Contributed paper to
the SOEIS meeting
at Bielefeld, March 27-28, 1998

CHAOTIC ANALYSIS OF INTERNET PING DATA: Just a Random Number Generator?

Dimitris Kugiumtzis * & Moses A. Boudourides †

Department of Electrical and
Computer Engineering
Democritus University of Thrace
67 100 Xanthi, Greece

Abstract

The Internet performance is characterized by successive periods of congestion and subsidence according to the bandwidth consumption. To investigate the dynamics of the Internet traffic, we have collected a time series composed of the round trip times for ping packets and we have analyzed it by linear and nonlinear statistical methods, the latter of which are commonly utilized in the analysis of deterministic chaos. Our analysis aims to examine whether this series possesses any structure, i.e. any correlations in the data, or whether the series is completely random. Many of the employed methods suggest that the ping data are white noise of a non-symmetrical distribution. However, some methods, such as the global and local linear prediction as well as the estimation of the largest Lyapunov exponent, indicate that there exists some structure in the data.

1 Introduction

One way to estimate the traffic on the Internet is by measuring the round-trip times for ICMP ping packets sent from one host to a second one. In this way we can obtain a time series (herein called ”ping time series”), which we are going to process through linear statistical methods and methods of chaotic nonlinear analysis. In this way, we intend to investigate whether certain patterns emerge, i.e. there exists some kind of structure in the data, or the data are just the outcome of some random number generator. Certainly, one expects that, if any, the underlying mechanism is a complicated one that cannot be identified using the classical statistical tools, such as autocorrelation and Fourier spectrum analyses.

The advances in chaos theory offer new methods (summed under the term ”nonlinear or chaotic time series analysis”) to handle random-like data, such as the ping time series, without giving up the idea of deterministic structure in the data. This analysis assumes that the underlying mechanism of the

*dimitris.kugiumtzis@iname.com
†mboudour@duth.gr

data is low dimensional and deterministic but with a stochastic outlook, i.e., it is a chaotic mechanism. However, the methods initially stemming from nonlinear dynamics and chaos theory are now used not only to identify pure deterministic nonlinear and chaotic mechanisms but also to pinpoint deterministic elements that are mixed with other stochastic elements in the data. In this case, the assumption is that the mechanism generating the data follows some deterministic dynamic laws, possibly chaotic, but other stochastic factors corrupt its evolution, and the measured data is a mixture of deterministic and stochastic elements. Thus, the tools of nonlinear analysis can be used to verify either of the two hypotheses and to provide some evidence directing towards determinism or stochasticity.

In Section 2 we present a short list of the employed methods and in Section 3 we describe the analyzed ping time series. Subsequently, we proceed to analyze this series by linear methods in Section 4 and nonlinear methods in Section 5. Finally, we discuss the results in Section 6.

2 The Methods

To investigate linear correlation in the ping data we use the following methods from linear analysis:

Autocorrelation $R(\tau)$ **/ Power Spectrum** $P(f)$: The estimation of $R(\tau)$ or equivalently $P(f)$ identifies linear correlations in the data, pronounced either as significant autocorrelations between lagged samples, or respectively, characteristic peaks in the $P(f)$ spectrum.

AR modeling or prediction : If the data exhibit any linear correlations at all, the AR model should show some predictability (give better predictions than the mean value) ([1]).

To investigate nonlinear / general correlations, we use the following tools from nonlinear analysis, typically used to investigate also chaos in the data ([2], [3]).

Mutual Information $I(\tau)$: The function $I(\tau)$ measures the general correlation between lagged samples ([4]). If the correlation in the ping data is nonlinear, $I(\tau)$ should be significantly greater than zero for small lag values (although the $R(\tau)$ may be zero for the same lags).

False Nearest Neighbors (FNN) : This method finds an estimate for the embedding dimension m by counting false nearest neighbors for successively larger m ([5]). If the data are white noise, one expects that the percent of false nearest neigbors does not vanish with the increase of m but it remains at a high level.

Correlation dimension ν : This is an estimate of the dimension of the attractor of the measured system (fractal dimension for strange attractors, i.e., generated from chaotic systems) ([6]). The estimation of ν is valid if there exists a scaling of the correlation integral $C(r)$ for small r and a saturation of the scaling for increasing m. For deterministic systems, both scaling and saturation are expected. For white noise, scaling is found at the order of the applied m, i.e., $\nu \sim m$.

Largest Lyapunov Exponent (LLE) : This is a measure of the divergence of two trajectories when they evolve from close starting points. Meaningful estimations are drawn whenever there exists dynamics in the data. The estimates are found to be sensitive to the method parameters, both regarding the state space reconstruction and the tracking of the trajectories divergence ([7]). For white noise, the LLE should get arbitrary values, and should not be sensitive to parameter changes.

Local Linear Prediction (LLP) : This is a nonlinear prediction model. For each target point, a linear map based only on a given number of neighbors k is implemented to give the predicted value (local AR model) ([8]). As k increases, the model approaches the classical AR model. For nonlinear systems corrupted with noise, the best predictions are expected for moderate values of k ([9]).

