms for a given site (including the differencing-lag  $s$ ) appear to remain the same across time. Also, the parameter estimates are close to each other. While this is not likely if measurements are spaced far apart in time (because new applications and new technology may impact traffic correlation structures), parameter shifts may be expected to move slowly with time in the time-scale at which a forecasting algorithm needs to re-fit its model. The pros and cons of on-line re-fitting were discussed in Section 4.

## 6 Generating Synthetic Sequences

It is desirable to be able to generate synthetic data sequences that exhibit features of measured traffic. For instance, with the help of a parametric model, a congestion control algorithm may be

studied for parametric sensitivities using generated sequences (e.g., to investigate questions such as what happens if  $\phi$  or  $\theta$  were slightly different from that observed in a given trace).

In this section, we study an algorithm for generating synthetic sequences based on the multiplicative ARIMA models developed in Section 4. The experimentation leads to identification of a numerical integration problem for which, we have only a partial solution.

Suppose that  $\{a_t\}$  are the uncorrelated residuals of a model-fit. These have a known empirical distribution. Then, at first sight, it appears that one should be able to generate first  $\{W_t\}$ , and subsequently, from it  $\{Y_t\}$  :

$$W_t = \phi^{-1}(B)\Phi^{-1}(B^s)\theta(B)\Theta(B^s)a_t,$$

and

$$(1 - B)^d(1 - B^s)^D Y_t = W_t,$$

where the  $a_t$ 's are now sampled from the empirical distribution (so they do not necessarily appear in the same order as in the original data). However, experiments show that this does not work very well. There are two issues: (i) the boundary condition  $Y_t \geq 0$ , and (ii) the problem of noise that gets induced in numerical integration (summation) in computing the  $Y_t$ 's from  $W_t$ 's.

The boundary condition  $Y_t \geq 0$  may be enforced by setting  $Y_t$  to zero if it becomes negative — we shall study its effect shortly. The problem of noise is perhaps best seen with an example. Figure 9 shows the SMTP-III dataset and its ACF in (a) and (b), and a generated sequence and its ACF in (c) and (d). In this figure, we have (i) enforced the boundary condition, and (ii) rounded the  $Y_t$ 's to their closest integer counterpart — in order to reflect an integer dataset. The patterns for the generated sequences are similar with or without these two operations — they do not look like the data, and have ACFs that decay much slower than that observed in the dataset.

One algorithm is to truncate  $Y_t$  rather than rounding it. Figure 10 shows the resulting plots. We observe that the generated sequence 'looks' closer to the measured sequence; however, its ACF decays faster. The reason for both of these observations can be traced to noise suppression in the truncation procedure as discussed below.

## Noise Suppression

For ease of exposition, we assume the ARIMA(0, 1, 1) model in the following discussion. We shall see that the argument is more generally applicable. For the ARIMA(0, 1, 1) model, we have

$$(1 - B)Y_t = a_t - \theta a_{t-1},$$

or,

$$Y_t = Y_{t-1} + a_t - \theta a_{t-1}.$$

In the algorithm involving either truncation or rounding, with or without the boundary condition, we can write

$$Z_t = Y_{t-1} + a_t - \theta a_{t-1}, \tag{5}$$

$$Y_t = b_t Z_t, \tag{6}$$

where, if we use truncation, we have  $0 \leq b_t \leq 1$  for all  $t$ , and if we use rounding,  $b_t$  may be greater than one at times. Further,  $b_t$  equals zero if  $Z_t < 0$ .

