Statistical Modelling 2012; **12**(**1**): 29–65

Estimation of traffic matrices in the presence of long memory traffic

PL Conti[1], L De Giovanni[2] and M Naldi[3]
[1]Sapienza Università di Roma, Dipartimento di Statistica, Probabilità e Statistiche Applicate
[2]Università LUMSA, Facoltà di Scienze della Formazione and Università LUISS Guido Carli
[3]Università di Roma 'Tor Vergata', Dipartimento di Informatica, Sistemi e Produzione

Abstract: The estimation of traffic matrices in a communications network on the basis of a set of traffic measurements on the network links is a well-known problem, for which a number of solutions have been proposed when the traffic does not show dependence over time, as in the case of the Poisson process. However, extensive measurements campaigns conducted on IP networks have shown that the traffic exhibits long range dependence. Here a method is proposed for the estimation of traffic matrices in the case of long range dependence, and its theoretical properties are studied. Its merits are then evaluated via a simulation study. Finally, an application to real data is provided.

Key words: network tomography; traffic estimation; self-similarity; long range dependence

Received September 2009; **revised** August 2010; **accepted** November 2010

1 Introduction

Traffic matrices play a crucial role in network management and provisioning. They describe the amount of bits (packets) transmitted between every Source–Destination (S–D) pair. If compared to other forms of network traffic representation (such as path matrices or measures on links) traffic matrices have the important advantage to be invariant under changes of either the network topology or routing (see Bear, 1988).

Direct measurement of the traffic matrix elements is not usually feasible. It is customary to overcome this inconvenience by resorting to indirect estimates of the traffic matrix elements via measurements of the traffic on the links. Of course, this requires the knowledge of the routing configuration.

Address for correspondence: Pier Luigi Conti, Dipartimento di Statistica, Probabilità e Statistiche Applicate; Sapienza Università di Roma; P.le A. Moro, 5; 00185 Roma; Italy. E-mail: pierluigi.conti@uniroma1.it

 10.1177/1471082X1001200103

In a network with n nodes there are typically (at most) $N = n(n-1)$ S–D pairs, but only M links, with M considerably smaller than N. Hence, there is a many-to-one mapping relating the (expected) traffic on links to the (expected) S–D traffic. In a sense, the information produced by observations on links is not enough in order to identify the S–D traffic. This means we are facing with an incomplete information (or, equivalently, an under-constrained problem).

The approaches to S–D traffic matrix estimation under incomplete information are either based on optimization techniques or on statistical inference techniques. Proposals based on optimization techniques rely on the idea of reducing the space of solutions by appropriate constraints on S–D traffic (see Goldschmidt, 2000; Juva *et al.*, 2006). Proposals based on statistical inference (known as *network tomography* techniques) are based on probabilistic models for S–D traffic, and aim at estimating appropriate parameters via either the maximum likelihood method (Vardi, 1996; Cao *et al.*, 2000; Bermolen *et al.*, 2006) or via Bayesian methods (Tebaldi and West, 1988; Vaton and Gravey, 2002). In Cao *et al.* (2000) a functional mean variance relation of S–D traffic guarantees identifiability under special assumptions on the network topology.

All the above-mentioned works assume that S–D counts are independent Gaussian over S–D pairs and independent and identically distributed (*i.i.d.*) within a S–D pair over successive measurements periods. More formally they are based on the following assumptions.

1. S–D pairs are independent.
2. The traffic produced by a single S–D pair is stationary Gaussian.
3. The traffic produced in different time intervals by a S–D pair is uncorrelated.

As it will be seen in Section 2, empirical analysis of observed traffic, as well as theoretical studies, suggests that traffic data are strongly correlated over time, and hence that the third assumption listed above is false.

Essentially motivated by the presence of strong correlation in real traffic data, in the present article a model for S–D traffic incorporating such a feature is considered. Its formal assumptions are listed in Section 2, and discussed in detail. In Section 3, statistical inference problems for our model are studied. In Section 4, a simulation study is performed. Finally, in Section 5 an application to real data is provided.

Before closing this section, we must say that, as far as we know, the first paper where the independence assumption is removed is Conti *et al.* (2009), where a model with long memory is introduced. The estimation method considered in that paper is essentially the maximum likelihood method, based on the reconstruction of unobserved data via the EM algorithm. However, in the above-mentioned paper there are several critical statistical aspects. First of all, the model used in Conti *et al.* (2009) is not necessarily identifiable. Second, the theoretical properties of the estimators are not studied. Third, the estimators developed in the present article are more efficient, both computationally and in terms of mean square error (see Section 4).

2 The model

As mentioned in the introduction, both empirical evidence and theoretical analysis show that there is a strong correlation (slowly decreasing over time) among traffic observed in non-overlapping time intervals. This motivates the introduction of a new model based on long range dependence (LRD, for short) in the statistical analysis of S–D traffic.

2.1 Model description

Let X_i^t be the traffic for the S–D pair i at time slot t, and let

$$X^t = (X_1^t, \ldots, X_N^t) \tag{2.1}$$

be the vector of traffic for all N S–D pairs at time t. Our assumptions are listed below.

- $A1$ The stochastic process $(X^t;\ t \geq 1)$ is a stationary Gaussian process, with

$$E[X_i^t] = \mu_{Xi},\ i = 1, \ldots, N;\ t \geq 1,$$
$$V[X_i^t] = \sigma_{Xi}^2,\ i = 1, \ldots, N;\ t \geq 1.$$

- $A2$ Different S–D pairs generate independent traffic

$$C[X_i^t,\ X_j^{t+k}] = 0,\ i \neq j,\ i,\ j = 1, \ldots, N;\ t \geq 1,\ k \geq 0.$$

- $A3$ The autocorrelation function of lag k for S–D pairs possesses the form:

$$C[X_i^t,\ X_i^{t+k}] = \gamma_{Xi}(k) = \sigma_{Xi}^2\, \rho_X(k),\ i = 1, \ldots, N;\ t \geq 1,\ k \geq 0$$
$$\rho_X(k) = \frac{1}{2}\left\{(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}\right\},\ i = 1, \ldots, N;\ k \geq 0, \tag{2.2}$$

where $1/2 \leq H < 1$ is the Hurst parameter. The basic dichotomy is between $H = 1/2$ (short range dependence) and $H > 1/2$ (long range dependence). Equation (2.2) contains an important assumption: all S–D pairs have the same value of the Hurst parameter H. This *homogeneity assumption* will be carefully discussed in the sequel.

The spectral function at a single S–D pair level is unbounded at the origin, and takes the form:

$$f_{X_i}(\omega;\ H) = \sigma_{Xi}^2\ f_X(\omega;\ H)$$
$$= \sigma_{Xi}^2\ \frac{1}{\pi}\ \sin(\pi H)\,\Gamma(2H+1)\,(1-\cos\omega)C_0(H,\ \omega),\ i = 1, \ldots, N, \tag{2.3}$$

where

$$C_l(H,\ \omega) = \sum_{k=-\infty}^{+\infty} (\log|2\pi k + \omega|)^l |2\pi k + \omega|^{-(2H+1)},\quad l = 0,\ \pm1,\ \pm2,\ \ldots \tag{2.4}$$

When ω is close to zero, the following well-known approximation holds:

$$f_{X_i}(\omega;\ H) \approx \sigma_{Xi}^2 \frac{1}{2\pi} \sin(\pi H)\,\Gamma(2H+1)\,|\omega|^{1-2H} \ \text{ as } \omega \to 0,\quad i = 1,\ \ldots,\ N. \tag{2.5}$$

The stochastic process (X_i^t) admits the following backward expansion:

$$X_i^t - \mu_{Xi} = \sigma_{Xi} \sum_{\tau=0}^{\infty} c_\tau(H)\, u_i^{t-\tau}, \tag{2.6}$$

where $(u_i^t;\ t \geq 1)$ are *i.i.d.* standard normal random variates.

For the sake of simplicity, from now on we will use the following notation:

$$\begin{aligned} u^t &= (u_1^t,\ \ldots,\ u_N^t) \\ \mu_X &= (\mu_{X1},\ \ldots,\ \mu_{XN}) \\ \theta_i &= \sigma_{Xi}^2,\ i = 1,\ \ldots,\ N, \\ \theta &= (\sigma_{X1}^2,\ \ldots,\ \sigma_{XN}^2). \end{aligned}$$

Due to the independence among S–D pairs, the covariance matrix of the r.v. X^t is equal to

$$\Gamma_X(k;\ \theta,\ H) = \rho_X(k) \begin{bmatrix} \theta_1 & 0 & \cdots & 0 \\ 0 & \theta_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \theta_N \end{bmatrix} = \rho_X(k)\, \Sigma_X(\theta).$$

From equation (2.6) we immediately obtain the following expansion for multidimensional process $(X^t;\ t \geq 1)$:

$$X^t - \mu_X = \Sigma_X(\theta)^{1/2} \sum_{\tau=0}^{\infty} c_\tau(H)\, u^{t-\tau}, \tag{2.7}$$

where $(u^t;\ t \geq 1)$ are *i.i.d.* multinormal random vectors with zero mean vector and covariance matrix I_N, the $N \times N$ identity matrix.

Denote now by

$$Y^t = (Y_1^t,\ \ldots,\ Y_M^t)$$

the (column) vector containing the traffic for the M links. Denoting by $A = (a_{kl})$ the $M \times N$ matrix with a_{kl} equal either to 1 or to 0 according to whether link k does or does not belong to the directed path of the S–D pair l, the following relationship

$$Y^t = AX^t,\quad t = 1,\ 2,\ \ldots \tag{2.8}$$

holds.

In view of equation (2.8) and assumptions A1–A3, it is obvious that the (multivariate) process $(Y^t;\ t \geq 1)$ is a stationary Gaussian process, with mean function $\mu_Y = A\mu_X$ and covariance matrix of lag k $\Gamma_Y(k;\ \theta,\ H) = A\,\Gamma_X(k;\ \theta,\ H)\,A'$. More explicitly, the covariance matrix of lag k, $\Gamma_Y(k;\ \theta,\ H)$, possesses the following structure:

$$\Gamma_Y(k;\ \theta,\ H) = \rho_X(k)G(\theta), \tag{2.9}$$

where $G(\theta)$ is the $M \times M$ matrix

$$G(\theta) = A\Sigma_X(\theta)A'. \tag{2.10}$$

From equation (2.9) it appears that the (univariate) stochastic processes $(Y_m^t;\ t \geq 1)$, $m = 1, \ldots, M$, are still stationary Gaussian processes, with Hurst parameter H. Furthermore, the cross-covariance between Y_j^t and Y_l^{t+k} is long memory, too.

The cross-spectrum matrix of the process $(Y^t;\ t \geq 1)$ takes the following form:

$$\Phi_Y(\omega;\ \theta,\ H) = f_X(\omega;\ H)\,G(\theta), \tag{2.11}$$

where $f_X(\omega;\ H)$ and $G(\theta)$ are given by equations (2.3) and (2.10), respectively.

Relationship (2.11) shows that the cross-spectrum matrix (2.11) factorizes into the product of two terms: a scalar only depending on H and a $M \times M$ matrix only depending on θ.