Recently, it has been shown in [10] that regularization of the standard ordinary least squares solution (OLS) of the local linear maps, e.g., using the principal component regression (PCR) concept, enhances the predictability of LLP applied to noise data, especially when it is used with large m. The level of regularization is determined by the projection parameter p in PCR.

In our work, we also use this regularization and denote the two models of LLP as OLS and PCR(p), respectively.

Most of the aforementioned nonlinear methods make use of the state space reconstruction, in which m-dimensional points are constructed from the scalar data. Here, we simply use the **Method of Delays (MOD)** setting the delay time parameter τ to 1, since successive data do not seem to have strong correlations ([2], [3]).

3 The Ping Time Series

In this work, we use a time series composed of the round trip times for 64-byte ICMP ping packets sent every 15sec, from the host `demokritos.cc.duth.gr` in Greece to the host `max.eng.uci.edu` in USA. Of the 15000 ping packets sent, 1063 timed out and were removed from the series. Of the remaining 13927 round trip times, five were extremely large and were also neglected (considered to be due to exceptional causes). The derived data set of 13922 samples is denoted hereafter as the p_1 time series and it is shown in Fig.1a.

From Fig.1a, one can notice a data window with characteristically larger values, starting approximately at sample with time index 7600 and lasting over roughly 150 samples. This epoch of large round trip times corresponds to a long congestion period and it is probably due to exogenous factors. We have removed this data window to get the *stationary* ping time series of 13753 samples, denoted hereafter as the p_2 time series (average 295.25 and SD 36.74).

The histogram of the p_1 and p_2 ping data sets, shown in Fig.2, resembles the graph of the lognormal distribution, as pointed out in [11]. The tail corresponds to congestions of varying intensity. Note that the histograms of p_1 and p_2 are identical apart from the tail.

4 Linear Methods of Analysis of the Ping Data

The classical linear tools are applied to both p_1 and p_2 ping data sets.

4.1 Power Spectral Density and Autocorrelation

The data window of long congestion period seems to affect the estimation of the Power Spectral Density $P(f)$ and the autocorrelation $R(\tau)$. For the p_1 time series, $P(f)$ decays gradually for small frequencies (< 0.05) and then it levels out (see Fig.3a). $R(\tau)$ falls to 0.5 with $\tau = 1$ and then it decays very slowly (to 0.4 for $\tau = 50$), indicating significant linear correlations, even over large delays (see Fig.3b). This seems to be just an artifact due to the inclusion of the congested data window. This signature does not occur when $P(f)$ and $R(\tau)$ are estimated on the p_2 time series, and the results rather indicate no linear correlations in the ping data. As shown in Fig3a for the p_2 time series, $P(f)$ is almost flat and $R(\tau)$ falls abruptly to small values, essentially close to zero, even for $\tau = 1$.

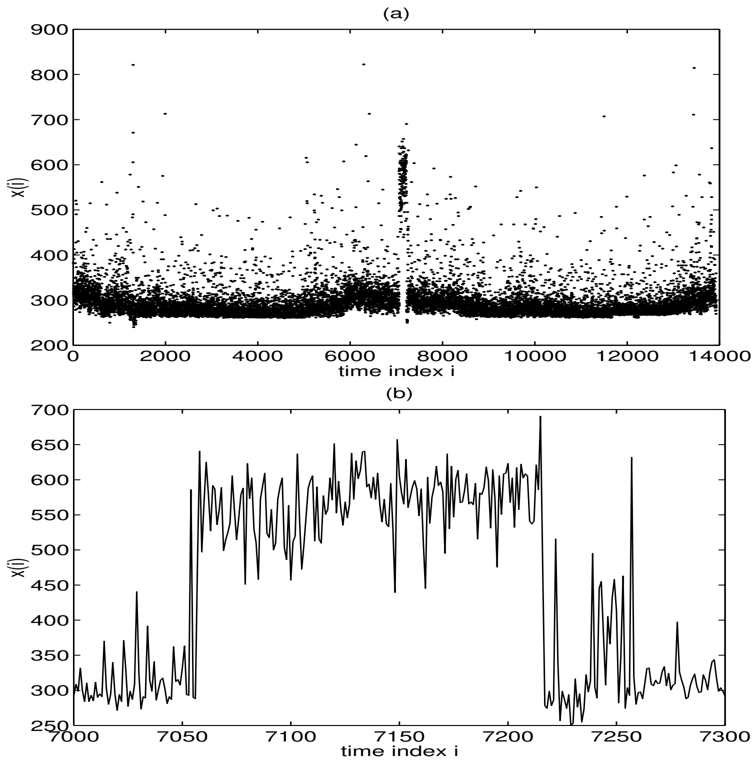

Figure 1: (a) The ping time series of the 13922 samples of travel time of ping packets sent every 15sec (average $= 298.4$, SD $= 46.9$). (b) The segment of the time series in (a) including the data window of long congestion period.