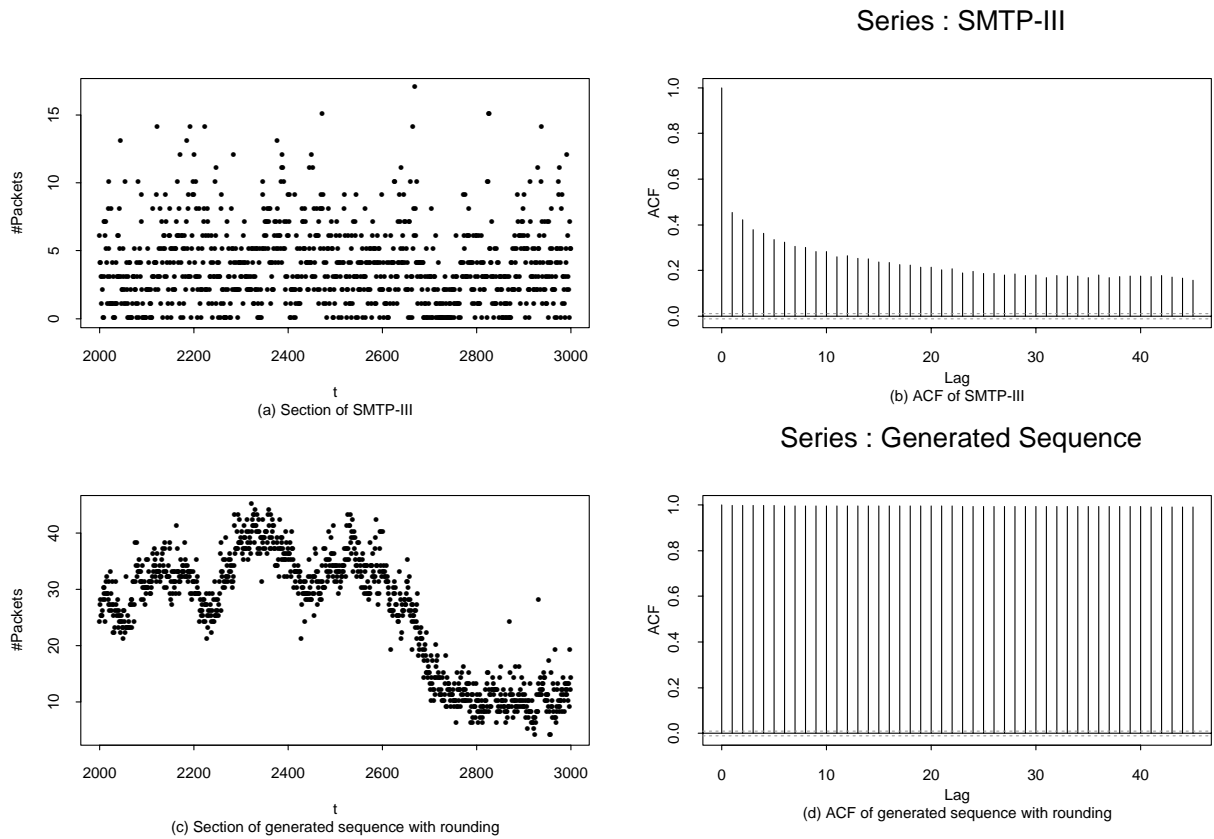


Figure 9: The SMTP-III dataset and a faulty generated sequence. (a) Section of traffic data, SMTP-III. (b) ACF of SMTP-III. (c) Section of generated sequence using rounding. (d) ACF of generated sequence.

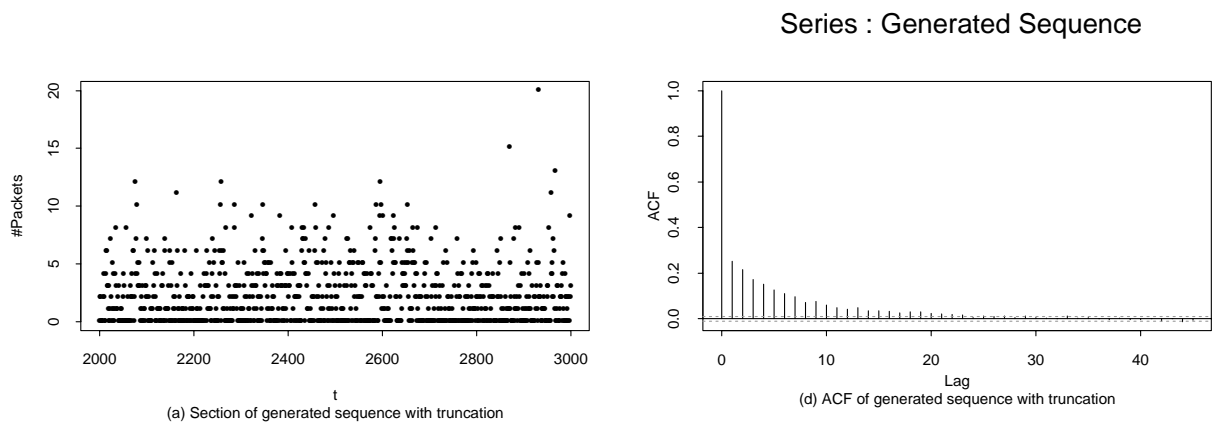


Figure 10: (a) Section of generated sequence using truncation. (b) ACF of generated sequence.

From (5) and (6),

$$\begin{aligned}
Y_t &= b_t(Y_{t-1} + a_t - \theta a_{t-1}), \\
&= b_t b_{t-1}(Y_{t-2} + a_{t-1} - \theta a_{t-2}) + b_t a_t - \theta b_t a_{t-1}, \\
&= b_t a_t + b_t b_{t-1} a_{t-1} - \theta(b_t a_{t-1} + b_t b_{t-1} a_{t-2}) + b_t b_{t-1} Y_{t-2},
\end{aligned}$$

which recursively expands to

$$Y_t = \sum_{i=0}^t a_{t-i} \prod_{j=0}^i b_{t-j} - \theta \sum_{i=0}^t a_{t-1-i} \prod_{j=0}^i b_{t-j}. \quad (7)$$

Now, we do not have the true value of  $\theta$  but its estimate  $\hat{\theta}$ , where  $\hat{\theta} = \theta + \epsilon$ ,  $\epsilon$  being an error term. Similarly, we potentially introduce noise in the  $a_t$ 's themselves (while generating them using the algorithm discussed in Section 4), so  $a_t$  in (7) is in fact  $a_t + \delta_t$ , where  $\delta_t$  represents the error in  $a_t$ . Therefore, (7) is in fact

$$\begin{aligned}
Y_t &= \sum_{i=0}^t (a_{t-i} + \delta_{t-i}) \prod_{j=0}^i b_{t-j} - (\theta + \epsilon) \sum_{i=0}^t (a_{t-1-i} + \delta_{t-1-i}) \prod_{j=0}^i b_{t-j} \\
&= \sum_{i=0}^t a_{t-i} \prod_{j=0}^i b_{t-j} - \theta \sum_{i=0}^t a_{t-1-i} \prod_{j=0}^i b_{t-j} + \text{error terms}.
\end{aligned} \quad (8)$$

Now, with truncation,  $0 \leq b_t \leq 1$  for all  $t$ , and with high probability,  $b_t < 1$ , so

- (i) the noise introduced further away from the present time  $t$ , get damped out quickly because they are weighted by  $\prod_j b_{t-j}$ ;
- (ii) the signal corresponding to the the  $a_{t-i}$ 's also get damped.