From equation (2.7) a backward representation formula for the process $(Y^t;\ t \geq\ 1)$ is obtained. First of all, using equations (2.8) and (2.7) we may write

$$Y^t - A\mu_x = \sum_{\tau=0}^{\infty} c_\tau(H)\ A\Sigma_X(\theta)^{1/2}\epsilon^{t-\tau}, \tag{2.12}$$

where $A\Sigma_X(\theta)^{1/2}e^{t-\tau}$ are *i.i.d.* M-variate multinormal random vectors with zero mean vector and covariance matrix $G(\theta)$. In the second place, if $Q(\theta)$ denotes the $M \times M$ orthogonal matrix whose column are the normalized eigenvectors of $G(\theta)$, and $\Lambda(\theta)$ is the diagonal matrix composed by the eigenvalues of $G(\theta)$, taking $G(\theta)^{1/2} = Q(\theta)\,\Lambda(\theta)^{1/2}$ and using the spectral decomposition $G(\theta) = G(\theta)^{1/2}\,(G(\theta)^{1/2})'$, it is immediate to see that the relationship

$$A\Sigma_X(\theta)^{1/2}\epsilon^t = G(\theta)^{1/2}\epsilon^t,\ t \geq 1$$

holds, where $(\epsilon^t;\ t \geq 1)$ are *i.i.d.* multinormal random vectors with zero mean vector and covariance matrix I_M. As a consequence, the process $(Y^t;\ t \geq 1)$ possesses the following backward representation

$$Y^t - A\mu_X = G(\theta)^{1/2} \sum_{\tau=0}^{\infty} c_\tau(H)\,\epsilon^{t-\tau}. \tag{2.13}$$

Since M is usually smaller than N, the model introduced so far is not identifiable. The most common approach consists in using a mean–variance relationship, such as

$\mu_{Xi} = \theta_i$, o, more generally, $\mu_{Xi} = const \times \theta_i^q$, $q > 0$. In the sequel, we will assume that

$$\mu_{Xi} = h(\theta_i),\ i = 1, \ldots, N,$$

where $h(\cdot)$ is a strictly monotone, known function. In the sequel, the notation

$$h(\theta) = [h(\theta_1), \cdots, h(\theta_N)] \tag{2.14}$$

will be used.

2.2 Discussion and justification of the assumptions

In the sequel, the assumptions on which our model rests are discussed in detail.

Independence between S–D pairs. The hypothesis of independence between S–D pairs has been studied for real data in Susitaival *et al.* (2006), via the correlation between the standardized residuals of the bits (packets) arrival process (bit/packet network traffic) of two different S–D pairs at various time aggregation levels. The data at hand are measurements observed on one link in the Finnish University Network (Funet), and partitioned into S–D traffic on the basis of source and destination IP address; time aggregation varies from 1 second to 300 seconds. The main conclusion is that there is no particular evidence against the assumption of independence among S–D pairs.

Gaussianity. The assumption of Gaussianity of the S–D bits (packets) arrival process is justified by (space/time) aggregation obtained by superimposing independent traffic processes (independent sources) satisfying the usual conditions of the functional central limit theorem (see Norros and Pruthi, 1996; Taqqu *et al.*, 1997). The assumption has been considered in Cao *et al.* (2000), Norros and Kilpi (2002), Juva *et al.* (2005), Susitaival *et al.* (2006) and Juva *et al.* (2007), and validated via QQ-plots and related correlation tests that compare the empirical distribution with a fitted Gaussian distribution. The data at hand is traffic observed on one link partioned into S–D traffic based on source and destination IP address (time aggregation from 1 second to 300 seconds), except in Cao *et al.* (2000) where data is traffic observed on the links and S–D pairs of a one-router network with 5 minutes time aggregation, as provided by Simple Network Management Protocol (SNMP). The data are consistent with a normal based modelling approach under suitable (space/time) aggregation.

Stationarity. The assumption of stationarity is studied in Cao *et al.* (2000) and Norros and Kilpi (2002). In Cao *et al.* (2000) the time-varying nature of network traffic is visually observed in Lucent data (traffic observed on the links and S–D pairs of a one-router network with 5 minutes time aggregation as provided by SNMP). Stationarity of S–D traffic (with respect to mean and variance) is assumed to hold in a window lasting up to 21 five minutes time intervals. To take into account the time-varying nature of S–D traffic, Cao *et al.* (2000) propose to estimate S–D traffic (via *EM*) using a local *i.i.d.* model within a moving data window. In Norros and Kilpi (2002) the assumption of stationarity is considered via the correlation coefficient between packet network traffic in adjacent time periods (time aggregation from milliseconds to seconds). The main conclusion is that stationarity can be reasonably assumed to hold within a period of 30–90 minutes. The empirical studies performed in Juva *et al.* (2005), Susitaival *et al.* (2006) and Juva *et al.* (2007) for Funet data essentially confirm the stationarity assumption, with a time aggregation from 1 second to 300 seconds.

As far as the constancy of H over time is concerned (with reference to time scales relevant for traffic matrix estimation in a Wide Area Network [WAN] environment) it has been studied in Norros and Kilpi (2002). The main conclusion is that the stationarity of H can be reasonably assumed to hold within periods of 30–90 minutes.

Long range dependence. The assumption of LRD among traffic packets is more delicate. Since the seminal paper by Leland *et al.* (1994), a number of studies of measurements of packet networks have shown that the arrival process of (bits) packets is self-similar, with increments exhibiting a strong temporal correlation (LRD or long memory). Theoretical analysis is in Taqqu *et al.* (1997), where a functional central limit theorem for aggregated traffic is proved. It provides some fundamental theoretical motivations to use Gaussianity, long range dependence and homogeneity of H, as well. In more detail, Theorems 1 and 2 in Taqqu *et al.* (1997) can be interpreted as follows.

1. The superposition of independent ON/OFF sources with heavy-tailed ON and/or OFF periods produces a limiting Gaussian self-similar aggregate cumulative packets arrival process.
2. Under appropriate assumptions, the expected traffic level provides the main term of the observed traffic; fluctuations around expected traffic approximately behave like a rescaled fractional Brownian motion.
3. If a finite number of independent heterogeneous sources, possibly with different values of the Hurst parameter H, are superimposed, then the term with the highest value of H tends to be dominant. Equivalently, the application with the highest value of the Hurst parameter determines the value of the Hurst parameter for the aggregate traffic. This point is also raised in Fonseca *et al.* (2000).

Further theoretical results, again supporting the presence of LRD, are in Resnick and Samorodnitsky (2001), where infinite source Poisson models are considered.

Due to the difficulty to identify the tail of distributions from limited data, Gong *et al.* (2005) propose a generative model for network traffic (MHOP Markovian Hierarchical ON-OFF Process) that produces long memory traffic without relying on heavy tails. Self-similarity of the simulated traffic fits (by means of different statistical techniques) self-similarity of Bellcore real data.

In order to explore self-similarity of the aggregate cumulative arrival process, some studies have analyzed traffic at the level of individual source (or S–D pair) in Local Area Networks (host-to-host) to validate the assumption that ON and/or OFF periods have a heavy-tailed distribution (Paxson and Floyd, 1994; Willinger *et al.*, 1995, 1998; Park and Willinger, 2000). Similar studies have been performed for WANs, to validate the assumption that session durations possess a heavy-tailed probability distribution (Crovella and Bestavros, 1997; Park and Willinger, 2000).

Other studies have analyzed directly the cumulative traffic arrival process in WANs, to validate the assumption that the corresponding (traffic) increment process is Gaussian long range dependence (through several different estimates of the long memory parameter; see Park and Willinger (2000), Fay *et al.* (2008)). The relevant time scales for these works is of the order of a few seconds or less. The autocorrelation of the cumulative aggregate (bits/packets) arrival process at the level of S–D pair has been recently analyzed at a time scale either relevant for traffic matrix estimation or suitable to detect long memory (time aggregation varying from 1 second to 300 seconds). In particular, the presence of long range dependence of traffic data is shown in Juva *et al.* (2005) and Susitaival *et al.* (2006) through the analysis of the correlation between the standardized residuals of the bits (packets) arrival process (bit/packet network traffic) at lag k of a S–D pair at various time aggregation levels, as well as in Park *et al.* (2005a), again with a time aggregation from 10 milliseconds to 60 seconds, through wavelet analysis (as developed in Veitch and Abry, 1999). In Norros and Kilpi (2002) long range dependence is detected through a visual analysis. All the above-mentioned papers show that S–D traffic is characterized by the presence of long range dependence (with values of the Hurst parameter H ranging between 0.65 and 0.9 for aggregation time 1–300 seconds).

Homogeneity. The main theoretical justification of the homogeneity hypothesis (all S–D pair possess the same value of H) is in Taqqu *et al.* (1997) (see the third point). Empirical analyses supporting the homogeneity assumption are in Park *et al.* (2005b) and Susitaival *et al.* (2006).

From a practical point of view, the homogeneity assumption is based on a simple idea: applications run by customers connected to a given node do not significantly differ from those used by the customers of different nodes. In other words, nodes (representing routers) are not specialized by service. Although we do not know of any thorough work devoted to the relation between the Hurst parameter and the network application, it is known in practice that the feedback behaviour built in the Transmission Control Protocol (TCP) transport protocol (and hence in all the network application relying on it) is a source of long range dependence. More recently, it has been observed that the appearance of a particular P2P protocol (namely Blubster, a.k.a. Piolet) has been related to a variation of the measured Hurst parameter (Park *et al.*, 2005a). Hence, the Hurst parameter can be reasonably considered as determined by the mix of applications generating the observed traffic. Therefore, there should be no difference among nodes roughly generating similar application mix.

Mean–variance relationship. Finally, as far as the mean–variance relation is concerned, the simplest (and the most widely used as well) assumption is that $\mu_{Xi} = h(\theta_i) = const \times \theta_i^q$, i.e., a power law. In the seminal paper by Vardi (1996), the values $const = 1$, $q = 1$ are assumed, whilst in Cao *et al.* (2000) the value $q = 1/2$ is also considered. With reference to S–D data on time scales relevant for traffic matrix estimation, the mean–variance relation has been studied in Susitaival *et al.* (2006) and Juva *et al.* (2007) under different space/time aggregation levels for Funet data. The main conclusion is that a power law $\mu_{Xi} = const \times \theta_i^q$ can be assumed, with $q = 2/3$.

2.3 Identifiability issues

Since the process $(Y^t;\ t \geq 1)$ is Gaussian, its probability law is specified by both its mean function and its covariance kernel (Karatzas and Shreve, 1991). As a consequence, the whole process $(Y^t;\ t \geq 1)$ is identifiable if and only if its two-dimensional distributions are, i.e., if and only if for every $t \geq 1$ and $k \geq 1$ the probability law of $(Y^t,\ Y^{t+k})$ is identifiable. The following result, which is essentially an adaptation of Theorem 1 in Cao *et al.* (2000), holds.

Proposition 1 *Let B be the* $(M(M+1)/2) \times N$ *matrix whose rows are the rows of A and the component-by-component products of all distinct pairs of rows of A. Under the mean–variance relationship summarized in* (14) $(Y^y;\ t \geq 1)$ *is identifiable if and only if B has full column rank.*

Proof See Appendix A.

The identifiability condition of Proposition 1 is expressed in terms of linear algebra. Next result, as expressed in Proposition 2, gives a simpler sufficient condition in terms of graph theory. Its essential merit is that it allows one to immediately establish the identifiability of the involved model, without any manipulation of the matrix B.

Proposition 2 is in the same spirit as the Corollary in Cao *et al.* (2000), where a sufficient condition is provided for the identifiability of the traffic matrix. Such a condition is that the only traffic flowing on the link outgoing from a node a and on the link incoming on a node b is the traffic pertaining to the OD couple $(a,\ b)$. This condition is tantamount to stating that all the nodes in the network act just as ‘end nodes’, and do not support transit traffic. Now, this is quite rare to be met in real networks: the only examples we can think of are pure hub networks (such as the network considered in Cao *et al.* (2000)) and fully meshed networks where routing is accomplished just by the direct one-link paths. Both represent quite special networks. From this point of view, Proposition 2 can be considered as a generalization of the Corollary in Cao *et al.* (2000). To state more transparently the main result, we need a slightly different terminology and notation. A telecommunication network can be thought of as a graph, where source/destination sites are nodes, and direct links are arcs. A sequence of arcs connecting two nodes is a path. Clearly, a path is a sequence of consecutive arcs connecting two nodes. The routing matrix A possesses elements a_{ij} equal to 1 if the arc j is a part of the path connecting the pair i of nodes, and equal to 0 otherwise.