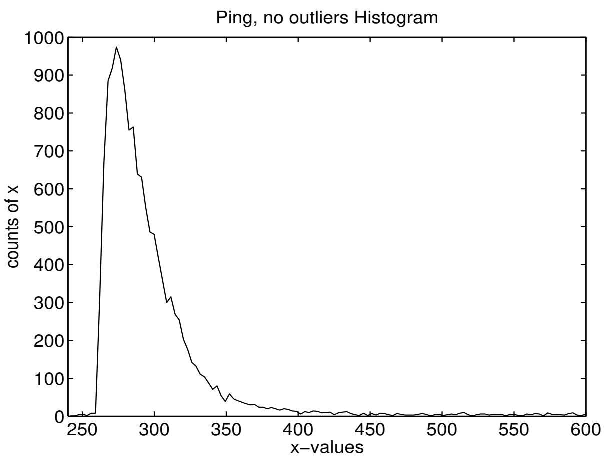

Figure 2: The histogram of the p_1 and p_2 time series using 200 bin resolution.

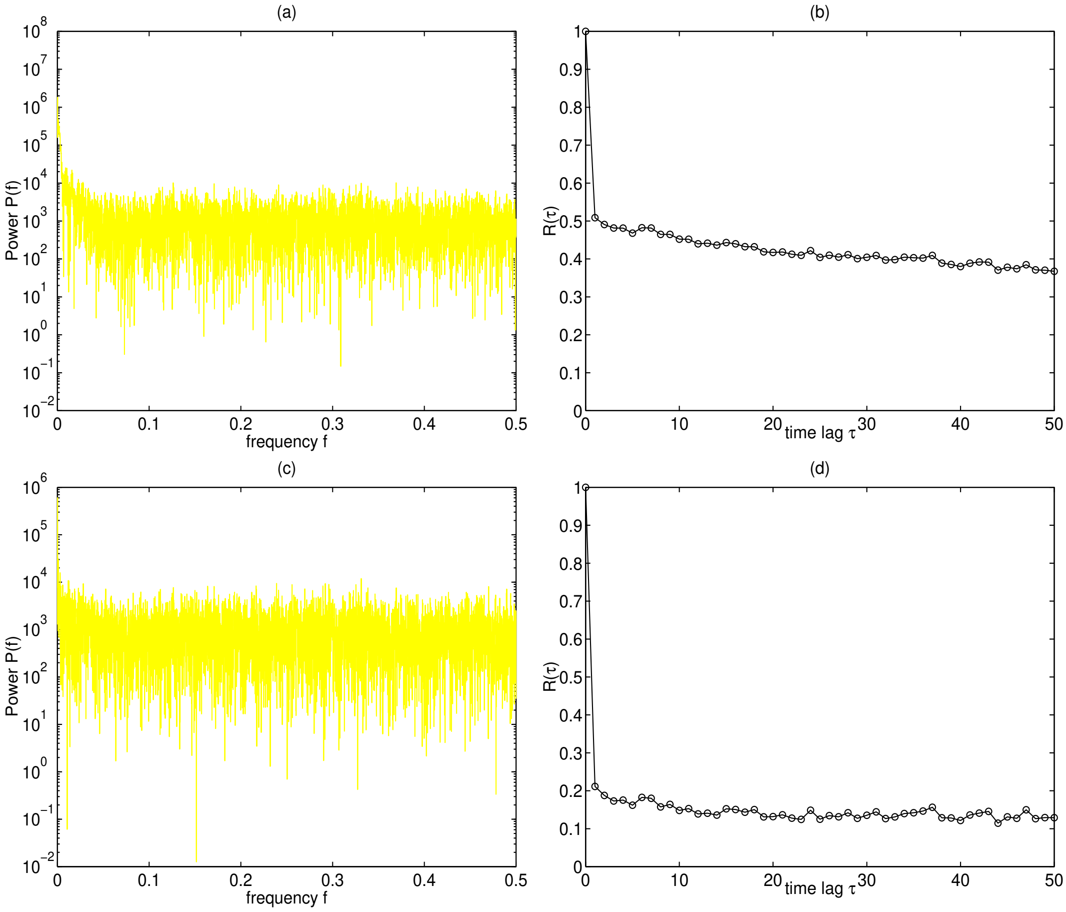

Figure 3: The Power Spectral Density (PSD) and the autocorrelation $R(\tau)$ for the p_1 time series in (a) and (b), and for the p_2 time series in (c) and (d), respectively.

4.2 AR-Model

The prediction with the AR model was applied both to the p_1 and the p_2 time series. The first 75% of each data set was used to estimate the parameters of the AR model, and the rest 25% was used for predictions. Here, only one step ahead prediction was made. The prediction error was measured both with the Normalized Root Mean Square Error (NRMSE) and the Correlation Coefficient (CC). Note that the data window of congestion belongs to the training set, and thus it is not expected to effect significantly the measures of predictability.