Once  $b_{t-j} = 0$ , the past values of  $a_{t-k-1}$  for  $k \leq j$  also do not appear in  $Y_t$  and neither do the noise terms. The boundary condition, therefore, damps both the signal and the noise further. With rounding, if over an interval  $I$ ,  $\{Y_t, t \in I\}$  is positive, the  $b_t$ 's are greater than one with probability 0.5, so noise suppression through  $\prod_j b_{t-j}$  is not necessarily effective. This is most likely the reason why the generated sequence does not resemble the measured data in Figure 9(c)-(d). With truncation, the damping of noise makes the generated series look closer to the measured data. However, it also dampens the dependence in the data across far lags [see Eq. (8)].

It is interesting to note that the noise suppression problem is similar to that in numerical integration of partial differential equations [12]. Finite-difference schemes for different classes of partial differential equations have been developed with the explicit objective of damping noise in their solution paths. Qualitative differences between the problem at hand and pde's are that (i) unlike a pde, there is no spatial dimension in our problem, only a time dimension, and (ii) there is a stochastic component in our problem that is not usually present in pde solutions.

One way to improve upon the ACF of the generated sequence is to generate  $Y_t$  as follows. Suppose the model written in its general form, including any difference that is needed, follows an ARMA( $p^*$ ,  $q^*$ ) model where  $p^* = p + sP + d + sD$  and  $q^* = q + sQ$ . For ease of notation, let us write the parameters for the AR and the MA parts as  $\phi_1, \dots, \phi_{p^*}$  and  $\theta_1, \dots, \theta_{q^*}$ , respectively. Then we can write the model as

$$Y_t = \phi_1 Y_{t-1} + \dots + \phi_{p^*} Y_{t-p^*} + a_t - \theta_1 a_{t-1} - \dots - \theta_{q^*} a_{t+q^*}.$$

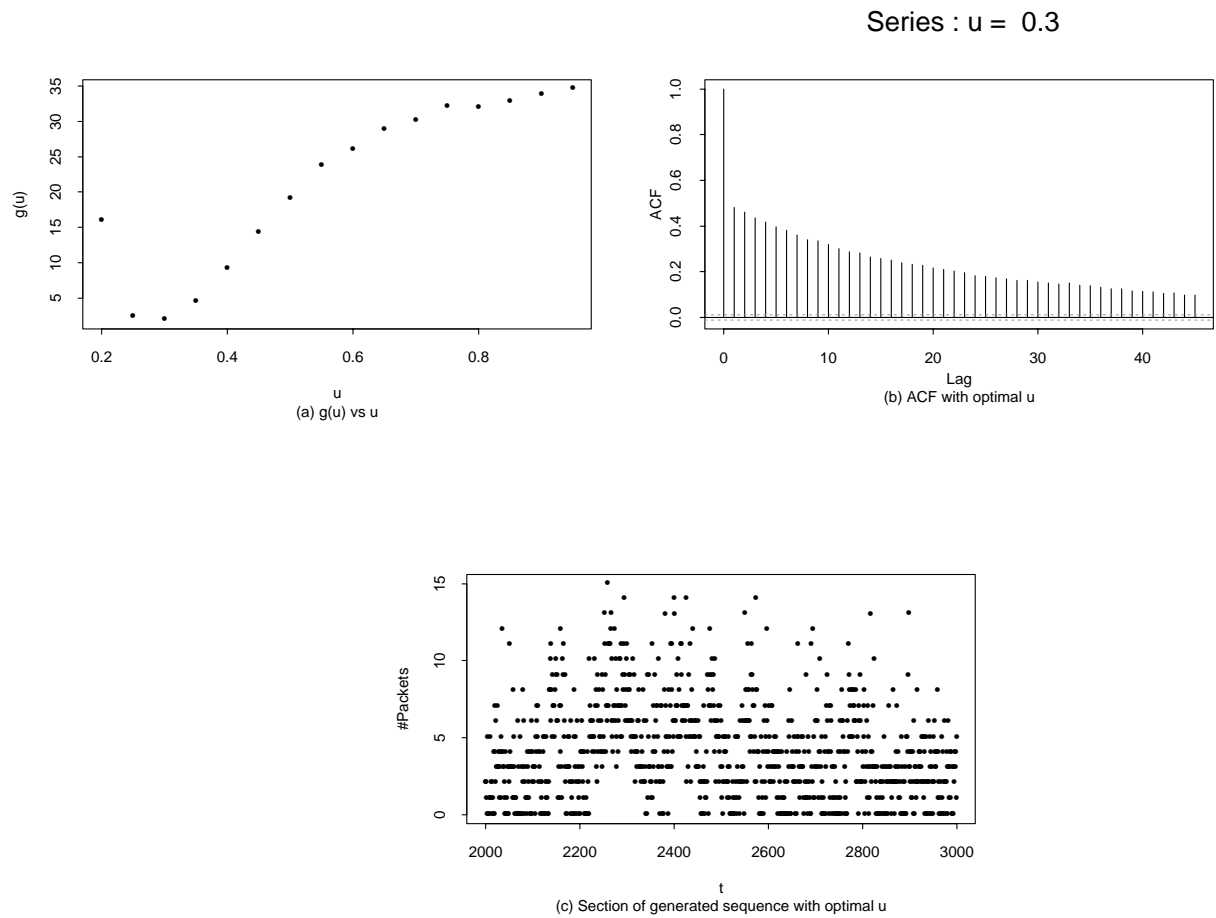


Figure 11: (a) Cost function,  $g(u)$  vs  $u$  for  $0.2 \leq u \leq 1.0$ . (Values of  $u$  smaller than 0.2 yield much larger  $g(u)$  and are, therefore, omitted.) (b) ACF of generated sequence with optimal  $u$ . (c) Section of corresponding generated sequence.