The *length* of a path is the number of arcs that define the path. The simplest algorithm to construct S–D paths consists in using the *minimum length* rule: paths are composed by the smallest number of arcs.

A sub-path of a given path is a sub-sequence of the arcs of the path. Of course, a sub-path is still a path, connecting two nodes. More formally, and with a different notation, let us indicate nodes by letters, and arcs by pairs of letters, those of the two nodes connected by arcs. Suppose that two nodes a and b are connected through the path composed by the l arcs $(a, a_1), (a_1, a_2), \dots, (a_{l-1}, b)$. Then such a path also contains all paths from a node a_i to a node a_j, $0 \leq i < j \leq l$, with $a_0 = a$, $a_l = b$. Such paths will be called *sub-paths* of that connecting the nodes a and b. Clearly, every sub-path connects two nodes. In the sequel, we will assume that the paths in the routing matrix possess the following property.

$P1$ For every pair a_i, a_j, $i < j$, of nodes, the sub-path $(a_i, a_{i+1}), \dots, (a_{j-1}, a_j)$ is also the path connecting the Source-node a_i to the Destination-node a_j, as it appears in the routing matrix A.

Proposition 2 *Assume that all S–D pairs can be connected through a path, and that $P1$ holds. Then the matrix B does have rank N.*

Proof See Appendix A.

Remark 1 Property $P1$ is satisfied when paths are obtained via the minimum length rule. This is true, in particular, for two widely deployed Internet routing algorithms, namely Routing Information Protocol (RIP) and Open Shortest Path First (OSPF), which represent two alternative ways to provide the shortest path: see, for instance, Huitema (1999). In RIP, by construction each node obtains the optimal path to a destination by passing through its adjacent node providing the shortest path to that destination. In OSPF all the transit nodes are chosen to provide the shortest path to the final destination.

3 Statistical analysis: basic aspects

In this section we mainly concentrate on the estimation of the parameters θ, H, μ_X. The statistical data at hand are a multivariate time series $Y^1, Y^2, \dots, Y^T$, each Y^t being a M-dimensional vector of components $Y_1^t, \dots, Y_M^t$.

3.1 General theoretical results

A general framework for the asymptotic theory of parameter estimates in case of linear, scalar-valued stationary processes with long range dependence is in Whittle (1952), Yajima (1985), Fox and Taqqu (1986) and Dahlhaus (1997) for Gaussian processes, and in Giraitis and Surgailis (1990) for non-Gaussian processes. Whittle's results were extended to vector-valued processes by Hosoya and Taniguchi (1982) and Hosoya (1997). In the sequel, we give a short summary of Hosoya's results. Let $(Z(t);\ t \geq 0)$ be a (discrete-time) zero mean vector-valued linear process:

$$Z(t) = \sum_{l=0}^{\infty} G(l;\ \psi)\, e(t-l), \quad t \geq 0,$$

where $Z(t)$s are q-dimensional random vectors, and $e(t)$s are p-dimensional random vectors such that

$$E[e(u)\, e(v)^*] = \delta(u,\ v)\, K(\psi),$$

K being a nonsingular $p \times p$ matrix. $G(l;\ \psi)$s are $q \times q$ matrices, $\delta(\cdot,\ \cdot)$ is the Kronecker delta, and the components of Z, E and G are all real. A^* denotes as usual the conjugate transpose of a matrix A, and the same notation is retained for the transpose of a real A; $tr(A)$ and $det(A)$ are the trace and the determinant of A, respectively. ψ is a s-dimensional vector of unknown parameters. Under the condition

$$\sum_{j=0}^{\infty} tr(G(l;\ \psi)\, K(\psi)\, G(l;\ \psi)^*) < \infty, \tag{3.1}$$

the process $(Z(t);\ t \geq 0)$ is a second order stationary process, with spectral density matrix

$$g(\omega;\ \psi) = \frac{1}{2\pi}\, k(\omega;\ \psi)\, K(\psi)\, k(\omega;\ \psi)^*, \quad -\pi < \omega \leq \pi \tag{3.2}$$

and

$$k(\omega;\ \psi) = \sum_{l=0}^{\infty} G(l;\ \psi)\, e^{i\omega l}.$$

Assume further that the process $(Z(t);\ t \geq 0)$ is Gaussian, and choose frequencies ω_js $(j = 1,\ \ldots,\ T)$ equally spaced on the torus $(-\pi,\ \pi]$, i.e., $\omega_j = (2\pi j)/T - \pi$. The finite Fourier transforms

$$w_T(\omega_j) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} Z(t)\, e^{it\omega_j}, \quad j = 1,\ \ldots,\ T$$

based on a finite segment $Z(1), \ldots, Z(T)$ are sufficient statistics for ψ. The random vectors $w_T(\omega_j)$s are approximately independent for large T, with complex-valued multivariate normal distribution. Their density functions are proportional to

$$(det(g(\omega_j;\ \psi))^{-1/2} \exp\left\{-\frac{1}{2} tr\left(g^{-1}(\omega_j;\ \psi)\, w_T(\omega_j)\, w_T(\omega_j)^*\right)\right\},\quad j = 1, \ldots, T.$$

The sufficiency of $w_T(\omega_1), \ldots, w_T(\omega_T)$ implies that an approximate log-likelihood for ψ, up to a multiplicative factor, is

$$\bar{l}_T(\psi) = -\sum_{i=1}^{T} \left\{\log\det(g(\omega_j;\ \psi)) + tr(g^{-1}(\omega_j;\ \psi)\, I_T(z,\ \omega_j)\right\}, \tag{3.3}$$

where $I_T(z, \omega_j)$ is the periodogram matrix, defined by $w_T(\omega_j)\, w_T(\omega_j)^*$. The approximate log-likelihood $\bar{l}_T(\psi)$ is constructed in Hosoya (1997). It is essentially the multivariate version of the approximate log-likelihood for scalar-valued processes first considered by Whittle (1952).

Assume now that $\log\det(g(\omega;\ \psi))$ and $tr\left(g^{-1}(\omega_j;\ \psi)\, w_T(\omega)\, w_T(\omega)^*\right)$ are differentiable w.r.t. ψ, and denote by $\partial\bar{l}_T(\psi)/\partial\psi$ the vector of partial derivatives of $\bar{l}_T(\psi)$ w.r.t. ψ. A root $\widehat{\psi}_T$ of the equations $\partial\bar{l}_T(\psi)/\partial\psi = 0$ is a quasi maximum likelihood (QML) estimate of ψ. Under appropriate regularity conditions, consistency and asymptotic normality of $\widehat{\psi}_T$ are proved in Hosoya (1997). The asymptotic covariance matrix of $\widehat{\psi}_T$ has a complicated form, involving the second partial derivatives of $\log\det(g(\omega;\ \psi))$ and $tr\left(g^{-1}(\omega_j;\ \psi)\, w_T(\omega)\, w_T(\omega)^*\right)$ (Hosoya, 1997, Theorem 2.2).

3.2 Statistical analysis for telecommunication networks

The results in the previous sub-section are used here to construct point estimates for the parameters of the model developed in Section 2.

In our problem, $Y^t - A\mu_X = Y^t - Ah(\theta)$ essentially plays the role of $Z(t)$ in Sub-section 3.1. The parameter to be estimated on the basis of the approximate log-likelihood (3.3) is $\psi = \eta = (\theta,\ H)$. The cross-spectrum matrix $g(\omega,\ \psi)$ is given by equation (16). Under conditions $A1$–$A3$, it is not difficult to see that Y^t satisfies all conditions of Sub-section 3.1. Hence, the approximate log-likelihood for (θ, H) is given by

$$\bar{l}_T(\theta,\ H) = -\sum_{j=1}^{T} \left\{\log\det(f_X(\omega_j;\ H)\, G(\theta)) + tr((f_X(\omega_j;\ H)\, G(\theta))^{-1})\, I_T(\theta,\ \omega_j))\right\}, \tag{3.4}$$

where $I_T(\theta,\ \omega_j) = w_T(\theta,\ \omega_j)\, w_T(\theta,\ \omega_j)^*$,

$$w_T(\theta,\ \omega_j) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} (Y^t - Ah(\theta))\, e^{it\omega_j}, \quad j = 1,\ \ldots,\ T.$$

Now, it is not difficult to show that the approximate likelihood (3.4) can be considerably simplified, because the term $Ah(\theta)$ essentially disappears. In fact, the term $I_T(\theta,\ \omega_j)$ can be written as

$$\begin{aligned} I_T(\theta,\ \omega_j) &= \frac{1}{2\pi T} \sum_{t=1}^{T} \sum_{u=1}^{T} (Y_t - Ah(\theta))(Y_u - Ah(\theta))' e^{it\omega_j} e^{-iu\omega_j} \\ &= \frac{1}{2\pi T} \left\{ \sum_{t=1}^{T} \sum_{u=1}^{T} Y_t Y_u' e^{it\omega_j} e^{-iu\omega_j} - \sum_{t=1}^{T} Y_t (Ah(\theta))' e^{it\omega_j} \sum_{u=1}^{T} e^{-iu\omega_j} \right. \\ &\left. - \sum_{t=1}^{T} Y_t (Ah(\theta))' e^{-it\omega_j} \sum_{u=1}^{T} e^{iu\omega_j} + Ah(\theta)(Ah(\theta))' \sum_{t=1}^{T} e^{it\omega_j} \sum_{u=1}^{T} e^{-iu\omega_j} \right\}, \end{aligned} \tag{3.5}$$

Using the well-known trigonometric relationship

$$\sum_{t=1}^{T} \exp\left\{ i \frac{2\pi}{T} tj \right\} = \begin{cases} T, & j = 0 \\ 0, & j \neq 0 \end{cases}$$

j being an integer such that $|j| < T$, and taking into account that $\omega_j = 2\pi j/T - \pi$, it is not difficult to see that the relationships

$$\sum_{t=1}^{T} e^{-iu\omega_j} = \sum_{t=1}^{T} e^{iu\omega_j} = 0, \quad j = 1,\ \ldots,\ T-1 \tag{3.6}$$

hold true. In view of equations (3.5) and (3.6), it is easy to see that

$$\begin{aligned} I_T(\theta,\ \omega_j) &= \frac{1}{2\pi T} \sum_{t=1}^{T} \sum_{u=1}^{T} Y_t Y_u' e^{it\omega_j} e^{-iu\omega_j} \\ &= \widehat{I}_T(\omega_j); \quad j = 1,\ \ldots,\ T-1. \end{aligned}$$

Since the term corresponding to $j = T$ is essentially negligible for large T, the approximate log-likelihood (3.4) can be written in the form:

$$\bar{l}_T(\theta,\ H) = -\sum_{j=1}^{T} \left\{ \log \det(f_X(\omega_j;\ H)\, G(\theta)) \right.$$
$$\left. + tr((f_X(\omega_j;\ H)\, G(\theta))^{-1})\, \widehat{I}_T(y,\ \omega_j)) \right\}. \tag{3.7}$$

Note that the 'empirical periodogram' $\widehat{I}_T(\omega)$ can be written as

$$\widehat{I}_T(\omega) = \widehat{w}_T(\omega)\, \widehat{w}_T(\omega)^*,$$

where

$$\widehat{w}_T(\omega) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} Y^t\, e^{it\omega}.$$

We have now to maximize equation (3.7) w.r.t. H and θ. To this purpose, we first compute the corresponding partial derivatives. They are computed according to the conventions and rules in Magnus and Neudecker (1999) and Magnus and Neudecker (1985).