The AR model of different order m, applied both to the p_1 and the p_2 time series, did not show any signficant predictability (see Fig.4). However, marginal enhancement of the prediction error can

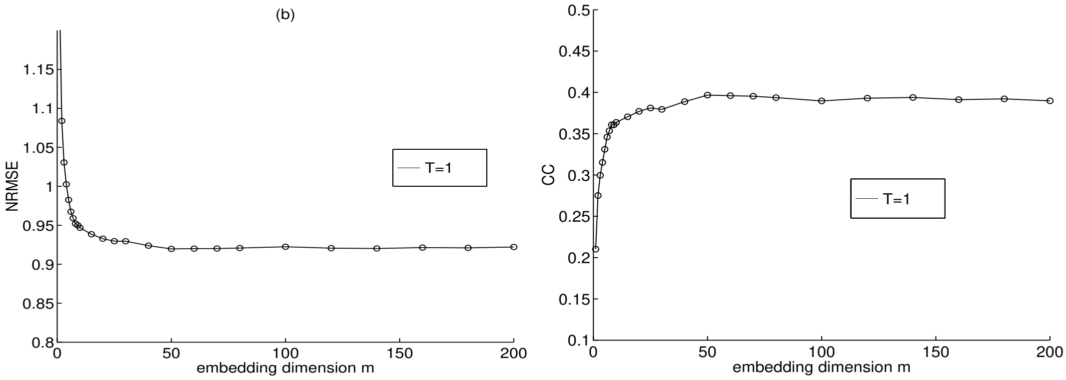

Figure 4: Prediction with the AR-model of different orders for the p_1 and the p_2 time series, as shown in the legend. In (a) the prediction error is measured with NRMSE, and in (b) with CC.

be noticed for larger m.

Certainly, a CC of the order of 0.4 indicates that the data are not white noise, either Gaussian or uniform. However, it is questionable whether this result suggests that there are linear correlations in the ping data, or it is just due to a lognormal-type distribution of the data, i.e., that the data are actually white noise but not with a distribution symmetrical around the mean value. To investigate this, we look closer to the results in Fig.4, choosing the p_2 time series and $m = 50$, which turns out to give marginally the best predictions. In Fig.5a, the one step ahead predictions $\hat{x}_{k+1}$ together with the real values x_{k+1} are shown for a segment of the test set, as well as the prediction error $e_{k+1} = \hat{x}_{k+1} - x_{k+1}$. Obviously, the AR model cannot track the congestions, i.e., the peaks in the real data set, but rather it fluctuates slightly around the mean value. This is also demonstrated in Fig.5b, where the cross correlation is shown as a scatter plot of $\hat{x}_{k+1}$ vs. x_{k+1}. The points in Fig.5b are indeed more concentratedly close to the diagonal but only for the samples around the mean value, however, resulting to $CC \simeq 0.4$. Moreover, this result is to be compared with $CC = 0.21$ derived from the consistent prediction estimate, i.e., $\hat{x}_{k+1} = x_k$ ([2]).

5 Nonlinear Methods

Here, the analysis of the p_1 and the p_2 time series is implemented by the nonlinear models.

5.1 Mutual Information

The estimation of the mutual information $I(\tau)$ shows that there exist no correlations in the ping data. In particular, $I(\tau)$ estimated both for p_1 and p_2 falls abruptly to the zero plateau, even after one lag (see Fig.6). For the p_2 time series, the estimated $I(\tau)$ is similar to the estimated $R(\tau)$ (compare Fig.6

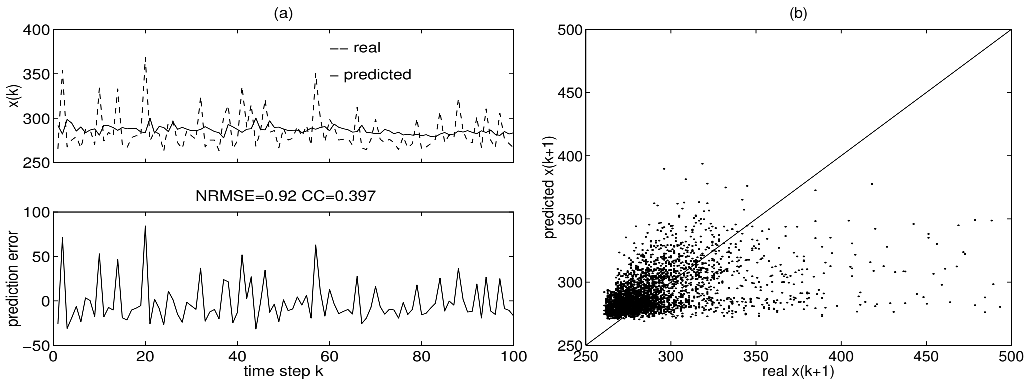

Figure 5: Prediction with the AR-model of the p_2 time series with $m = 50$. (a) Predicted and real values in the first plot and prediction error in the second for a segment of the test set. (b) Scatter plot of the predicted vs. the real values over the whole test set. The magnitude scale at the abscissa is up to 500 to demonstrate better the scattering of points, excluding some large values (> 500).