Set  $a_j = 0$  and  $Y_j = 0$  for  $j \leq 0$ . Let

$$Z_t = \phi_1 Y_{t-1} + \dots + \phi_{p^*} Y_{t-p^*} + a_t - \theta_1 a_{t-1} - \dots - \theta_{q^*} a_{t-q^*}. \quad (9)$$

If  $Z_t \leq 0$ , then define  $Y_t = 0$ . However, if  $Z_t > 0$ , then generate  $Y_t$  as follows:

$$Y_t = \begin{cases} [Z_t], & \text{with probability } u \\ (Z_t), & \text{with probability } 1 - u \end{cases} \quad (10)$$

where  $[Z_t]$  is the integer part of  $Z_t$  and  $(Z_t)$  is the integer obtained by rounding  $Z_t$ . Next, select  $u$  that minimizes

$$g(u) = \sum_k \left( r_{data}(k) - r_{gen}(u, k) \right)^2 w_k, \quad (11)$$

where  $r_{data}(k)$  and  $r_{gen}(u, k)$  are the estimated auto-correlations at lag  $k$  for the measured data and the generated data for a given  $u$ , respectively. Experiments show that when the probability of truncation,  $u$ , is close to one,  $(r_{data}(k) - r_{gen}(u, k))$  increases with  $k$  — because of the damping effect of truncation. However, with  $u$  close to zero,  $(r_{data}(k) - r_{gen}(u, k))$  is negative for all  $k$ , because in this case, more values are rounded up. Therefore, we expect  $u$  to be between 0 and 1. Experiments further suggest that  $w_k$  should be an increasing function of  $k$  because for the case where  $u$  serves to damp the noise, it also damps the ACF, and for this case,  $(r_{data}(k) - r_{gen}(u, k))^2$  increases with  $k$ . Figure 11 shows  $g(u)$  versus  $u$  for  $w_k = k$ , and the corresponding ACF and a section of the generated sequence for the value of  $u$  that minimizes  $g(u)$  for this specific dataset. The generated data is close in ACF to that of the measured sequence. However, it still is a damped series, and does not preserve the ACF at larger time-scales — the damping appears to make it a short memory process. More study is required in this regard.

## 7 Concluding Remarks

Internet traffic from a number of networks and for a range of popular TCP port numbers (with the notable exception of FTP-data), lend themselves to seasonal ARIMA models. In many cases, they show non-random periodicities — which may be time-scale sensitive.

ARIMA models have typically been applied with Gaussian variates. However, this is not a necessity. If one is willing to adopt a non-linear least squares algorithm for parameter estimation, one can compute the variates of the distribution and obtain their tail percentiles directly from the computation — using compute power of current machines where maximum likelihood based methods become intractable. There are of course limitations of such an approach: no longer are the standard acceptance/rejection tests based on known distribution forms available. The analyst does have plots of the ACF of residuals to point to potential model inadequacies.

The distribution of the variates can be used for forecasting tail percentiles of traffic data. Such a forecast is useful if the cost of overshooting the forecast is significantly higher than undershooting it — as is the case with dropped packets over the Internet. The algorithm may have potential application in dynamic bandwidth and buffer allocation strategies for connectionless servers — especially if they are required to reserve bandwidth in advance in future broadband networks. While it is not clear that this approach will be adopted by the networking community, the key is that seasonal ARIMA models can be fit to Internet traffic in many cases, and in such cases, a corresponding forecast may be obtained.

Synthetic traffic generation is another application of interest — especially in simulation studies of resource management algorithms, for example, in background traffic modeling, and sensitivity

study of parameter estimates on a given network configuration or a new algorithm being tested. Preliminary results on generating such sequences indicate that noise suppression in the generation process is an important consideration. Partial results based on truncation and rounding with probabilities  $u$  and  $(1 - u)$ , respectively, indicate that the noise is indeed sufficiently damped, but so is the signal. More work on this is necessary, and is currently in progress.

## Acknowledgments

We wish to thank Dr. Hans-Werner Braun for his support of this work and for providing us with the SDSC and ENSS datasets, Dr. Will Leland for providing us with the Bellcore datasets, and Prof. John Limb for his support and encouragement for our traffic analysis work.