First of all, it is easy to see the equality

$$\frac{\partial}{\partial H} \left\{ \sum_{j=1}^{T} \log \det(f_X(\omega_j;\ H)\, G(\theta)) \right\}$$
$$= \sum_{j=1}^{T} \frac{\partial}{\partial H} \left\{ \log(f_X(\omega_j;\ H)^M \det(G(\theta)) \right\}$$
$$= \sum_{j=1}^{T} \frac{d}{d H} \left\{ M \log f_X(\omega_j;\ H) \right\}$$
$$= M \sum_{j=1}^{T} \frac{1}{f_X(\omega_j;\ H)} \frac{1 - \cos\omega_j}{\pi} \left\{ \frac{d}{dH} (\sin(\pi H)\, \Gamma(2H+1)) \sum_{k=-\infty}^{+\infty} |2k\pi + \omega_j|^{-(2H+1)} \right.$$
$$\left. + (\sin(\pi H)\, \Gamma(2H+1)) \sum_{k=-\infty}^{+\infty} \frac{d}{dH} \left(|2k\pi + \omega_j|^{-(2H+1)} \right) \right\}$$
$$= M \sum_{j=1}^{T} \frac{1}{f_X(\omega_j;\ H)} \frac{1 - \cos\omega_j}{\pi} \left\{ (\pi\ \cos(\pi H)\, \Gamma(2H+1) + 2\ \sin(\pi H)\, \Gamma'(2H+1)) \right.$$
$$\left. \times C_0(H,\ \omega_j) - 2\sin(\pi H)\, \Gamma(2H+1) C_1(H,\ \omega_j) \right\}, \tag{3.8}$$

where $\Gamma'(x) = d\Gamma(x)/dx$ and $C_l(H,\ \omega)$ is given by equation (2.4).

In the second place, the derivative w.r.t. θ of the log det term appearing in equation (3.7) is

$$\frac{\partial}{\partial \theta}\left\{\sum_{j=1}^{T} \log\det(f_X(\omega_j;\ H)\,G(\theta))\right\} = \sum_{j=1}^{T} \frac{\partial\ \log\det G(\theta)}{\partial \theta}$$
$$= T\,(\det G(\theta))^{-1}\frac{\partial\ \det G(\theta)}{\partial \theta}$$
$$= T\,(vec((G(\theta)^{-1})')'\,\frac{\partial\ G(\theta)}{\partial \theta}, \qquad (3.9)$$

where $\frac{\partial\, G(\theta)}{\partial\, \theta}$ is a $M^2 \times N$ matrix having the following structure:

$$\left[\left(\frac{d\,vec(G(\theta))}{d\,\sigma^2_{X1}}\right)\cdots\left(\frac{d\,vecG(\theta)}{d\,\sigma^2_{XN}}\right)\right] \qquad (3.10)$$

and

$$\frac{d\,vecG(\theta)}{d\,\sigma^2_{Xi}} = \boldsymbol{a}_i = vec\begin{bmatrix} a^2_{1i} & a_{1i}a_{2i} & \cdots & a_{1i}a_{Mi} \\ \cdots & \cdots & \cdots & \cdots \\ a_{Mi}a_{1i} & a_{Mi}a_{2i} & \cdots & a^2_{Mi} \end{bmatrix};\ \ i = 1,\ \ldots,\ N. \qquad (3.11)$$

Next steps consists in computing the partial derivatives of the *tr* term in equation (3.7). We have first

$$\frac{\partial}{\partial H}\left\{\sum_{j=1}^{T} tr((f_X(\omega_j;\ H)\,G(\theta))^{-1})\,I_T(y,\ \omega_j))\right\}$$
$$= \sum_{j=1}^{T} tr(G(\theta)^{-1}\widehat{I}_T(\omega_j))\,\frac{d}{d\,H}\,f_X(\omega_j;\ H)^{-1}. \qquad (3.12)$$

Since

$$\frac{d}{d\,H}\,f_X(\omega_j;\ H)^{-1} = -\frac{1}{f_X(\omega_j;\ H)^2}\frac{d}{d\,H}\,f_X(\omega_j;\ H)$$
$$= -\frac{1}{f_X(\omega_j;\ H)^2}\,\frac{1-\cos\omega_j}{\pi}$$
$$\times\{(\pi\ \cos(\pi H)\,\Gamma(2H+1) + 2\ \sin(\pi H)\,\Gamma'(2H+1))$$
$$\times\, C_0(H,\ \omega_j) - 2\sin(\pi H)\,\Gamma(2H+1)C_1(H,\ \omega_j)\} \qquad (3.13)$$

from equations (3.12) and (3.13) we easily obtain

$$\frac{\partial}{\partial H}\left\{\sum_{j=1}^{T} tr((f_X(\omega_j;\ H)\,G(\theta))^{-1})\,\widehat{I}_T(\omega_j))\right\}$$
$$= -\sum_{j=1}^{T} tr(G(\theta)^{-1}\widehat{I}_T(\omega_j))\,\frac{1}{f_X(\omega_j;\ H)^2}\,\frac{1-\cos\omega_j}{\pi}$$
$$\times\{(\pi\ \cos(\pi H)\,\Gamma(2H+1)+2\ \sin(\pi H)\,\Gamma'(2H+1))C_0(H,\ \omega_j)$$
$$-2\sin(\pi H)\,\Gamma(2H+1)C_1(H,\ \omega_j)\}\,. \tag{3.14}$$

Finally, as far as the partial derivatives w.r.t. θ of the *tr* term are concerned, using the foregoing notation we first have

$$\frac{\partial}{\partial\theta}\left\{\sum_{j=1}^{T} tr((f_X(\omega_j;\ H)\,G(\theta))^{-1})\,\widehat{I}_T(\omega_j))\right\}$$
$$= \sum_{j=1}^{T} f_X(\omega_j;\ H)^{-1}\,(vec(\widehat{I}_T(\omega_j)))'\left(\frac{\partial\, G(\theta)^{-1}}{\partial\,\theta}\right), \tag{3.15}$$

where, as shown in Magnus and Neudecker (1999, p. 208, Table 7),

$$\frac{\partial\, G(\theta)^{-1}}{\partial\,\theta} = -((G(\theta)')^{-1}\otimes G(\theta)^{-1})\,\frac{\partial\, G(\theta)}{\partial\,\theta} \tag{3.16}$$

and $\partial\, G(\theta)/\partial\,\theta$ is given by equations (3.10) and (3.11).

If $\widehat{H}$, $\widehat{\theta}$ denote the values of H, θ maximizing $\bar{l}_T(\theta,\ H)$ (3.7), the mean vector μ_X is then estimated by inverting the relationship $\mu_X = h(\theta)$, i.e., by taking

$$\widehat{\mu}_{Xi} = h^{-1}(\widehat{\theta}_i),\quad i = 1,\ \ldots,\ N.$$

3.3 Properties of the obtained estimators

The asymptotic properties of the estimators $\widehat{H}$, $\widehat{\theta}$ can be obtained by applying Theorem 2.2 in Hosoya (1997). Denote by H_0, θ_0 the 'true values' of H, θ, respectively, and let $Q_1(H, \theta)$, $Q_2(H, \theta)$, $Q(H, \theta)$ be the functions:

$$\begin{aligned} Q_1(H, \theta) &= \int_{-\pi}^{\pi} \log\det(f_X(\omega, H)\, G(\theta))\, d\omega \\ &= M \int_{-\pi}^{\pi} \log(f_X(\omega, H))\, d\omega + 2\pi \log\det(G(\theta)) \end{aligned} \tag{3.17}$$

$$\begin{aligned} Q_2(H, \theta) &= \int_{-\pi}^{\pi} tr\left(\{ f_X(\omega, H)^{-1}\, G(\theta)^{-1}\}\, \{ f_X(\omega, H_0)\, G(\theta_0)\}\right) d\omega \\ &= tr(G(\theta)^{-1} G(\theta_0)) \int_{-\pi}^{\pi} \frac{f_X(\omega, H_0)}{f_X(\omega, H)}\, d\omega \end{aligned} \tag{3.18}$$

$$Q(H, \theta) = Q_1(H, \theta) + Q_2(H, \theta) \tag{3.19}$$

and by W the $(N+1) \times (N+1)$ matrix having elements

$$w_{kl}(H, \theta) = \frac{\partial^2 Q(H, \theta)}{\partial\theta_k\, \partial\theta_l}, \quad k, l = 1, \ldots, N \tag{3.20}$$

$$w_{N+1,l}(H, \theta) = w_{l,N+1}(H, \theta) = \frac{\partial^2 Q(H, \theta)}{\partial H\, \partial\theta_l}, \quad l = 1, \ldots, N \tag{3.21}$$

$$w_{N+1,N+1}(H, \theta) = \frac{\partial^2 Q(H, \theta)}{\partial H^2}. \tag{3.22}$$

If $x = (x_1 \ldots x_{m^2})$ is a vector of m^2 elements, denote further by $Ma(x)$ the $m \times m$ matrix whose hth row is composed by $x_{m(j-1)+1}, \ldots, x_{mj}$, $j = 1, \ldots, m$. Finally, let $\gamma_k(\omega; H, \theta)$ be the $M \times M$ matrices

$$\begin{aligned} \gamma_k(\omega; H, \theta) &= f_X(\omega, H)^{-1} \left\{ Ma\left(\frac{\partial\, G(\theta)^{-1}}{\partial\, \theta_k}\right)\right\} \\ &= -f_X(\omega, H)^{-1}\, Ma\left((G(\theta)^{-1} \otimes G(\theta)^{-1})\boldsymbol{a}_k\right), \quad k = 1, \ldots, N \end{aligned} \tag{3.23}$$

$$\begin{aligned} \gamma_{N+1}(\omega; H, \theta) &= \left(\frac{d\, f_X(\omega, H)^{-1}}{d\, H}\right) G(\theta)^{-1} \\ &= -\frac{f'_X(\omega, H)}{f_X(\omega, H)^2}\, G(\theta)^{-1} \end{aligned} \tag{3.24}$$

and let $V(H, \theta)$ the $(N+1) \times (N+1)$ matrix with elements

$$v_{kl}(H, \theta) = 4\pi \int_{-\pi}^{\pi} f_X(\omega, H_0)^2 tr\left(\gamma_k(\omega; H, \theta)\, G(\theta_0)\, \gamma_l(\omega; H, \theta)\, G(\theta_0)\right) d\omega. \tag{3.25}$$

Explicit expressions for the terms w_{kl} and γ_k are obtained in Appendix B.

As a consequence of Theorem 2.2 in Hosoya (1997), it is not difficult to see that, as T increases

$$\sqrt{T}\begin{bmatrix}\widehat{\theta}-\theta_0\\ \widehat{H}-H_0\end{bmatrix}\xrightarrow{d}\mathcal{N}_{N+1}(0,\, W(H_0,\,\theta_0)V(H_0,\theta_0)W(H_0,\,\theta_0)'), \tag{3.26}$$

where $\mathcal{N}_p(\mu, D)$ denotes a p-variate normal distribution with mean vector μ and covariance matrix D.

The matrices $V(H_0, \theta_0)$ and $W(H_0, \theta_0)$ can be consistently estimated. In fact, let $\widehat{Q}_1(H, \theta)$ and $\widehat{Q}_2(H, \theta)$ be defined exactly as (3.17) and (3.18), where integrals are replaced by Riemann sums evaluated at Fourier frequencies $\omega_j = 2\pi j/T - \pi$, $j = 1, \ldots, T$, and H_0 and θ_0 are replaced by $\widehat{H}$ and $\widehat{\theta}$, respectively. Moreover, let $\widehat{Q}(H, \theta) = \widehat{Q}_1(H, \theta) + \widehat{Q}_2(H, \theta)$, and let $\widehat{w}_{kl}(H, \theta)$ be defined exactly as in (3.20)–(3.22), with Q replaced by $\widehat{Q}$. Similarly, define $\widehat{v}_{kl}(H, \theta)$ as (3.25), but again with integrals replaced by Riemann sums evaluated at Fourier frequencies ω_j, and with $\widehat{H}$ and $\widehat{\theta}$ in the place of H_0 and θ_0, respectively. Finally, let $\widehat{W}(H, \theta)$, $\widehat{V}(H, \theta)$ the $(N+1) \times (N+1)$ matrices of elements $\widehat{w}_{kl}(H, \theta)$, $\widehat{v}_{kl}(H, \theta)$, respectively.