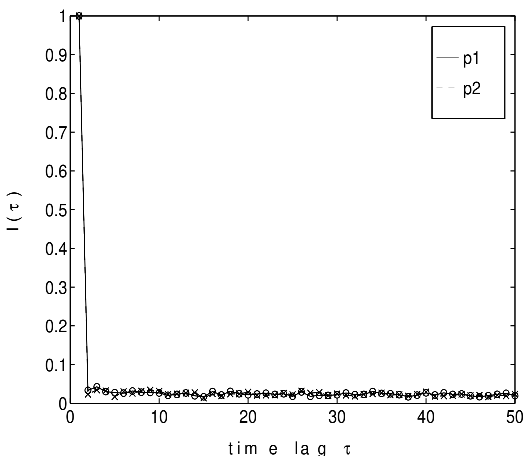

Figure 6: The Mutual Information $I(\tau)$ for the p_1 and the p_2 time series.

with Fig.3b) and both suggest no correlations. Apparently, the inclusion of the congested data window does not alter $I(\tau)$, as it does for $R(\tau)$.

5.2 False Nearest Neighbors

Typically, the algorithm of the False Nearest Neighbors (FNN) for any time series (stochastic or not) gives a decreasing percentage of FNN, as m increases, which finally reaches a plateau at a critical value m^*. Depending on the input time series, differences are seen with respect to the value of m^* and the level of the plateau. The levelling below 1% at a value m^* indicates that the data have prominent deterministic structure and m^* is the proper choice of the embedding dimension, as it involves no false neighbors, i.e., no bad projections of the assumed attractor. A higher levelling indicates significant stochasticity in the data, which dominates over a possible deterministic structure, because regardless of the value of $m > m^*$ a significant percentage of FNN persists. The latter is the case with the p_1 and the p_2 time series; the percentage of FNN falls with increasing m and it stabilizes for $m^* = 6 - 7$ at a level of about 5% for the p_1 data and $7 - 8\%$ for the p_2 data regardless of the choice of τ (see Fig.7a and Fig.7b).

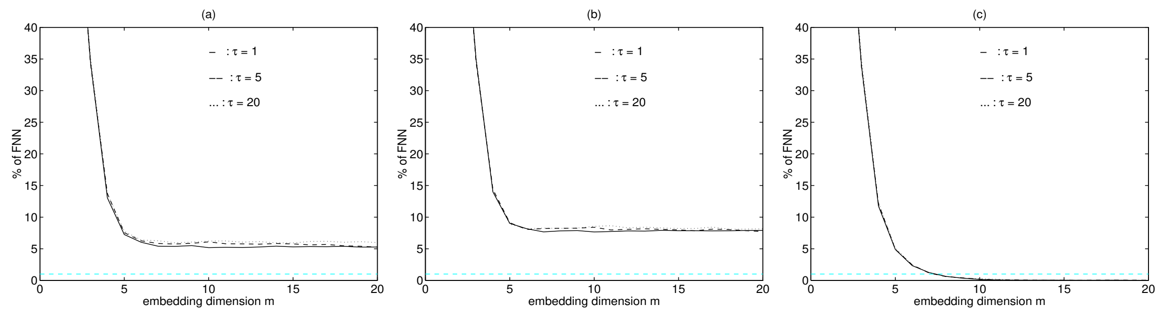

Figure 7: The $\%$ of FNN for increasing m and for three different choices of τ, as shown in the legend, applied to the p_1 data set in (a), the p_2 data set in (b), and the initial data set including the extremely large values in (c).

This result alone may plausibly suggest that the p_1 data are "less stochastic" than the p_2, due to the lower level of plateau. However, knowing the real difference of the p_1 and p_2 data sets, we attribute their discrepancy to the presence of the data window of long congestion period. Note that, when the same setup for FNN is applied to the initial time series of 13927 samples (including the 5 extremely large round trip times), an erronous fall of FNN to zero is observed, reminiscent of the "deterministic case."

In conclusion, FNN is not particularly informative for the detection of any correlations and structure in the ping data. However, the consistent levelling of FNN at $m = 6 - 7$, for a variation of τ values and for both the p_1 and p_2 data sets, suggests that, if "embedding techniques" are to be used (such as those for the estimation of the correlation dimension and the largest Lyapunov exponent), the choice of $m = 6 - 7$ should be the most appropriate. With smaller m, the embedding constitutes a "bad projection," since it gives large $\%$ of FNN and with larger m no improvement is gained, since the $\%$ of FNN remains the same.