## References

- [1] Box, G.E.P, G.M. Jenkins, and G.C. Reinsel, "Time Series Analysis: Forecasting and Control," Third Edition, Prentice Hall, 1994.
- [2] Caceres, R., P.B.Danzig, S.Jamin and D.J.Mitzel, "Characteristics of wide-area TCP/IP conversations," *Proc of ACM Sigcomm*, 101-112, Zurich, Sept 1991.
- [3] Claffy, K. C., H. W. Braun, and G. C. Polyzos, "A parameterizable methodology for Internet traffic profiling," *IEEE Journal on Selected Areas in Communications*, (to appear).
- [4] Claffy, K.C., G.C. Polyzos and H.W. Braun, "Traffic characteristics of the T1 NSFNet backbone," *Proc. IEEE Infocom*, San Francisco, 1993.
- [5] Danzig, P. B., S. Jamin, R. Caceres, D.J. Mitzel, and D. Estrin, "An Empirical Workload Model for Driving Wide-Area TCP/IP Network Simulations", *Journal of Internetworking: Practice and Experience*, 3, No. 1, March 1992.
- [6] Klivansky S., A. Mukherjee and C. Song, "On long-range dependence in NSFNET traffic," College of Computing Tech Report GIT-CC-94-61, Georgia Institute of Technology, December 1994.
- [7] Leland, W.E. and D. Wilson, "High time resolution measurement and analysis of LAN traffic: implications for LAN interconnection," *Proc. IEEE Infocom*, 1991.
- [8] Leland, W.E., M.S. Taqqu, W. Willinger and D.V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE Trans. Networking*, **2**, (1), 1-15, February, 1994.
- [9] Paxson, V. and S. Floyd, "Wide-area traffic: the failure of Poisson modeling," *IEEE/ACM Trans. on Networking*, **3**, (3), 226-244, June 1996.
- [10] Paxson, V., "Empirically-derived analytical models of wide-area TCP connections: extended report," *IEEE/ACM Transactions on Networking*, **2**, (4), 316-336, August 1994.
- [11] Priestley, M.B., "Spectral analysis and time series," Academic Press, 1988.
- [12] Strikwerda, J. C., "Finite difference schemes and partial differential equations," *Wadsworth and Brooks/Cole*, 1989.

- [13] W. Willinger, M.S. Taqqu, W.E. Leland, and D.V. Wilson, “Self-similarity in high-speed packet traffic: analysis and modeling of ethernet traffic measurements”, *Statistical Science*, **10**, (1), 67-85, 1995.

## A Summary of Seasonal Traffic Models

In the following,  $Y_t$  denotes the number of packets in time-interval  $(t-1, t]$ .  $W_t$  denotes the differenced series  $Y_t - Y_{t-s}$ , for which an ARMA model is studied. The value of  $s$  for each data-set is indicated in the first equation for each model. An ARMA( $p, q$ ) model has the form

$$W_t - \phi_1 W_{t-1} - \dots - \phi_p W_{t-p} = a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}.$$

The values of  $p$  and  $q$  are indicated through the second equation in each model's summary. Parameter estimates are shown next, followed by the estimated error variance. Each model is fit for a fixed interval of length  $T$  which is shown as the last entry. We studied models with  $T = 0.1s$  or  $T = 0.01s$ . For most cases, adequate models were found to fit both values of  $T$ , although the model forms differed. Only one  $T$  per dataset is reported below. The value of  $T$  to use in a resource-management study will depend on the application.

For each traffic trace, a set of viable models were identified, based on ACF of datasets, and ACF of differenced datasets for different values of  $s$ . The models given below were found to fit the traffic data best — i.e., the ACF of  $\{a_t\}$  did not show appreciable auto-correlations, and for a fixed  $T$ , its variance was the smallest among the alternatives. In some cases, differences between two or more models were not appreciable. In such cases, two of the models are given.

Seasonal ARIMA models appear to fit a fairly large class of Internet traffic. However, their exact forms and parameter estimates differ. For new traffic traces where seasonal traffic models may apply, one would need to (i) investigate the model form, (ii) obtain parameter estimates, and (iii) generate the error distribution. If the model form at a multiplexor is known (e.g., does not change appreciably with time), steps (ii) and (iii) may be computed on-line with previous data seen at the multiplexor.

### A.1 Campus FDDI Traffic

- GT-I:

(a)

$$\text{Model form: } \begin{cases} W_t = Y_t - Y_{t-2} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \theta_4 a_{t-4} - \theta_5 a_{t-5} - \theta_6 a_{t-6} \end{cases}$$

Parameter estimates:  $\hat{\theta}_1 = -0.3688, \hat{\theta}_2 = 0.6814, \hat{\theta}_3 = 0.1416, \hat{\theta}_4 = 0.1025, \hat{\theta}_5 = 0.1080, \hat{\theta}_6 = 0.0855.$

Diagnostics:  $\hat{\sigma}^2 = 674.3073. \underline{T} = 0.1s.$

(b)

$$\text{Model form: } \begin{cases} W_t = Y_t - Y_{t-2} \\ W_t - \phi_1 W_{t-1} - \phi_2 W_{t-2} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \end{cases}$$

Parameter estimates:  $\hat{\phi}_1 = 0.3250, \hat{\phi}_2 = 0.1562, \hat{\theta}_1 = -0.0495, \hat{\theta}_2 = 0.9367.$

Diagnostics:  $\hat{\sigma}^2 = 674.3928. \underline{T} = 0.1s.$

- GT-II:

(a)

$$\text{Model form : } \begin{cases} W_t = Y_t - Y_{t-2} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \theta_4 a_{t-4} - \theta_5 a_{t-5} - \theta_6 a_{t-6} \end{cases}$$

Parameter estimates :  $\hat{\theta}_1 = -0.3201, \hat{\theta}_2 = 0.8138, \hat{\theta}_3 = 0.2218, \hat{\theta}_4 = 0.1052, \hat{\theta}_5 = 0.0529, \hat{\theta}_6 = 0.0293$ .  
Diagnostics :  $\hat{\sigma}^2 = 829.5069$ .  $\underline{T} = 0.1s$ .