Since (as it is easily shown)

$$\widehat{W}(\widehat{\theta},\ \widehat{H})\xrightarrow{p} W(H_0,\,\theta_0),\ \ \widehat{V}(\widehat{\theta},\ \widehat{H})\xrightarrow{p} V(H_0,\,\theta_0)\ \ as\ T\to\infty, \tag{3.27}$$

the asymptotic covariance matrix appearing in equation (3.26) can be consistently estimated.

3.4 Additional results on statistical inference on the Hurst parameter

Although the primary goal of the present article is the estimation of the (expected) traffic matrix, an important point, which needs to be developed, consists in testing for the presence of LRD, as well as in constructing a confidence interval for H. Of course, these two problems are strictly related, since in order to test the null hypothesis $H = 0.5$ (LRD is absent) versus $H > 0.5$ (LRD is present), the procedure usually adopted in the literature (Beran, 1994), consists in constructing a confidence interval for H with level $1 - \alpha$. The null hypothesis $H = 0.5$ is rejected whenever the point 0.5 does not lie in the confidence interval.

In principle, the results of the previous section could be used to construct a Wald-type test-statistic, of the form $T(\widehat{H} - H)'\widehat{\sigma}^2(\widehat{H}, \widehat{\theta})^{-1}(\widehat{H} - H)$, where $\sigma^2(H, \theta)$ is the asymptotic variance of $\widehat{H}$, and $\widehat{\sigma}^2(\widehat{H}, \widehat{\theta})$ is its estimated version. The foregoing statistic does possess an asymptotic Chi-square distribution with 1 degree of freedom, as T goes to infinity. This procedure, although intuitive, does have some relevant drawbacks. First of all, the estimation of the (asymptotic) variance of $\widehat{H}$ is computationally heavy. Second, and more importantly, the actual (finite sample) distribution of $\widehat{H}$ is asymmetric, so that the asymptotic approximation (based on the asymptotic normality of $\widehat{H}$) could be inaccurate. For this reason, we resort to a completely different approach based on sub-sampling, as used, for instance, in Hall *et al.* (1998a), Lahiri (2003) and Conti *et al.* (2008).

Let

$$\zeta_T = \sqrt{T}\,(\widehat{H} - H) \tag{3.28}$$

be the 'centred version' of the estimator $\widehat{H}$, based on T observations, and denote by

$$D_T(x) = P(\zeta_T \leq x) \tag{3.29}$$

its distribution function (d.f.). As a consequence of (3.26), $D_T(x)$ converges to a normal $N(0,\, \sigma^2(H,\, \theta))$ d.f. as T goes to infinity. Convergence is uniform w.r.t. x.

Let further $B^t = (Y^t,\, \ldots,\, Y^{t+l-1})$, $t = 1,\, \ldots,\, N$, be a collection of $N = T - l + 1$ overlapping blocks of length l, for some given integer $l = l_T$ ($1 \leq l \leq T$). Finally, let $\widehat{H}_{l,t}$ be the estimator of H based on the data in block B^t.

A 'sub-sample copy' of ζ_T, based on B^t, is given by

$$\widehat{\zeta}_{l,t} = \sqrt{l}\,(\widehat{H}_{l,t} - \widehat{H}). \tag{3.30}$$

The sub-sampling estimator of the distribution function $D_T(x) = P(\zeta_T \leq x)$ of ζ_T, based on the sub-samples B^t's, is the empirical distribution function (e.d.f.) of the $\widehat{\zeta}_{l,t}$'s in equation (3.30), namely,

$$\widehat{D}_T(x) = \frac{1}{N} \sum_{t=1}^{N} I_{(\widehat{\zeta}_{l,t} \leq x)}, \quad x \in I\!R. \tag{3.31}$$

By repeating *verbatim* the reasoning in Proposition 1, and taking into account that the (multivariate) Gaussian process $(Y^t;\ t \geq 1)$ is completely regular (see, for instance, Ibragimov and Rozanov, 1978), it is not difficult to show that

$$\sup_{x \in \mathbb{R}} \left| \widehat{D}_T(x) - D_T(x) \right| \xrightarrow{p} 0, \quad \text{as } T \to \infty. \tag{3.32}$$

As a consequence, the quantiles of $\widehat{D}_T$ are asymptotically equivalent to the quantiles of D_T. Precisely, if $D_T^{-1}(u) = \inf\{x :\ D_T(x) \geq u\}$ and $\widehat{D}_T^{-1}(u) = \inf\{x : \widehat{D}_T(x) \geq u\}$, then

$$\widehat{D}_T^{-1}(u) - D_T^{-1}(u) \xrightarrow{p} 0, \ \ \text{as } T \to \infty,\ \forall\, u \in (0,\, 1). \tag{3.33}$$

Results 3.32 and 3.33 suggest that the d.f. D_T can be approximated by the e.d.f. $\widehat{Q}_T$ obtained through sub-sampling, and that the quantile D_T^{-1} can be approximated by the sample quantile $\widehat{D}_T^{-1}$. Hence, for $0 < \alpha < 1$ the interval

$$\left(\widehat{H} - \frac{1}{\sqrt{T}}\, \widehat{D}_T^{-1}(\alpha/2),\ \widehat{H} + \frac{1}{\sqrt{T}}\, \widehat{D}_T^{-1}(1 - \alpha/2) \right) \tag{3.34}$$

is a confidence interval for H with asymptotic coverage probability $1 - \alpha$. The accuracy of the interval (3.34) will be evaluated in Section 4.2.2.

Finally, in order to evaluate the possible presence of LRD, it is enough to check whether the confidence interval (3.34) contains the point 0.5 (LRD is absent) or does not contain the point 0.5 (LRD is present). Of course, such a test possesses asymptotic size α.

4 Simulation study

4.1 Traffic model and simulation assumptions

In order to evaluate the performance of the traffic estimation technique presented in the previous sections, we consider here a ‘toy network’, already used by Vardi in his seminal paper Vardi (1996). Such a network is reported in Figure 1.

The nodes in Figure 1 are labelled by letters, and the links by numbers. The network consists of 4 nodes (hence 12 S–D pairs) and 7 unidirectional links. The routing matrix (where the S–D pairs listed on the columns are sorted in lexicographic

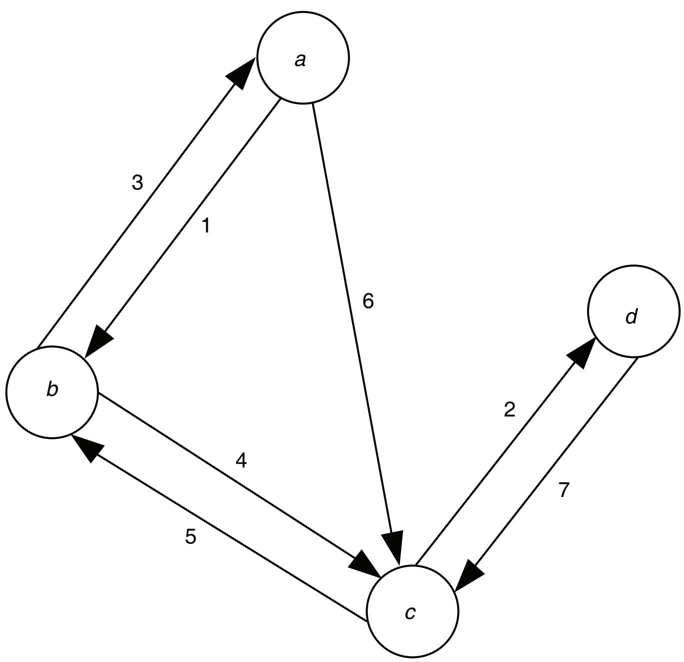

Figure 1 ‘Toy network’ used in the simulation study

order) is reported as in the following.

$$\mathbf{A} = \begin{pmatrix} 1&0&0&0&0&0&0&0&0&0&0&0\\ 0&1&1&0&0&1&0&0&0&0&0&0\\ 0&0&0&1&0&1&1&0&0&1&0&0\\ 0&0&0&0&1&0&0&0&0&0&0&0\\ 0&0&0&0&0&0&1&1&0&1&1&0\\ 0&0&1&0&0&1&0&0&1&0&0&0\\ 0&0&0&0&0&0&0&0&0&1&1&1 \end{pmatrix}$$

For this network, the performance of our traffic estimation technique is evaluated by simulation. In the following sub-sections we review the scenarios employed in the simulation procedure.

In order to simulate the traffic measured on the links and apply our matrix estimation algorithm, we make the following assumptions on the traffic generated.

1. The traffic generated by each node is assumed to follow a long range dependence process, and the autocorrelation functions obey the relationship (2.2).
2. The stochastic processes associated to any two S–D pairs are stochastically independent.
3. All the S–D pairs have the same value of the Hurst parameter.
4. The expected value of traffic follows a Zipf rank–size relationship.
5. The mean–variance relationship for the traffic intensity follows a power law.
6. The coefficient of preference, which determines how the traffic generated by a given origin node distributes among all the destinations, is proportional to the traffic generated by the destination node.

Note that only Assumptions 1–3 are actually used in the proposed estimation method. Assumptions 4–6 (that are well supported in the literature) are used to generate the synthetic traffic data employed to feed the estimation algorithm.

As remarked in the Introduction, Assumptions 1–3 are well supported in the literature. As far as Assumption 4 is concerned, suppose the n nodes are sorted according to the average traffic generated. Denote by μ_{O_j} the expected traffic generated by jth node, so that $\mu_{O_1} \leq \mu_{O_2} \leq \cdots \leq \mu_{O_n}$. Assumption 4 means that μ_{O_j}s obey the Zipf law

$$\mu_{O_j} \propto \frac{1}{i^{\alpha}} \qquad i = 1, \ldots, n. \tag{4.1}$$

This law, originally formulated in the context of linguistics in Zipf (1949), is found in many different contexts to describe rank–frequency relationships. The parameter α determines the imbalance of the traffic distribution: the larger α, the larger the differences in the traffic intensity between the highest and the lowest ranked nodes.

As far as its applications in telecommunications are concerned, Zipf law is supported by measurements conducted on the telephone network and on Internet users (see Naldi and Salaris, 2006). Assumption 5 is expressed by the relationship

$$\sigma^2_{X_i} = \phi\, \mu^c_{X_i} \qquad i = 1, \ldots, N. \tag{4.2}$$

It was put forward in the seminal paper by Cao *et al.* (2000) and is supported by several measurements campaigns: Medina *et al.* (2002), Susitaival *et al.* (2006) and Gunnar *et al.* (2004). Assumption 6 is common in teletraffic studies, e.g., Chapter 13 in the reference book by Bear (1988). Consider the ith S–D pair, and denote by O_l its origin node, and by O_m its destination node. The expected traffic intensity for such S–D pair is then

$$\mu_{X_i} = \mu_{O_l} \frac{\mu_{O_m}}{\sum_{k=1}^{n} \mu_{O_k}} \qquad i = 1, \ldots, N. \tag{4.3}$$

As a consequence of Assumptions 4 and 6, the resulting matrix of expected traffic intensities is asymmetric.

Finally, the long-range-dependent traffic traces are generated by using the Choleski method (see Hall *et al.*, 1998b). This method, though computationally heavy, is exact and has become the reference method for such task (De Giovanni and Naldi, 2008).

In our simulation the following set of parameter values have been used.

- $H = 0.5,\ 0.6,\ 0.8$;
- Zipf parameter $\alpha = 1$;
- Sample size (traffic traces length) $T = 30,\ 90,\ 120$;
- Parameter of the power law relationship between mean and variance $\phi = 1$ and $c = 1,\ 1.5$.

