On Appropriate Intervals for Active Measurement of Packet Transmission Delays

Katsunori Ori[1], Takafumi Okuyama[1], and Yutaka Takahashi[2]

[1]NTT Service Integration Laboratories, Musashino-shi, Tokyo 180-8585, Japan
[2]Graduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan
E-mail: {k.ori,okuyama.takafumi}@lab.ntt.co.jp, takahashi@i.kyoto-u.ac.jp

***Abstract*—Quality of service requirements for IP networks are becoming higher. Accordingly, managing, measuring, and monitoring IP networks are also becoming more important. In particular, for real-time communication, the quality of packet transmission (packet loss, delay, and delay variation) needs to be carefully maintained. In managing these performance parameters, active measurements have been often used, and measured samples are periodically evaluated statistically. Although performance measurements are standardized by several organizations, such as ITU-T and IETF, the length of the measurement intervals have not been mentioned. In this paper, in order to consider its appropriate length, we conduct periodogram analysis of packet transmission delay and investigate its stochastic characteristics in the frequency domain. After considering long-range dependency and randomness arising in packet transmission delay, we develop a method to determine the optimal measurement period for statistification. We applied our method to measured samples in an actual network and verified the effectiveness of our method.**

I. Introduction

In IP networks in the past, the main application was transmission of large files, and as a result, the principal performance measure was file transmission time. Quality of each packet transmission (delay and/or loss) was not an important concern because network impairments for short intervals did not significantly affect performance. Today, quasi-real-time traffic, such as that associated with WWW pages, and real-time traffic, such as transmission of voice and/or video, are becoming popular in IP networks. These applications are severely affected by packet-level quality of service (QoS), and monitoring QoS is therefore becoming important.

In conventional QoS management, resource utilization of links and/or nodes is a major concern. Collecting the actual data for these performance measures is relatively easy and enables us to recognize the occurrence of congestion and locate the cause of impairment. However, current performance-aware applications require packet-level QoS monitoring. For this purpose, active and end-to-end measurements are often used to facilitate more detailed resource management. Sampled delay and loss statistics are usually used as performance measures of packet-transmission quality. As ITU and IETF pointed out, the average and variation of packet transmission delay and packet loss rate are strongly concerned with packet-transmission quality [1], [2], [3], [4].

For active measurement, many packets are sent sequentially and whether each packet is lost or not is measured. Unless the packet is lost, transmission delay is measured and periodically analyzed as statistics. To this end, there are several important parameters to be decided, such as the generation rate and total number of probe packets as well as the unit interval for statistical analysis. However, to the best of the authors' knowledge, no standardized methods have been recommended. Of these important parameters, we focus on the length of a unit observation interval for statistical analysis. We call this length a statistification interval. So far, the length has been determined according to the experience of network managers and/or system-specific restrictions. However, the length plays a crucial role in balancing the sensitivity of detecting impairment in networks and the incurred traffic of probe packets. If it is set too long, congestion and/or malfunctions will not be detected promptly. If it is too short, the bandwidth used for probe packets will affect the performance of regular packets significantly because the number of probe packets needs to be kept constant for each statistification interval regardless of the length of the interval.

Among different types of traffic, TCP and video traffic is bursty and its volume has long-range dependency [5], [6], [7], [8] and these stochastic characteristics are known to degrade the performance of networks [9], [10], [11], [12]. In those works, the strength of long-range dependency in traffic volume is estimated from long-haul traffic measurements. As for packet transmission quality, loss and delay are evaluated under the long-range dependent traffic on condition that the network is in a steady state. Packet delay measured in the Internet also has long-range dependency in a steady state [13], [14], [15]. On the other hand, the short-range characteristics of Internet traffic can be approximated by white noise or the Poisson process [16], [17].

We describe a method to determine the length of a statistification interval for monitoring based on periodogram analysis and demonstrate its effectiveness through numerical examples. First, we evaluate the short-range characteristics of packet transmission quality of Internet traffic using periodogram analysis and show randomness within short intervals. Samples were collected using a network simulation with TCP self-similar background traffic. Using the short-range random characteristics of packet transmission delay, we develop a method for finding the optimal statistification interval.