(b)

$$\text{Model form : } \begin{cases} W_t = Y_t - Y_{t-2} \\ W_t - \phi_1 W_{t-1} - \phi_2 W_{t-2} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \end{cases}$$

Parameter estimates :  $\hat{\phi}_1 = 0.3031, \hat{\phi}_2 = 0.0804, \hat{\theta}_1 = -0.0207, \hat{\theta}_2 = 0.9715$ .  
Diagnostics :  $\hat{\sigma}^2 = 827.3565$ .  $\underline{T} = 0.1s$ .

- GT-III:

(a)

$$\text{Model form : } \begin{cases} W_t = Y_t - Y_{t-2} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \theta_4 a_{t-4} - \theta_5 a_{t-5} - \theta_6 a_{t-6} \end{cases}$$

Parameter estimates :  $\hat{\theta}_1 = -0.3679, \hat{\theta}_2 = 0.7974, \hat{\theta}_3 = 0.2437, \hat{\theta}_4 = 0.0836, \hat{\theta}_5 = 0.0690, \hat{\theta}_6 = 0.0542$ .  
Diagnostics :  $\hat{\sigma}^2 = 874.0525$ .  $\underline{T} = 0.1s$ .

(b)

$$\text{Model form : } \begin{cases} W_t = Y_t - Y_{t-2} \\ W_t - \phi_1 W_{t-1} - \phi_2 W_{t-2} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \end{cases}$$

Parameter estimates :  $\hat{\phi}_1 = 0.3425, \hat{\phi}_2 = 0.0505, \hat{\theta}_1 = -0.0284, \hat{\theta}_2 = 0.9587$ .  
Diagnostics :  $\hat{\sigma}^2 = 874.2110$ .  $\underline{T} = 0.1s$ .

- GT-IV:

(a)

$$\text{Model form : } \begin{cases} W_t = Y_t - Y_{t-2} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \theta_4 a_{t-4} - \theta_5 a_{t-5} - \theta_6 a_{t-6} \end{cases}$$

Parameter estimates :  $\hat{\theta}_1 = -0.4221, \hat{\theta}_2 = 0.7143, \hat{\theta}_3 = 0.2497, \hat{\theta}_4 = 0.1732, \hat{\theta}_5 = 0.1359, \hat{\theta}_6 = 0.0673$ .  
Diagnostics :  $\hat{\sigma}^2 = 961.8108$ .  $\underline{T} = 0.1s$ .

(b)

$$\underline{\text{Model form}} : \begin{cases} W_t = Y_t - Y_{t-2} \\ W_t - \phi_1 W_{t-1} - \phi_2 W_{t-2} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \end{cases}$$

Parameter estimates :  $\hat{\phi}_1 = 0.4091$ ,  $\hat{\phi}_2 = 0.0792$ ,  $\hat{\theta}_1 = -0.0136$ ,  $\hat{\theta}_2 = 0.9758$ .

Diagnostics :  $\hat{\sigma}^2 = 963.7665$ .  $\underline{T} = 0.1s$ .

- **SDSC-FDDI:**

(a)

$$\underline{\text{Model form}} : \begin{cases} W_t = Y_t - Y_{t-1} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} \end{cases}$$

Parameter estimates :  $\hat{\theta}_1 = 0.1884$ ,  $\hat{\theta}_2 = 0.1230$ ,  $\hat{\theta}_3 = 0.1038$ . Diagnostics :  $\hat{\sigma}^2 = 3.3687$ .  $\underline{T} = 0.01s$ .

(b)

$$\underline{\text{Model form}} : \begin{cases} W_t = Y_t - Y_{t-1} \\ W_t - \phi_1 W_{t-1} = a_t - \theta_1 a_{t-1} \end{cases}$$

Parameter estimates :  $\hat{\phi}_1 = 0.7912$ ,  $\hat{\theta}_1 = 0.9978$ . Diagnostics :  $\hat{\sigma}^2 = 3.1893$ .  $\underline{T} = 0.01s$ .

## A.2 NSFNET External Interface Traffic

- **SD-NSF:**

$$\underline{\text{Model form}} : \begin{cases} W_t = Y_t - Y_{t-1} \\ W_t - \phi_1 W_{t-1} = a_t - \theta_1 a_{t-1} \end{cases}$$

Parameter estimates :  $\hat{\phi}_1 = 0.3958$ ,  $\hat{\theta}_1 = 0.9657$ . Diagnostics :  $\hat{\sigma}^2 = 14.4814$ .  $\underline{T} = 0.01s$ .

- **NCSA:**

$$\underline{\text{Model form}} : \begin{cases} W_t = Y_t - Y_{t-1} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \theta_4 a_{t-4} - \theta_5 a_{t-5} - \theta_6 a_{t-6} \end{cases}$$

Parameter estimates :  $\hat{\theta}_1 = 0.5769$ ,  $\hat{\theta}_2 = 0.0802$ ,  $\hat{\theta}_3 = 0.0652$ ,  $\hat{\theta}_4 = 0.2264$ .

Diagnostics :  $\hat{\sigma}^2 = 5.6006$ .  $\underline{T} = 0.01s$ .