4.2 Simulation results

In this section we provide the results of the simulation-based evaluation. Namely, we report the overall error obtained in estimating the expected traffic intensity for S–D pairs, as well as of the Hurst parameter.

4.2.1 Traffic intensity estimation

In this section, the performance of the proposed estimator(s) of the expected S–D traffic is compared via simulation to the performance of other *statistical* methods proposed in the literature. Among them, techniques based on maximum likelihood have been shown to be the most effective. In detail, the following three traffic matrix estimation methods have been considered:

M1. Maximum likelihood (on Y^t) method based on maximizing the approximated log-likelihood (3.7).

M2. Blind EM maximum likelihood method proposed in Conti *et al.* (2009), and based on EM algorithm.

M3. Maximum likelihood method based on EM method, and that forces H to be equal to 1/2.

The estimation method in Conti *et al.* (2009) is based on a model for X_i^t $t \geq 1$, $i = 1, \ldots, N$ similar to one considered in the present article. The computation of the maximum likelihood estimate of the (expected) traffic matrix is performed via EM algorithm, but no explicit mean–variance relationship is used. Of course, the resulting model could be unidentifiable. The estimation method M3 is essentially the method proposed by Vardi (1996) and Cao *et al.* (2000), and ignores the possible presence of LRD.

For each combination of estimation method, value of T (sample size), value of H and the average error over all the S–D pairs (as in Juva, 2007) are reported in Table 1.

Estimation methods $M1$ and $M2$ outperform method $M3$ when LRD is actually present ($H > 0.5$). When $H = 0.5$, the performance of methods $M1$–$M3$ is comparable. Hence, our first conclusion is that 'taking into account the possible presence of LRD is better than ignoring, even when LRD is absent'.

As far as comparison between methods $M1$ and $M2$ is concerned, we stress that method $M1$ is considerably better than $M2$ from a computational point of view. In a statistical efficiency perspective, the two methods are essentially equivalent when $T = 30$. Method $M1$ is better than $M2$ when $T = 90$, 120. This is probably due to two effects: (i) the larger the T, the better the spectral approximation used by M1; (ii) M1 is based on the 'actual' (although approximated) likelihood function, with no attempts at reconstructing unobserved data.

Table 1 Average error over S–D pairs

	Estimation method								
	*M*1			*M*2			*M*3		
	$H = 0.5$	$H = 0.6$	$H = 0.8$	$H = 0.5$	$H = 0.6$	$H = 0.8$	$H = 0.5$	$H = 0.6$	$H = 0.8$
$T = 30$									
$c = 1.0$	13.7	15.4	15.9	12.6	13.2	15.7	12.6	13.4	16.5
$c = 1.5$	15.4	15.5	16.1	13.9	15.7	21.6	14.2	16.0	22.3
$T = 90$									
$c = 1.0$	11.9	12.2	12.8	12.0	12.4	14.6	11.9	12.6	15.1
$c = 1.5$	12.2	12.6	14.7	12.7	13.6	18.6	12.7	13.8	19.3
$T = 120$									
$c = 1.0$	11.6	11.3	12.4	11.9	12.1	14.5	11.9	12.9	15.0
$c = 1.5$	12.1	11.3	14.5	12.5	13.4	18.0	12.6	13.8	18.0

4.2.2 Hurst parameter estimation

The goal of the present section is to study the quality of both point and interval estimates of H, on the basis of the results obtained in the preceding section.

As far as the point estimation of H is concerned, the bias $B(\widehat{H})$ and the standard deviation $SD(\widehat{H})$ of the estimator $\widehat{H}$ maximizing (3.7) are shown in Table 2. The values reported are computed over a block of 500 simulation runs. We can observe a slight underestimation of the Hurst parameter. However, the bias is progressively reduced when longer traffic traces are used.

In the second place, the results of our simulation study have been used to evaluate the accuracy of confidence intervals (3.34). In particular, we have considered the sample sizes $T = 90,\ 120$, a nominal coverage probability $1 - \alpha = 0.95$, and two values of H, namely $H = 0.6,\ 0.8$. The block length l has been chosen according to the rule $l = 6 \times \sqrt{T}$, which gives good results as reported in Conti *et al.* (2008). In the present case, the value $l = 60$ has been used. The coverage probabilities and the average interval length, computed on the basis of simulated data, are reported in Table 3.

As it may be seen from Table 3, the coverage probability is rather close to its nominal level 0.95, and the average length of sub-sampling confidence interval is smaller than $H/10$.

Table 2 Estimated Hurst parameter

T	c	H	$B(\widehat{H})$	$SD(\widehat{H})$	T	c	H	$B(\widehat{H})$	$SD(\widehat{H})$
30	1	0.6	−0.016	0.014	90	1	0.6	−0.009	0.013
30	1	0.8	−0.008	0.024	90	1	0.8	−0.002	0.024
30	1.5	0.6	−0.009	0.021	90	1.5	0.6	−0.007	0.017
30	1.5	0.8	−0.02	0.028	90	1.5	0.8	−0.005	0.026

Table 3 Coverage probabilities and average length of sub-sampling confidence intervals (nominal level: 0.950)

	$H = 0.6$		$H = 0.8$	
	Coverage prob.	*Average length*	Coverage prob.	Average length
		Sample size $T = 90$		
$c = 1.0$	0.938	0.057	0.931	0.062
$c = 1.5$	0.934	0.071	0.929	0.074
		Sample size $T = 120$		
$c = 1.0$	0.940	0.050	0.933	0.056
$c = 1.5$	0.936	0.059	0.930	0.065

5 Application to real data

In order to evaluate the performance of the traffic estimation model presented in the previous sections, an application to real data has been developed. The primary goal of this analysis is to estimate the traffic matrix, i.e., the expected amounts of traffic μ_{Xi} os S–D pairs. Traffic data have been collected on a network belonging to a wider network of Tinet S.p.A. and Enter s.r.l., two Italian providers. The network considered here is made of 16 links and 9 nodes. Nodes are labelled by Rj, $j = 1, \ldots, 9$, and links by $i = 1, \ldots, 16$. The topology of this real network is shown in Figure 2, where the links 1–8 are directed to the hub router $R4$, while the links 9–16 go in the opposite direction.

For structural reasons related to the network architecture, the only allowed S–D pairs are either those with Source node R_i, $i = 1, \ldots, 3$ and Destination node R_j, $j = 5, \ldots, 9$, or those with Source node R_i and Destination node $R_j, i, j = 5, \ldots, 9$, $i \neq j$. The node R_4 is a transit node that does not generate its own traffic. As a consequence, there are $N = 50$ S–D pairs. The routing matrix A (where the S–D pairs listed on the columns are sorted in lexicographic order) is reported in the following.

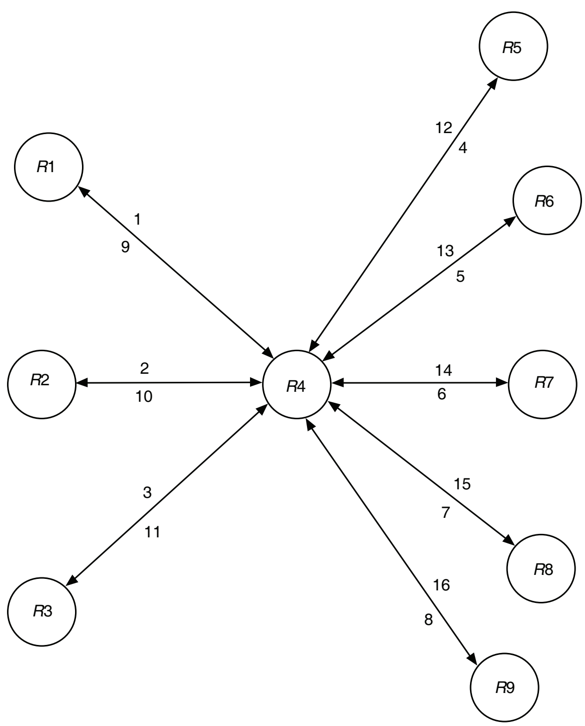

Figure 2 Network used in application

The last row has been deleted, in order to avoid linear dependence (exactly as in Cao *et al.*, 2000).

Links	*S–D pairs*
1	1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2	0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
3	0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
4	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
5	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
6	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
7	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 0 0 0 0 0 0 0
8	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1
10	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0
11	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0
12	1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0
13	0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0
14	0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0
15	0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1
16	0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0

The observed traffic on links consists of bytes per minute. For each link, 90 minutes of observed traffic are available, so that the sample size is $T = 90$. Due to the length of the observation period, the traffic produced by each S–D pair is assumed stationary. In Figure 3 the traffic (bytes per minute) generated in 90 minutes by all 16 links is reported.

In order to validate, via assumption of stationarity, 'empirical versions' of the variograms $\gamma(k) = E[(Y_j^{t+k} - Y_j^t)^2]$ (Beran, 1994) have been constructed for all links $j = 1, \ldots, 16$. They are shown in Figures 4 and 5.

As remarked in Beran (1994, p. 94), the behaviour of the depicted variograms shows the presence of LRD, and also the absence of trends.

In order to validate, again via an explorative analysis, the model assumptions of Gaussianity and long range dependence, the standardized residuals, as well as the corresponding (Gaussian) QQ-plots, have been computed. Since X^ts are not observed, the validation itself is based on Y^ts observations. For each link $j = 1, \ldots, 16$, the vector $[Y_j^1, \cdots, Y_j^{90}]$ has been standardized by taking into account its covariance structure (due to long range dependence). This requires a preliminary estimation of the expectation and of the covariance matrix of Y_j^ts, obtained by taking the sample mean and by estimating H via the Whittle method, respectively. Results support the assumptions of Gaussianity and long range dependence.

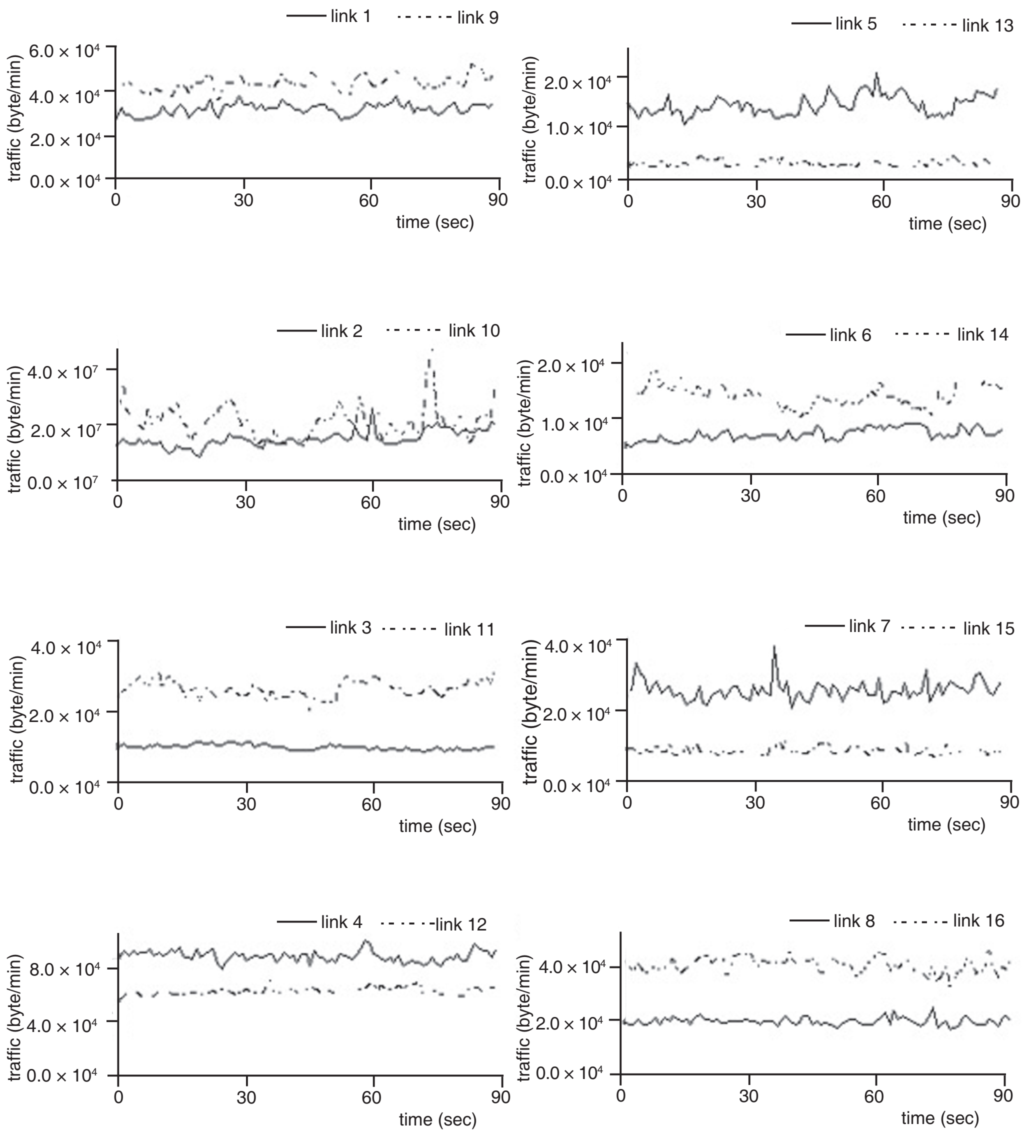

Figure 3 Byte counts per minute on all 16 links (90 minutes)