The remainder of this paper is organized as follows. In section II, we describe our simulation model for collecting sample data. We conduct periodogram analysis to concretely demonstrate the model's usefulness. Section III shows the periodogram analysis of packet transmission delay and describes a method to determine the statistification interval using the result of the analysis. We also apply our method to the measured samples in an actual network. In section IV, we show the numerical results and verify the feasibility of our method. Section V concludes this paper.

II. Simulations

Our simulation model to collect sample data for statistical analysis is shown in Fig. 1. Servers (S1-S10) and clients (C1-C32) are connected via routers (R1, R2). Each client downloads files from one of the servers using TCP(Reno) as the transport protocol. Traffic to clients is multiplexed at router R1. When the sum of traffic rates sent from servers is more than the bandwidth of the WAN, packets are buffered into an FIFO queue at router R1. We denote the bandwidth of WAN as B. Simulation experiments were carried out using ns2, and the performance of the model was monitored by active measurement.

To generate self-similar traffic, the downloaded file size, M, is assumed to be independent and identically distributed in accordance with a Pareto distribution with shape parameter α and location parameter β, $\mathrm{P}[M < x] = 1 - (\beta/x)^{\alpha}$. We use $\alpha = 1.1$ and set β so that $E[M] = m$. For each download, each client chooses a server randomly. After the download, each client starts an idle period, whose length is assumed to be exponentially distributed in accordance with parameter λ_i, before the next download. As λ_i increases, idle periods become longer and WAN utilization decreases. In the simulation results, we change WAN utilization by changing λ_i. Local access delays to R1 and R2 are assumed to be 1 and 2 ms, respectively.

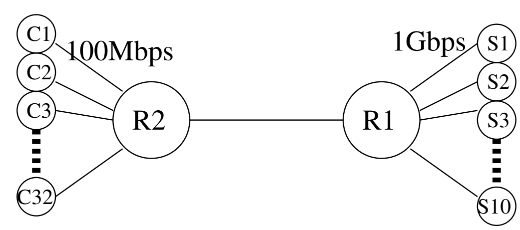

Fig. 1. Network model

We measure ping round trip time (RTT) between R2 and R1 and conduct periodogram analysis. The ping packets are sent from R1 every 10 ms. The major traffic is the downloading of files from servers by clients. The primary factor that increases RTT is the queueing delay at R1, and the variance of delay on each of the other hops is almost negligible. We use the RTT of ping measurements as the packet transmission delay at R1.

Round trip time is shown as a time series in Fig. 2 for the case $m = 1$ Mbyte and $B = 1$ Gbps. The utilization of the WAN is 16.1% when the idle time is set properly. It is difficult to grasp the long-term behavior of RTTs from the figure. Thus, we take the average RTT for every 10 sec, obtaining 1000 samples in total, as shown in Fig. 3. Macroscopic behavior of the RTTs, which is hard to identify in Fig. 2, can be seen. Although we use 10 sec for statistification, this may not be the best length.