## A.3 Ethernet Traffic

- **BC-I:**

$$\underline{\text{Model form}} : \begin{cases} W_t = Y_t - Y_{t-1} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \end{cases}$$

Parameter estimates :  $\hat{\theta}_1 = 0.5453$ ,  $\hat{\theta}_2 = 0.2568$ . Diagnostics :  $\hat{\sigma}^2 = 4.2111$ .  $\underline{T} = 0.01s$ .

- BC-II:

$$\text{Model form : } \begin{cases} W_t = Y_t - Y_{t-1} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \theta_4 a_{t-4} \end{cases}$$

Parameter estimates :  $\hat{\theta}_1 = 0.7276, \hat{\theta}_2 = 0.2945, \hat{\theta}_3 = -0.1393, \hat{\theta}_4 = 0.0036.$

Diagnostics :  $\hat{\sigma}^2 = 12.1431. \underline{T} = 0.01s.$

#### A.4 Selected TCP Protocol Traffic

- NNTP-I:

(a)

$$\text{Model form : } \begin{cases} W_t = Y_t - Y_{t-2} \\ W_t - \phi_1 W_{t-1} - \phi_2 W_{t-2} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \end{cases}$$

Parameter estimates :  $\hat{\phi}_1 = 0.5399, \hat{\phi}_2 = 0.0642, \hat{\theta}_1 = -0.0179, \hat{\theta}_2 = 0.9640.$

Diagnostics :  $\hat{\sigma}^2 = 237.9764. \underline{T} = 0.1s.$

(b)

$$\text{Model form : } \begin{cases} W_t = Y_t - Y_{t-2} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \theta_4 a_{t-4} - \theta_5 a_{t-5} \end{cases}$$

Parameter estimates :  $\hat{\theta}_1 = -0.5519, \hat{\theta}_2 = 0.6125, \hat{\theta}_3 = 0.3424, \hat{\theta}_4 = 0.2908, \hat{\theta}_5 = 0.1255.$

Diagnostics :  $\hat{\sigma}^2 = 240.3881. \underline{T} = 0.1s.$

- NNTP-II:

(a)

$$\text{Model form : } \begin{cases} W_t = Y_t - Y_{t-2} \\ W_t - \phi_1 W_{t-1} - \phi_2 W_{t-2} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \end{cases}$$

Parameter estimates :  $\hat{\phi}_1 = 0.5018, \hat{\phi}_2 = -0.0545, \hat{\theta}_1 = -0.0226, \hat{\theta}_2 = 0.9668.$

Diagnostics :  $\hat{\sigma}^2 = 256.5227. \underline{T} = 0.1s.$

(b)

$$\text{Model form : } \begin{cases} W_t = Y_t - Y_{t-2} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \theta_4 a_{t-4} - \theta_5 a_{t-5} \end{cases}$$

Parameter estimates :  $\hat{\theta}_1 = -0.5326, \hat{\theta}_2 = 0.7424, \hat{\theta}_3 = 0.4213, \hat{\theta}_4 = 0.2004, \hat{\theta}_5 = 0.0610.$

Diagnostics :  $\hat{\sigma}^2 = 256.0615. \underline{T} = 0.1s.$

- NNTP-III:

(a)

$$\underline{\text{Model form}} : \begin{cases} W_t = Y_t - Y_{t-2} \\ W_t - \phi_1 W_{t-1} - \phi_2 W_{t-2} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \end{cases}$$

Parameter estimates :  $\hat{\phi}_1 = 0.6640, \hat{\phi}_2 = -0.1398, \hat{\theta}_1 = 0.0012, \hat{\theta}_2 = 0.9890.$

Diagnostics :  $\hat{\sigma}^2 = 374.2799. \underline{T} = 0.1s.$

(b)

$$\underline{\text{Model form}} : \begin{cases} W_t = Y_t - Y_{t-2} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \theta_4 a_{t-4} - \theta_5 a_{t-5} \end{cases}$$

Parameter estimates :  $\hat{\theta}_1 = -0.6704, \hat{\theta}_2 = 0.7224, \hat{\theta}_3 = 0.6210, \hat{\theta}_4 = 0.2622, \hat{\theta}_5 = 0.0408.$

Diagnostics :  $\hat{\sigma}^2 = 374.4830. \underline{T} = 0.1s.$

- NNTP-IV:

(a)

$$\underline{\text{Model form}} : \begin{cases} W_t = Y_t - Y_{t-2} \\ W_t - \phi_1 W_{t-1} - \phi_2 W_{t-2} = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \end{cases}$$

Parameter estimates :  $\hat{\phi}_1 = 0.5842, \hat{\phi}_2 = -0.0134, \hat{\theta}_1 = -0.0006, \hat{\theta}_2 = 0.9878.$

Diagnostics :  $\hat{\sigma}^2 = 292.8905. \underline{T} = 0.1s.$

(b)

$$\underline{\text{Model form}} : \begin{cases} W_t = Y_t - Y_{t-2} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \theta_4 a_{t-4} - \theta_5 a_{t-5} \end{cases}$$

Parameter estimates :  $\hat{\theta}_1 = -0.5784, \hat{\theta}_2 = 0.7065, \hat{\theta}_3 = 0.4460, \hat{\theta}_4 = 0.2661, \hat{\theta}_5 = 0.1128.$

Diagnostics :  $\hat{\sigma}^2 = 296.2909. \underline{T} = 0.1s.$

- SMTP-I:

$$\underline{\text{Model form}} : \begin{cases} W_t = Y_t - Y_{t-1} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \end{cases}$$