5.3 Correlation Dimension

The estimation of a single ν for the ping data is not possible as there is no saturation of the estimated ν with increasing m. Actually, the estimated ν follows closely with m as long as a clear plateau is formed by the slope curves (see Fig.8). A similar feature would result from pure stochastic data of

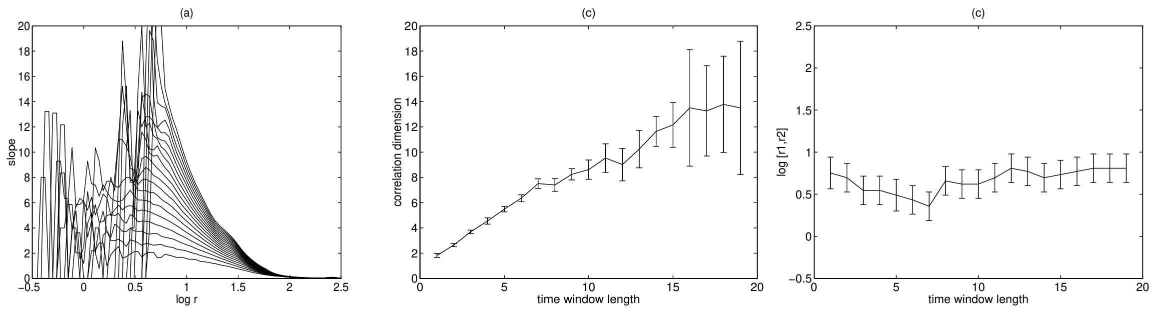

Figure 8: Estimation of ν for the p_1 data set. The state space reconstruction is done with MOD, $\tau = 1$, and $m = 2, \ldots, 20$. The inclusion of autocorrelation points in the computations does not alter the results. In (a), the slope curves for different m are shown. In (b), the estimates of ν computed from the "plateau" of each slope curve are shown (no plateau for large m). Each plotted ν-value is the average of the slope curve segment in the interval $[r_1, r_2]$, where $r_2 = 4 * r_1$, which is found to give the least SD, shown with error bars (r-values with $\log r > 1$ are excluded from the search for the "best" $[r_1, r_2]$). The corresponding $[r_1, r_2]$ intervals are plotted in (c).

limited length. The slope curves for the p_1 and p_2 data sets are identical for small r values and differ slightly for large r. Thus, the ν estimated from the slopes for each m are the same for the two data sets.

5.4 Local Linear Prediction

The local linear prediction (LLP) is applied to the p_1 and p_2 data sets using both the standard ordinary least squares solution (OLS) and the regularization with principal component regression (PCR) to estimate the parameters of each local linear model.

The LLP with OLS applied to the two ping data sets gives worse predictions than the mean value, as shown in Fig.9. The regularization of OLS, done here with PCR(1) and PCR(2), results an enhanced

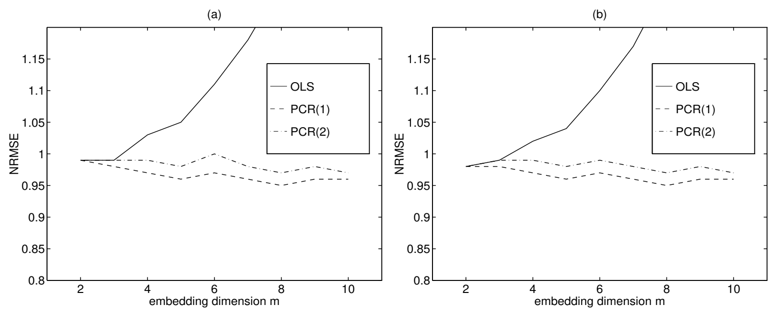

Figure 9: The NRMSE of three local linear predictors (LLP), as shown in the legend, for increasing m. The parameters are $\tau = 1$, $k = 15$, and $T = 1$. In (a) the p_1 data set is used, and in (b) the p_2 data set.

prediction, slightly better than the mean value estimate for increasing m, but still worse than the AR model (compare the results in Fig.9 and Fig.4 assigning the order of AR to the embeding dimension m). Readily, the linear model applied to small regions (which are formed by as few as $k = 15$ neighbors, as in Fig.9) has no predictive power, as one would expect to get if the data were pure chaotic.

A common approach used to assess nonlinearity in noisy data with respect to LLP is to monitor the prediction error sorting all possible number of neighbors k. If only linear correlations can be detected from the data (either because there are no nonlinear interactions or because they are masked by the stochastic components of the evolution mechanism), a global AR model is superior. (Note that the LLP with k at the order of the sample size is essentially a global AR model.) For this type of data, the prediction error of LLP either remains constant or it falls, as k increases. On the other hand, if there are detectable nonlinear correlations in the data, LLP is expected to do better than AR, for a range of k smaller than the sample size, with the prediction error reaching its minimum at a moderate k value, significantly smaller than the sample size. Moreover, this minimum depends on the level of stochasticity in the data as well as on the complexity of the underlying mechanism.