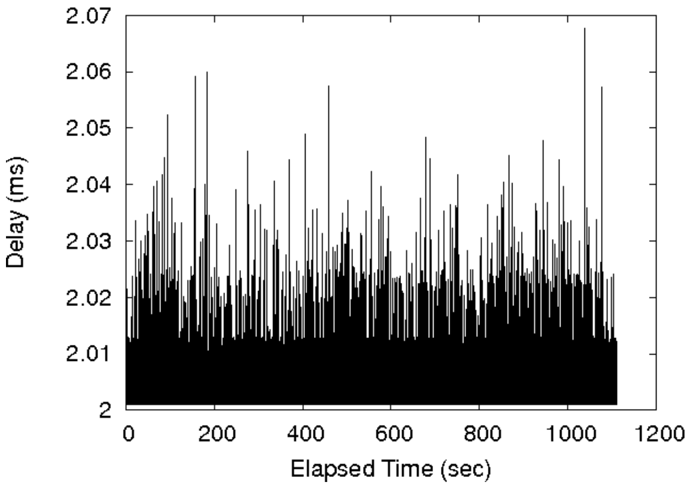

Fig. 2. Packet transmission delays: raw data

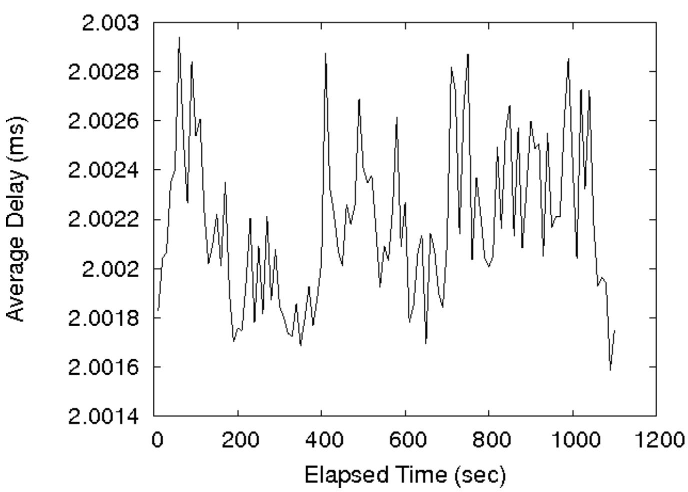

Fig. 3. Packet transmission delays: averaged for every 10 sec

III. Periodogram analysis of RTT

We conduct periodogram analysis of the measured RTT. Periodogram [18] analysis is often used to estimate the Hurst parameter [11], [19], which is a measure of the strength of long-range dependency.

A time series is defined to be long-range dependent if $\Gamma(\nu)$, which is its spectrum density function as a function of frequency ν of the set, has the following property

$$\Gamma(\nu) = c|\nu|^{-\alpha}, \qquad \nu \to 0, \tag{1}$$

where c denotes a constant and α represents the strength of the long-range dependency expressed as $\alpha = 2H - 1$ using the Hurst parameter H. Note that $-1 < \alpha < 1$ and long-range dependency is strong when α is close to 1.

A periodogram of the RTT shown in Fig. 2 is represented in Fig. 4. We apply a discrete Fourier transform [20] to the measured delay data and show it as a log-log plot with frequency on the x-axis and relative power on the y-axis. Local polynomial

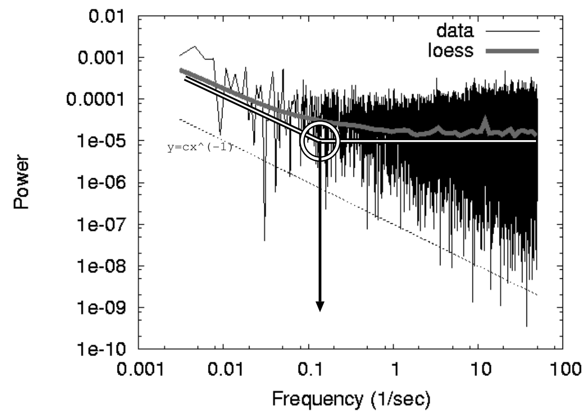

Fig. 4. Periodogram of measured RTT

regression (LOESS) and a line of $y = cx^{-1}$ are shown together. In the low-frequency range ($\nu \to 0$), the slope is almost -1 and strong long-range dependency is seen. This trend changes when ν is almost 1/8 (cycle length is 8 sec). When the frequency is higher than this, the spectrum power of the RTT remains almost constant, which means power is almost constant at these frequencies and the measured data exhibit randomness such as white noise [21]. For proper network management, long-term characteristics arising from long-range dependency are more important to monitor than short-range ones caused by randomness. For that purpose, it is helpful to find an appropriate statistification interval for on-line monitoring which aims for not microscopic randomness but macroscopic change arising in the performance measures of target traffic. In the above example, an appropriate statistification interval is 8 sec.

As shown in Fig. 5, the average and percentiles of RTT are calculated using the statistification interval and its original raw data is shown in Fig. 2. Although macroscopic change is difficult to see in Fig. 2, we can easily observe it in Fig. 5.