Parameter estimates :  $\hat{\theta}_1 = 0.7536, \hat{\theta}_2 = 0.0821. \underline{\text{Diagnostics}} : \hat{\sigma}^2 = 10.1946. \underline{T} = 0.1s.$

- SMTP-II:

$$\underline{\text{Model form}} : \begin{cases} W_t = Y_t - Y_{t-1} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \end{cases}$$

Parameter estimates :  $\hat{\theta}_1 = 0.7482, \hat{\theta}_2 = 0.1025. \underline{\text{Diagnostics}} : \hat{\sigma}^2 = 9.0317. \underline{T} = 0.1s.$

- SMTP-III:

$$\text{Model form : } \begin{cases} W_t = Y_t - Y_{t-1} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \end{cases}$$

Parameter estimates :  $\hat{\theta}_1 = 0.7547$ ,  $\hat{\theta}_2 = 0.0744$ . Diagnostics :  $\hat{\sigma}^2 = 7.6814$ .  $\underline{T} = 0.1s$ .

- SMTP-IV:

$$\text{Model form : } \begin{cases} W_t = Y_t - Y_{t-1} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \end{cases}$$

Parameter estimates :  $\hat{\theta}_1 = 0.7231$ ,  $\hat{\theta}_2 = 0.0823$ . Diagnostics :  $\hat{\sigma}^2 = 6.2600$ .  $\underline{T} = 0.1s$ .

- HUMAN-I:

$$\text{Model form : } \begin{cases} W_t = Y_t - Y_{t-2} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \theta_4 a_{t-4} - \theta_5 a_{t-5} \end{cases}$$

Parameter estimates :  $\hat{\theta}_1 = -0.2297$ ,  $\hat{\theta}_2 = 0.8017$ ,  $\hat{\theta}_3 = 0.1182$ ,  $\hat{\theta}_4 = 0.1388$ ,  $\hat{\theta}_5 = 0.0607$ .  
Diagnostics :  $\hat{\sigma}^2 = 14.5273$ .  $\underline{T} = 0.1s$ .

- HUMAN-II:

$$\text{Model form : } \begin{cases} W_t = Y_t - Y_{t-2} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \theta_4 a_{t-4} - \theta_5 a_{t-5} \end{cases}$$

Parameter estimates :  $\hat{\theta}_1 = -0.2077$ ,  $\hat{\theta}_2 = 0.7810$ ,  $\hat{\theta}_3 = 0.0781$ ,  $\hat{\theta}_4 = 0.1264$ ,  $\hat{\theta}_5 = 0.0521$ .  
Diagnostics :  $\hat{\sigma}^2 = 10.9298$ .  $\underline{T} = 0.1s$ .

- HUMAN-III:

$$\text{Model form : } \begin{cases} W_t = Y_t - Y_{t-2} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \theta_4 a_{t-4} - \theta_5 a_{t-5} \end{cases}$$

Parameter estimates :  $\hat{\theta}_1 = -0.2117$ ,  $\hat{\theta}_2 = 0.7441$ ,  $\hat{\theta}_3 = 0.0658$ ,  $\hat{\theta}_4 = 0.1593$ ,  $\hat{\theta}_5 = 0.0650$ .  
Diagnostics :  $\hat{\sigma}^2 = 9.5367$ .  $\underline{T} = 0.1s$ .

- HUMAN-IV:

$$\text{Model form : } \begin{cases} W_t = Y_t - Y_{t-2} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3} - \theta_4 a_{t-4} - \theta_5 a_{t-5} \end{cases}$$

Parameter estimates :  $\hat{\theta}_1 = -0.2535$ ,  $\hat{\theta}_2 = 0.7701$ ,  $\hat{\theta}_3 = 0.1299$ ,  $\hat{\theta}_4 = 0.1612$ ,  $\hat{\theta}_5 = 0.0741$ .  
Diagnostics :  $\hat{\sigma}^2 = 10.0471$ .  $\underline{T} = 0.1s$ .

- WWW-I:

$$\text{Model form : } \begin{cases} W_t = Y_t - Y_{t-1} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \end{cases}$$

Parameter estimates :  $\hat{\theta}_1 = 0.6605$ ,  $\hat{\theta}_2 = 0.0902$ . Diagnostics :  $\hat{\sigma}^2 = 4.76256$ .  $\underline{T} = 0.1s$ .

- WWW-II:

$$\text{Model form : } \begin{cases} W_t = Y_t - Y_{t-1} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \end{cases}$$

Parameter estimates :  $\hat{\theta}_1 = 0.6965$ ,  $\hat{\theta}_2 = -0.0460$ . Diagnostics :  $\hat{\sigma}^2 = 6.3027$ .  $\underline{T} = 0.1s$ .

- WWW-III:

$$\text{Model form : } \begin{cases} W_t = Y_t - Y_{t-1} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \end{cases}$$

Parameter estimates :  $\hat{\theta}_1 = 0.5263$ ,  $\hat{\theta}_2 = 0.0888$ . Diagnostics :  $\hat{\sigma}^2 = 6.5063$ .  $\underline{T} = 0.1s$ .

- WWW-IV:

$$\text{Model form : } \begin{cases} W_t = Y_t - Y_{t-1} \\ W_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \end{cases}$$

Parameter estimates :  $\hat{\theta}_1 = 0.5116$ ,  $\hat{\theta}_2 = 0.1089$ . Diagnostics :  $\hat{\sigma}^2 = 2.8577$ .  $\underline{T} = 0.1s$ .