Applying the above algorithm on the ping data shows that the NRMSE prediction error does not exhibit a significant minimum for any k less than the sample size. In fact, the general feature is that NRMSE flattens out at some moderate k value (see Fig.10). The use of regularization improves

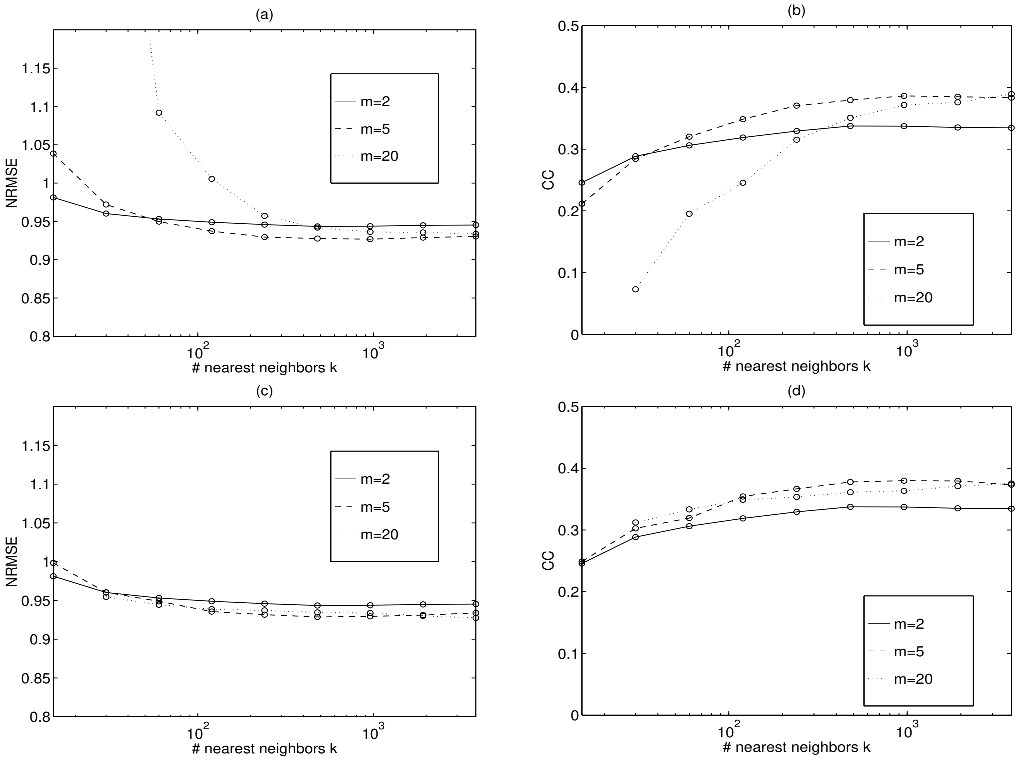

Figure 10: The NRMSE and CC of LLP with three different m, as shown in the legend, for increasing k. The parameters are $\tau = 1$, and $T = 1$, using the p_2 time series. In (a) and (b) the NRMSE and CC is shown for the LLP with OLS model and in (c) and (d) the same measures are shown for the LLP with the PCR(3) model.

prediction at large m, especially for small k, but otherwise it does not change the reults (compare Fig.10a and Fig.10b with Fig.10c and Fig.10d). For $m = 2$ and $m = 5$, best predictions are obtained for the largest k, i.e., the global AR model is superior. For $m = 20$, a slight enhancement of prediction is observed for k values between 10 to 50 times smaller than the sample size. However, this is not considered as a strong evidence of nonlinearity since the minimum is not significant.

5.5 Largest Lyapunov Exponent

The method of estimation of the largest Lyapunov exponent (LLE) is applied to the p_1 and p_2 data sets using different state space reconstructions as m and τ vary. (The parameters of the method are set as follows: minimum distance displacement = 1, maximum distance displacement = 30, evolution time = 3, maximum orientation error = 30).

The LLE is always found positive (as expected) for all reconstructions and for both data sets, but it exhibits a characteristic change with m. For any choice of τ, the LLE falls with m (starting with $m = 2$) at a minimum attained for $m = 6$ or 7, and then it increases (see Fig.11).