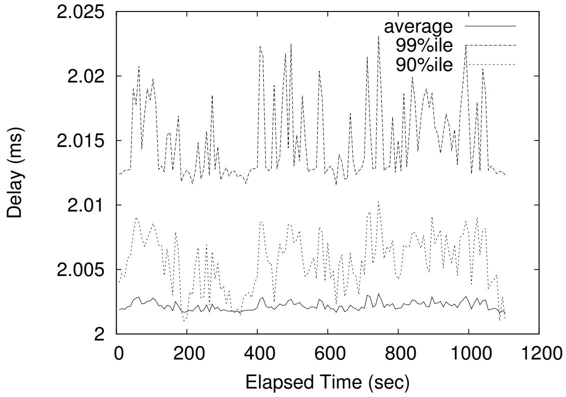

Fig. 5. RTT with statistification

When traffic volume from R1 to R2 is more than the available bandwidth, ping packet delay at R1 and RTT are expected to have a stronger correlation with higher utilization of the WAN. A periodogram of WAN downstream traffic volume (from R1 to R2) is shown in Fig. 6. In the low-frequency range, a slope near -1, which is a characteristic of long-range dependency, is observed. In contrast to Fig. 4, the slope remains the same in the region where ν is small. This means that, although packet transmission delay has the same characteristic as the traffic intensity with respect to long-range dependency, the short-range dependency of packet transmission delay is different from that of the traffic intensity. The short-range characteristics of Internet traffic

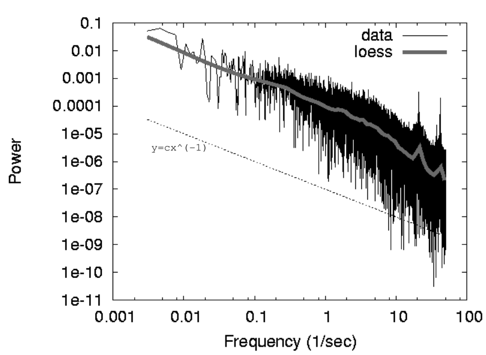

Fig. 6. Periodogram of WAN traffic intensity

can be approximated by white noise or the Poisson process [16], [17].

Comparing Figs. 4 and 6, it is found that the short-range dependency region is narrower in RTT than that in traffic volume.

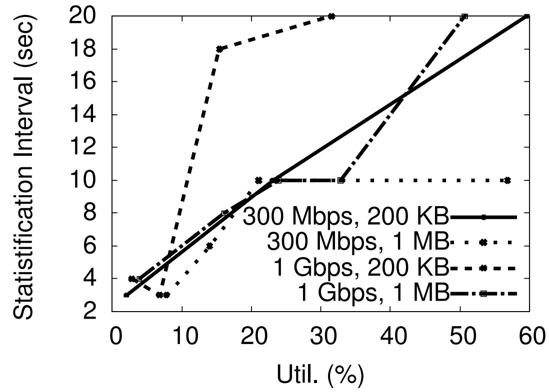

Fig. 7. Statistification interval with various parameters

Here, we examine how the boundary cycle length changes as simulation parameters, e.g., bandwidth and file size, change. Statistification intervals with various file sizes and bandwidths of the WAN are shown in Fig. 7, where the statistification interval ranges from 2 to 20 sec. When the utilization of a WAN increases, the boundary cycle length becomes longer. It is more clearly observed for a case where either the bandwidth or the file size increases.

IV. Application to Actual Network

We apply our analytical method to real RTT data measured in an actual operating network including an LAN gateway link to the Internet (Fig. 8). The RTT was measured every second for two days. Identifying a trend or sequential dependency in this figure is difficult. The periodogram of the data in Fig. 8 is shown in Fig. 9. A slope is observed when the cycle length is longer than 10 min, and there is long-range dependency. We can observe relatively strong power when the cycle length is 5 min, which means the RTT changes periodically with a 5-min interval. We also observe some peaks at 1 and 3 min. When the length is shorter than 10 sec, the power does not change and we can observe randomness. By calculating statistics every 10 sec, we can smooth the random variation and recognize the microscopic change with 1- or 3-min cycles. The changes in the performance seem to be caused by the effect of cyclic traffic for network management, i.e., exchanging dynamic routing information, and polling management information base (MIB), for example, which was not taken into consideration in our simulation experiments presented in section II.