As already remarked, the proposed model is not identifiable, unless a mean–variance relationship for S–D traffic is introduced. In particular, we consider here a power relation between mean and variance of S–D traffic, i.e.:

$$\mu_{X_i} = \phi^{-1/c}\sigma_{X_i}^{2/c}, \;\; c > 0; \;\; i = 1, \ldots, N. \tag{5.1}$$

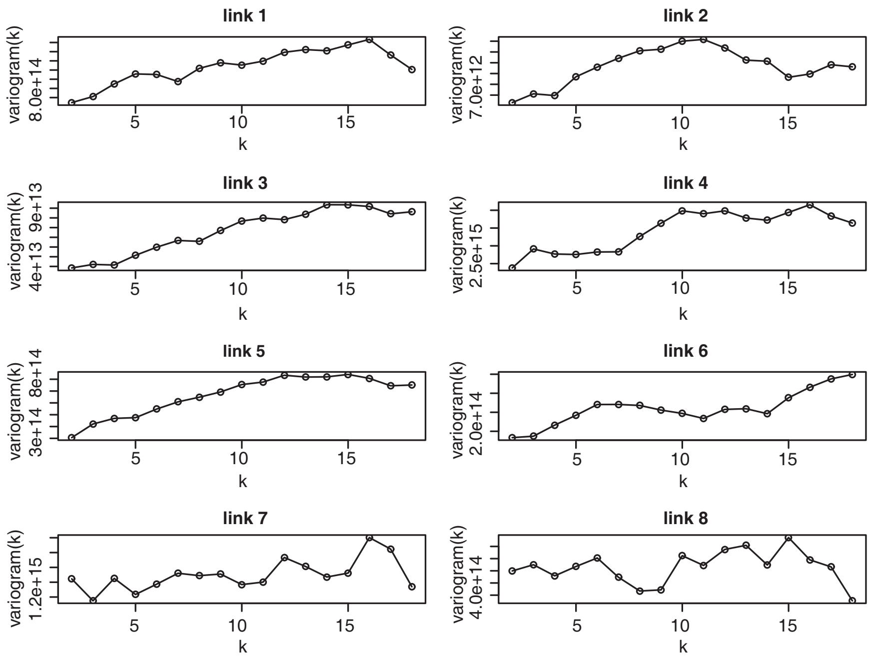

Figure 4 Empirical variograms for links 1–8

In order to get a good degree of flexibility of the proposed model, the quantities ϕ and c in (5.1) are obtained via a preliminary exploratory analysis. In more detail, a linear regression of $\log \widehat{s}_j^2$ against $\log \widehat{m}_j^2$ is considered

$$\log \widehat{s}_j^2 = \log \phi + c \log \widehat{m}_j^2 \tag{5.2}$$

$\widehat{s}_j^2$ and $\widehat{m}_j$ being the sample mean and variance of traffic on links, i.e., of Y_js ($j = 1, \ldots, 16$); see Figure 6. The estimated variance $\log \widehat{s}_j^2$ takes into account the presence of long memory dependence.

According to the arguments in Cao *et al.* (2000), $\log \phi$ is estimated with an error in between zero and $(1 - c) \log K$ where K is the number of S–D pairs that contribute to the byte count on link j. This term is small w.r.t. $c \log \widehat{m}_j^2$.

The obtained values for the parameters are $\log \phi = 9.89$, $c = 1.24$, with an R^2 goodness-of-fit value equal to 0.92, and a corresponding p-value 2.2×10^6. These quantities are used in the relationship (5.1) in order to estimate the mean S–D traffic. The estimated value of H, common to all S–D pairs, is 0.65. This denotes the presence of long memory on S–D traffic.

The S–D expected traffic values, μ_{X_i}s, have been estimated by maximizing the (approximated) log-likelihood (3.7). Their histogram is shown in Figure 7.

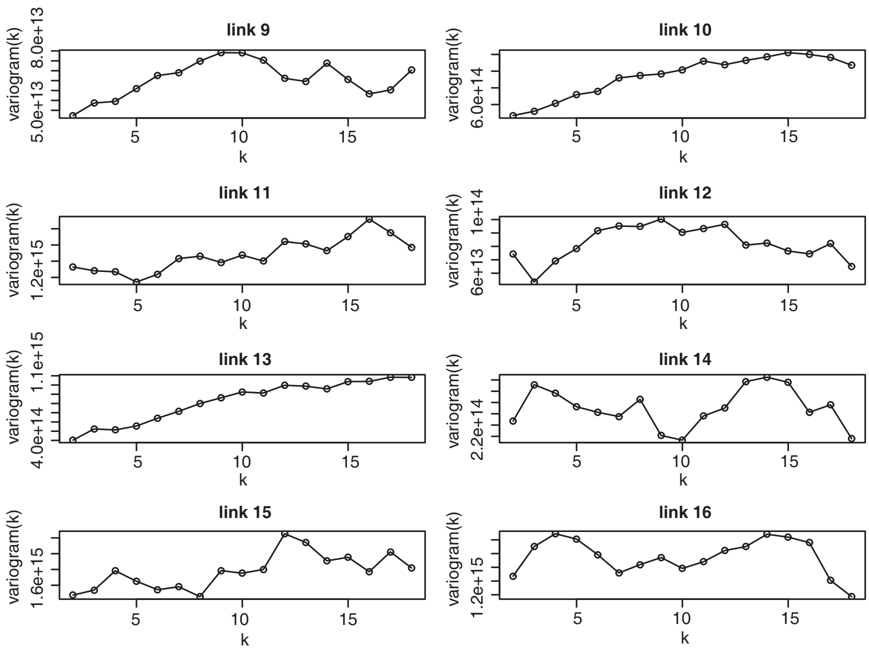

Figure 5 Empirical variograms for links 9–16

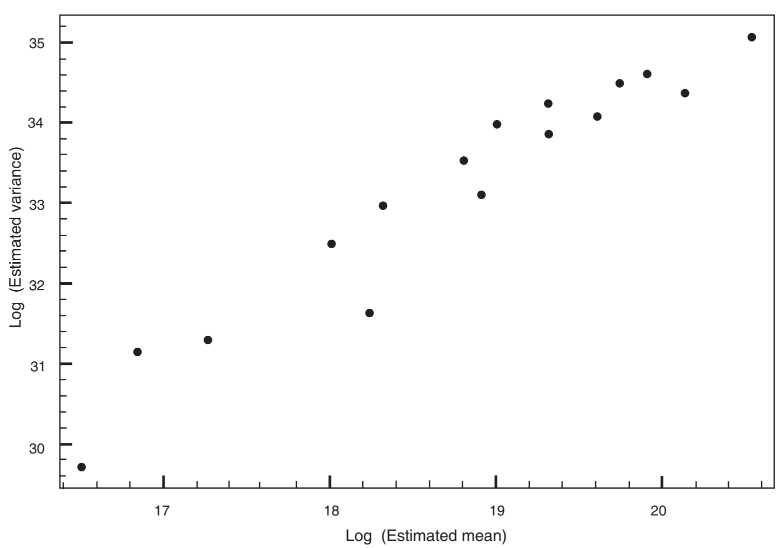

Figure 6 Linear regression of $\log \hat{s}_j^2$ against $\log \hat{m}_j$ for all links

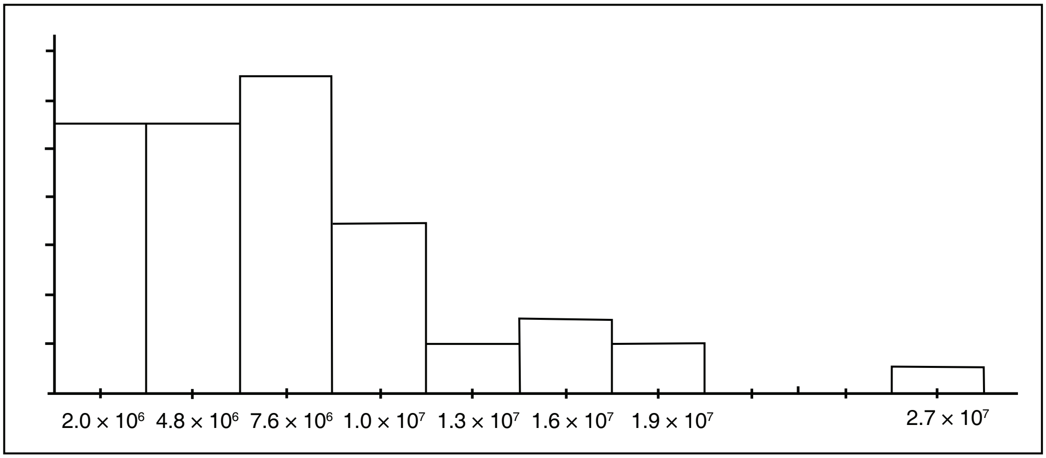

Figure 7 Histogram of S–D expected traffic values

Acknowledgements

This research has been partially supported by PRIN 2007 research project 'Analisi della dipendenza in problemi con struttura di informazione parziale'.

Thanks are due to an anonymous referee and to an associate editor, whose comments considerably improved an earlier version of this article.

Thanks are due to Dr Piergiorgio Aldo Lembo, Research and Development M.gr., Tiscali International Network S.p.A., and Dr Milko Ilari, Enter s.r.l., who kindly provided the data used in Section 5.

Appendix A

Proof of Proposition 1 As already remarked, the process $(Y^t,\ t \geq 1)$ is identifiable if and only if, for every $t \geq 1$, $k \geq 1$, the probability law of $(Y^t,\ Y^{t+k})$ is identifiable. Denote by $p(y^t, y^{t+k};\ H, \theta)$ the density function (d.f., for short) of $(Y^t,\ Y^{t+k})$, which turns out to be $2M$-dimensional normal with mean vector and covariance matrix

$$\begin{bmatrix} Ah(\theta) \\ Ah(\theta) \end{bmatrix}, \quad \begin{bmatrix} A\Sigma_X(\theta)A' & \rho_X(k)\ A\Sigma_X(\theta)A' \\ \rho_X(k)\ A\Sigma_X(\theta)A' & A\Sigma_X(\theta)A' \end{bmatrix}.$$

If $(H^{(1)},\ \theta^{(1)})$ and $(H^{(2)},\ \theta^{(2)})$ are two parameter values, then the identity

$$p(y^t, y^{t+k};\ H^{(1)},\ \theta^{(1)}) \equiv p(y^t, y^{t+k};\ H^{(2)},\ \theta^{(2)})$$

holds if and only if the three relationships

$$Ah(\theta^{(1)}) = Ah(\theta^{(2)}), \quad A\Sigma_X(\theta^{(1)})A' = A\Sigma_X(\theta^{(2)})A', \quad H^{(1)} = H^{(2)}$$

hold true. At this point, the proof is totally similar to the proof of Theorem 1 in Cao *et al.* (2000).