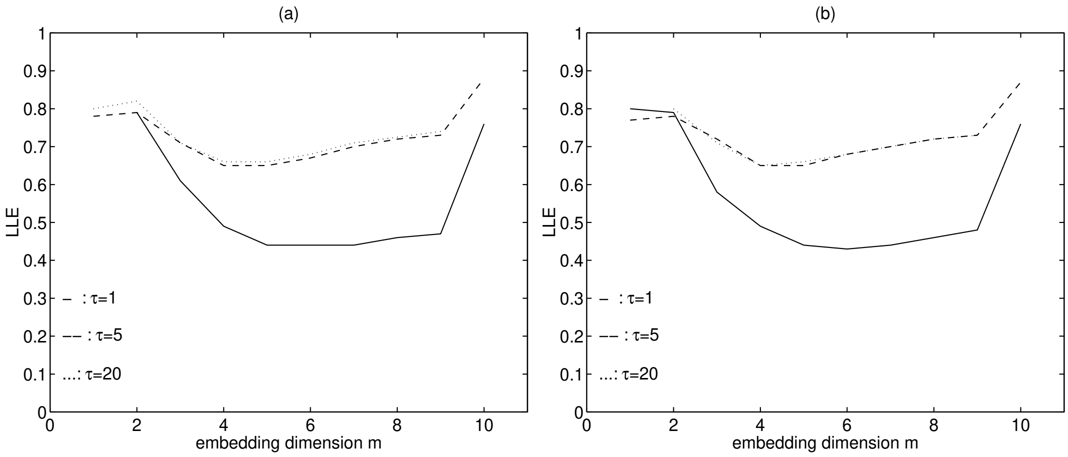

Figure 11: The LLE for increasing m for three different choices of τ, as shown in the legend, applied to the p_1 data set in (a) and to the p_2 data set in (c).

The results for the LLE do not support the hypothesis of white noise. If the data were white noise, such structural change with m should not be detected.

6 Discussion

The ping time series (measured at the scale of 15sec) follows a repeated pattern which is composed of a background, corresponding to a subsided Internet traffic, succeeded by a sudden peak, corresponding to a short congestion period of the order of a minute or less. In the measured set of round trip times, a data window of larger magnitude is detected corresponding to a large congestion period of more than half an hour, during which the Internet is overloaded. The presence of this data window introduces nonstationarity and, therefore, it alters the results of some of the applied methods. Particularly, the inclusion of the long congestion period results to substantially larger estimated linear correlations over many lags, which is apparently a misleading result. Nevertheless, removing the long congestion period, no correlation is detected. Moreover, the FNN method suggests erronously that there is more structure in the long congestion period data.

The results of the applied methods do not give a definite evidence whether the ping data are simply white noise or not. Obviously, the data follow a nonsymmetrical distribution that resembles the log-normal distribution and, so, the data are not white noise of a Gaussian or uniform type. Some methods, such as the estimation of the autocorrelation, mutual information and correlation dimension, suggest that there is no correlation and structure in the ping data. On the other hand, the prediction with AR and LLP, as well as the LLE method, indicate that there are small correlations and some structure in the data, as they give different results than those expected for the white noise case.

In order to further investigate the question of white noise, as well as to understand the results of the LLE method on the ping data, one should make a hypothesis testing, starting with the hypothesis that the ping data are white noise of lognormal distribution. To test this hypothesis, a number of white noise data with the same distribution as the original data should be generated, e.g., by scrambling the original ping time series. Then, one or more methods could be used as discriminating statistics to assess whether there is any statistically significant difference of the estimates for the original and the surrogate data. (This is currently an ongoing work.)

The ping dynamics at other time scales might be interesting, i.e., measuring the Internet traffic by pinging at the order of minutes, hours or even days. These time scales might be more suitable to find correlations in the data. For example, seasonal patterns are expected to be present in the ping time series measured over weeks, due to different traffic on weekends, and over days, due to different traffic during day and night.

Acknowledgements

We are grateful to Leonidas Karakatsanis for providing us with the ping data processed in this work.

References

[1] M. B. Priestley. *Non-linear and Non-stationary Time Series Analysis*. Academic Press, 1988.

[2] A. A. Tsonis. *Chaos: From Theory to Applications*. Plenum Press, New York, 1992.

[3] D. Kugiumtzis, B. Lillekjendlie, and N. Christophersen. Chaotic time series part I: Estimation of some invariant properties in state space. *Modeling, Identification and Control*, 15(4):205 – 224, 1994.

[4] A. M. Fraser and H. Swinney. Independent coordinates for strange attractors from mutual information. *Physical Review A*, 33:1134 – 1140, 1986.

[5] M. B. Kennel, R. Brown, and H. D. I. Abarbanel. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A*, 45:3403 – 3411, 1992.

[6] P. Grassberger and I. Procaccia. Measuring the strangeness of strange attractors. *Physica D*, 9:189 – 208, 1983.

[7] A. Wolf, J. Swift, H. Swinney, and J. Vastano. Determining Lyapunov exponents from a time series. *Physica D*, 16:285 – 317, 1985.

[8] J. D. Farmer and J. J. Sidorowich. Predicting chaotic time series. *Physical Review Letters*, 59:845 – 848, 1987.

[9] M. Casdagli. Chaos and deterministic versus stochastic nonlinear modeling. *Journal of the Royal Statistical Society Series B-Methodological*, 54:303 – 328, 1992.

[10] D. Kugiumtzis, O. C. Lingjærde, and N. Christophersen. Regularized local linear prediction of chaotic time series. *submitted to Physica D*, 1997.

[11] B. A. Huberman and R. M. Lukose. Social dilemmas and internet congestion. *Science*, 277:535 – 537, 1997.