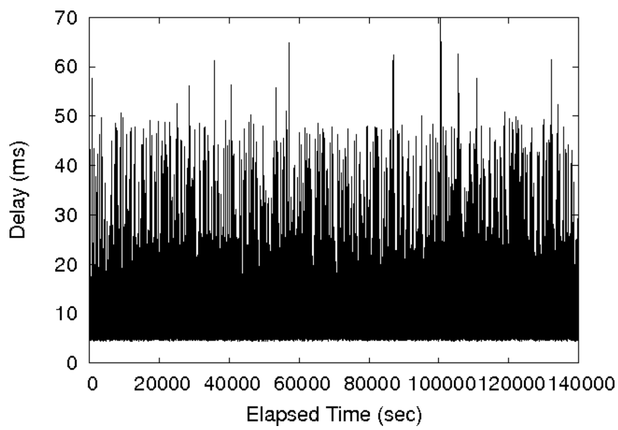

Fig. 8. Measured RTT in an actual network

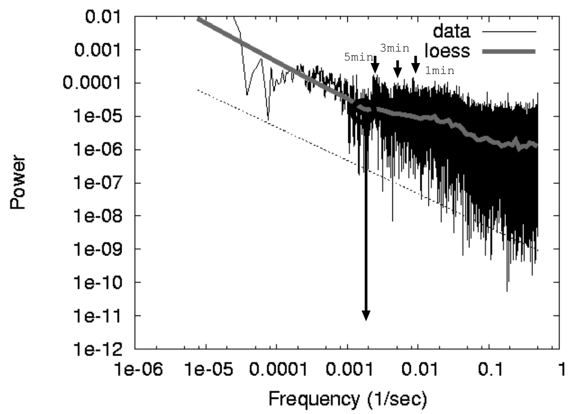

Fig. 9. Periodogram of measured RTT in actual network

The 5- and 60-min averages of RTT used in Fig. 8 are shown in Figs. 10 and 11, respectively. Although we can see the fluctuation of performance in Fig. 10, this is barely seen in Fig. 11 due to the smoothing effect.

Using the method introduced in this paper, we compare the relative power of each frequency factor. The data used in this section was measured for two days and the power was compared within that period. If we measure for a longer duration, we may not be able to observe long-range dependency clearly because of daily or weekly fluctuation. In this case, we focus only on randomness in short

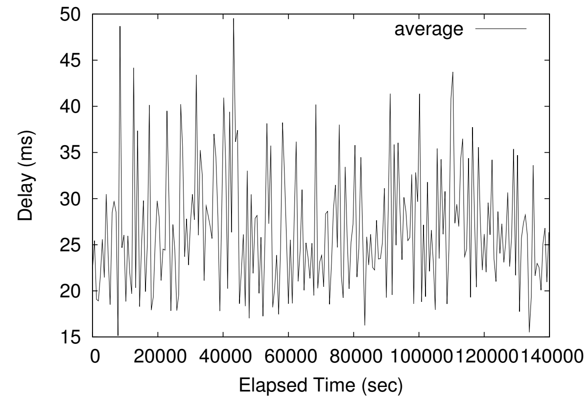

Fig. 10. 5-min average of RTT

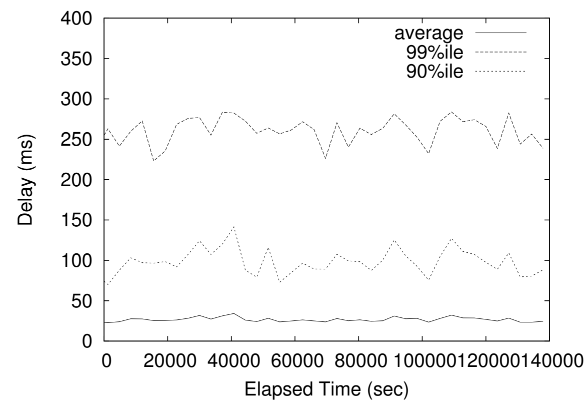

Fig. 11. 60-min average of RTT

intervals and use the maximum interval in which we observe randomness.

V. Conclusion

When conducting active measurement of packet transmission delay for network management, we carry out statistical analysis periodically for proper operation. We have proposed a method to determine the appropriate length of sample intervals for statistification. Our method uses the periodogram and analyzes the measured data in the frequency domain. We have divided the frequency into two parts: the longer range, which has characteristics of long-range dependency, and the shorter range, which exhibits the characteristics of white noise and random fluctuation. By calculating statistics with a statistification interval, it is possible to ignore random fluctuation and grasp the characteristics of long-range variation. We have analyzed the effect of network parameters, e.g., utilization, bandwidth, and file size of background traffic on performance measures by simulation. Finally, we have applied our method to an operating network and verified the usefulness of our method. Moreover, we have discussed the effect of periodic traffic on measured data and presented how to circumvent its effect to conduct proper performance measurement.

ACKNOWLEDGEMENT

One of the authors expresses his acknowledgement for the support by JSPS KAKENHI(18360181) Grant-in-Aid for Scientific Research (B).

REFERENCES

[1] "Internet protocol data communication service - IP packet transfer and availability performance parameters," ITU-T Recommendation Y.1540, 2007.

[2] G. Almes, S. Kalidindi, and M. Zekauskas, "A one-way delay metric for IPPM," IETF RFC 2679, Sep. 1999.

[3] C. Demichelis and P. Chimento, "IP packet delay variation metric for IP performance metrics (IPPM)," IETF RFC 3393, Nov. 2002.

[4] G. Almes, S. Kalidindi, and M. Zekauskas, "A one-way packet loss metric for IPPM," IETF RFC 2680, Sep. 1999.

[5] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Netw.*, vol. 2, no. 1, pp. 1–15, 1994.

[6] V. Paxson and S. Floyd, "Wide area traffic: the failure of Poisson modeling," *IEEE/ACM Trans. Netw.*, vol. 3, no. 3, pp. 226–244, 1995.

[7] M. Crovella and A. Bestavros, "Self-similarity in World Wide Web traffic: evidence and possible causes," *IEEE/ACM Trans. Netw.*, vol. 5, no. 6, pp. 835–846, Dec. 1997.

[8] J. Beran, R. Sherman, M. Taqqu, and W. Willinger, "Long-range dependence in variable-bit-rate video traffic," *IEEE Trans. Commun.*, vol. 43, no. 234, pp. 1566–1579, Feb./Mar./Apr. 1995.

[9] A. L. Neidhardt and J. L. Wang, "The concept of relevant time scales and its application to queuing analysis of self-similar traffic (or is Hurst naughty or nice?)," *SIGMETRICS Perform. Eval. Rev.*, vol. 26, no. 1, pp. 222–232, 1998.

[10] K. Park, G. Kim, and M. Crovella, "On the relationship between file sizes, transport protocols, and self-similar network traffic," *Proc. IEEE International Conference on Network Protocols*, pp. 171–180, 1996.

[11] K. Park and W. Willinger, *Self-similar Network Traffic and Performance Evaluation*. Wiley, New York, 2000.

[12] Y. Sumida, H. Ohsaki, M. Murata, and H. Miyahara, "Effects of upper-layer protocols on the self-similarity of network traffic," *IEICE Trans. Commun. (in Japanese)*, vol. J82-B, no. 6, pp. 1126–1137, 1999.

[13] Q. Li and D. Mills, "On the long-range dependence of packet round-trip delays in Internet," *Proc. IEEE ICC 98*, vol. 2, pp. 1185–1191 vol.2, Jun. 1998.

[14] H. Borella and G. Brewster, "Measurement and analysis of long-range dependent behavior of Internet packet delay," *Proc. IEEE INFOCOM98*, vol. 2, pp. 497–504 vol.2, Mar-2 Apr 1998.

[15] K. Papagiannaki, S. Moon, C. Fraleigh, P. Thiran, and C. Diot, "Measurement and analysis of single-hop delay on an IP backbone network," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 6, pp. 908–921, Aug. 2003.

[16] H. Furuya, H. Nakamura, S. Nomoto, and T. Takine, "Local Poisson property of aggregated IP traffic," *IEICE Trans. Commun.*, vol. 86, no. 8, pp. 2368–2376, 2003.

[17] J. Cao, W. S. Cleveland, D. Lin, and D. X. Sun, "Internet traffic tends toward Poisson and independent as the load increases," in *Nonlinear Estimation and Classification*. Springer, 2002, pp. 83–110.

[18] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*. Springer-Verlag, New York, 1996.

[19] T. Karagiannis, M. Faloutsos, and M. Molle, "A user-friendly self-similarity analysis tool," *SIGCOMM Comput. Commun. Rev.*, vol. 33, no. 3, pp. 81–93, 2003.

[20] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C (2nd ed.): The Art of Scientific Computing*. Cambridge University Press, New York, 1992.

[21] D. Pollock, *A Handbook of Time-series Analysis, Signal Processing and Dynamics*. Academic Press, 1999.