Proof of Proposition 2 First of all, observe that the first M rows of the matrix B possess elements b_{ij} equal to 1 if the arc corresponding to the row i is in the path connecting the S–D pairs of nodes corresponding to the column j, and is equal to 0 otherwise. As far as the remaining $M(M-1)/2$ rows of B are concerned, if the lth row ($l = m+1, \ldots, M(M+1)/2$) is obtained by the component-wise product of rows (i, p ($i \neq p$; i, $p = 1, \ldots, M$)), then the element b_{lj} is equal to 1 if the two arcs corresponding to the rows i and p are in the path connecting the S–D pairs of nodes corresponding to the column j, and is equal to 0 otherwise.

Next, observe that every S–D pair connected by an arc is uniquely identified by that arc. Similarly, all S–D pairs connected by a path of two or more arcs are uniquely identified by the pair of arcs (*first arc in the path*, *last path in the path*). In this way, a one-to-one map T is defined. The map T associates to each S–D pair (a, b) of nodes the arc (a, b) is it exists, and the two arcs (a, a_1) and (a_{l-1}, b) if the two nodes a and b are connected through the path composed by the arcs (a, a_1), (a_1, a_2), $\ldots$, (a_{l-1}, b).

Reorder now the columns of B by considering first the $c_1 = M$ S–D pairs connected by a single arc (i.e., by path of length 1), then the c_2 S–D pairs connected by a path of length 2, then the c_3 S–D pairs connected by a path of length 3, and so on, up to the c_k S–D pairs connected by a path of length k. Clearly, $c_1 + c_2 + \cdots + c_k = N$.

Next, reorder the rows of B in this way:

- the ith row of B, $i = 1, \ldots, N$, represents either the single arc or the pair of arcs corresponding, via the map T, to the S–D pair identified by the ith column of B;
- the remaining $M(M+1)/2 - N$ rows are arbitrarily ordered.

Let now B_1 be the sub-matrix of B composed by its first N rows and $c_1 = M$ columns, B_2 the sub-matrix of B composed by its first N rows and by the next c_2 columns, and so on, up to B_k, which is composed by the first N rows and the last c_k columns of B. Since $c_1 + \cdots + c_k = N$, it is easy to see that

$$\overline{B} = [B_1\ B_2\ \cdots\ B_k] \tag{A1.1}$$

is the sub-matrix of B composed by its first N rows and columns. We now show that the determinant of $\overline{B}$ is equal either to 1 or to -1.

The matrix B_1 does have the following structure:

$$B_1 = \begin{bmatrix} \Pi_{c_1} \\ 0_1 \end{bmatrix}, \tag{A1.2}$$

where Π_{c_1} is a $c_1 \times c_1$ matrix whose rows are obtained by permuting the rows of an identity matrix of order c_1, and 0_1 is a $(N - c_1) \times c_1$ matrix having all elements equal to zero.

Similarly, the matrix B_2 can be written as

$$B_2 = \begin{bmatrix} Q_2 \\ \Pi_{c_2} \\ 0_2 \end{bmatrix},$$

where Q_2 is a $c_1 \times c_2$ matrix having elements equal either to 0 or to 1, Π_{c_2} is a $c_1 \times c_1$ matrix whose rows are obtained by permuting the rows of an identity matrix of order c_2, and 0_2 is a $(N - c_1 - c_2) \times c_2$ matrix having all elements equal to zero.

Similarly, the general sub-matrix B_j in (A1.1) does possess the following structure:

$$B_j = \begin{bmatrix} Q_j \\ \Pi_{c_j} \\ 0_j \end{bmatrix}, \quad j = 2, \ldots, k, \tag{A1.3}$$

where Q_j is a $(c_1 + \cdots + c_{j-1}) \times c_j$ matrix having elements equal either to 0 or to 1, Π_{c_j} is a $c_j \times c_j$ matrix whose rows are obtained by permuting the rows of an identity matrix of order c_j, and 0_j is a $(N - c_1 - \cdots - c_{j-1}) \times c_j$ matrix having all elements equal to zero.

In view of (A1.2) and (A1.3), using the Laplace expansion rule for determinants, it is immediate to conclude that det $(\overline{B}) = \pm 1$. As a consequence, the N columns of B are linearly independent.

Appendix B

We compute here separately the second derivatives of $Q_1(H, \theta)$, $Q_2(H, \theta)$. We have first:

$$\begin{aligned} \frac{\partial^2 Q_1(H, \theta)}{\partial \theta_k \, \partial \theta_l} &= \frac{\partial^2}{\partial \theta_k \, \partial \theta_l} \left\{ \int_{-\pi}^{\pi} \log \det(f_X(\omega, \ H) G(\theta)) \, d\omega \right\} \\ &= 2\pi \frac{\partial}{\partial \, \theta_k} \left\{ \frac{\partial}{\partial \theta_l} \left(\log \det(G(\theta)) \right) \right\} \\ &= 2\pi \frac{\partial}{\partial \, \theta_k} \left\{ vec(G(\theta)^{-1})' \, \boldsymbol{a}_k \right\} \\ &= 2\pi \, \boldsymbol{a}_k' \frac{\partial \, vec(G(\theta)^{-1})}{\partial \, \theta_k} \\ &= -2\pi \, \boldsymbol{a}_k' \{ G(\theta)^{-1} \otimes G(\theta)^{-1} \} \boldsymbol{a}_l . \end{aligned} \tag{B1.1}$$

In the second place, from the equality $\partial G(\theta)^{-1}/\partial\theta_l = -\{G(\theta)^{-1} \otimes G(\theta)^{-1}\}\boldsymbol{a}_l$, using the matrix differentiation rule for Kronecker products (Magnus and Neudecker, 1985), it is seen that

$$\frac{\partial}{\partial\theta_k}\left\{tr\left(\frac{\partial G(\theta)^{-1}}{\partial\theta_l}G(\theta_0)\right)\right\} = \frac{\partial}{\partial\theta_k}\left\{\frac{\partial}{\partial\theta_l}\, tr\left(G(\theta)^{-1}G(\theta_0)\right)\right\}$$
$$= \frac{\partial}{\partial\theta_k}\left\{(vec(G(\theta_0)))'\frac{\partial\, vec(G(\theta)^{-1})}{\partial\,\theta_l}\right\}$$
$$= -(vec(G(\theta_0)))'\frac{\partial}{\partial\theta_k}\left(\{G(\theta)^{-1}\otimes G(\theta)^{-1}\}\boldsymbol{a}_l\right)$$
$$= -(vec(G(\theta_0)))'\left(\boldsymbol{a}_l'\otimes I_{M^2}\right)\frac{\partial}{\partial\theta_k}\left(\{G(\theta)^{-1}\otimes G(\theta)^{-1}\}\right)$$
$$(vec(G(\theta_0)))'\left(\boldsymbol{a}_l'\otimes I_{M^2}\right)\left\{(I_M\otimes C_1 + C_2\otimes I_M)(G(\theta)^{-1}\otimes G(\theta)^{-1})\boldsymbol{a}_k\right\} \tag{B1.2}$$

where

$$C_1 = (K_{M,M}\otimes I_M)(I_M\otimes vec(G(\theta)^{-1})),\quad C_2 = (I_M\otimes K_{M,M})(vec(G(\theta)^{-1})\otimes I_M),$$

I_p being the identity matrix of order p, and $K_{p,p}$ being the commutation matrix (Magnus and Neudecker, 1999) that transforms $vec(A)$ into $vec(A')$. From (B1.2) it is easy to see that

$$\frac{\partial^2 Q_2(H,\,\theta)}{\partial\theta_k\,\partial\theta_l} = (vec(G(\theta_0)))'\left(\boldsymbol{a}_l'\otimes I_{M^2}\right)\left\{(I_M\otimes C_1 + C_2\otimes I_M)(G(\theta)^{-1}\otimes G(\theta)^{-1})\boldsymbol{a}_k\right\}$$
$$\times\int_{-\pi}^{\pi}\frac{f_X(\omega,\,H_0)}{f_X(\omega,\,H)}\,d\omega, \tag{B1.3}$$

and hence:

$$w_{kl}(H,\,\theta) = \frac{\partial^2 Q_1(H,\,\theta)}{\partial\theta_k\,\partial\theta_l} + \frac{\partial^2 Q_2(H,\,\theta)}{\partial\theta_k\,\partial\theta_l},\quad k,\, l = 1,\,\ldots,\,N. \tag{B1.4}$$

where the partial derivatives of Q_1 and Q_2 are given by (B1.1) and (B1.3), respectively.

The element $w_{N+1,N+1}$ is obtained by differentiating twice the function Q w.r.t. H:

$$w_{N+1,N+1}(H,\,\theta) = \frac{\partial^2 Q_1(H,\,\theta) + \partial^2 Q_2(H,\,\theta)}{\partial H^2}, \tag{B1.5}$$

where

$$\frac{\partial^2 Q_1(H,\,\theta)}{\partial H^2} = M\int_{-\pi}^{\pi}\frac{d^2}{dH^2}\log f_X(\omega,\,H)\,d\omega$$
$$= M\int_{-\pi}^{\pi}\frac{f_X''(\omega,\,H)\,f_X(\omega,\,H) - f_X'(\omega,\,H)^2}{f_X(\omega,\,H)^2}\,d\omega \tag{B1.6}$$

and

$$\frac{\partial^2 Q_2(H,\,\theta)}{\partial H^2} = tr(G(\theta)^{-1}G(\theta_0))\int_{-\pi}^{\pi}\frac{d^2}{dH^2}\left(\frac{f_X(\omega,\ H_0)}{f_X(\omega,\ H)}\right)d\omega$$

$$= tr(G(\theta)^{-1}G(\theta_0))\int_{-\pi}^{\pi} f_X(\omega,\ H_0)\left\{2\frac{f'_X(\omega,\ H)^2}{f_X(\omega,\ H)^3} - \frac{f''_X(\omega,\ H)}{f_X(\omega,\ H)^2}\right\}d\omega. \tag{B1.7}$$

Finally, taking into account that

$$\frac{\partial^2 Q_1(H,\,\theta)}{\partial\theta_k\,\partial H} = 0$$

it is easy to see that the terms $w_{k,N+1}$, $k = 1, \ldots, N$ are equal to

$$w_{k,N+1}(H,\,\theta) = \frac{\partial^2 Q_2(H,\,\theta)}{\partial\theta_k\,\partial H}$$

$$= \frac{\partial}{\partial\theta_k}\left\{tr(G(\theta)^{-1}G(\theta_0))\right\}\frac{d}{dH}\left\{\int_{-\pi}^{\pi}\frac{f_X(\omega,\ H_0)}{f_X(\omega,\ H)}d\omega\right\}$$

$$= \left\{(vec(G(\theta_0)))'\left\{G(\theta)^{-1}\otimes G(\theta)^{-1}\right\}a_k\right\}\int_{-\pi}^{\pi}\frac{f'_X(\omega,\ H)}{f_X(\omega,\ H)^2}f_X(\omega,\ H_0)\,d\omega. \tag{B1.8}$$

References

Bear D (1988) *Principles of telecommunication traffic engineering*. London. IEEE/Peter Peregrinus.

Beran J (1994) *Statistics for long-memory processes*. Boca Raton, Florida: CRC Press.

Bermolen P, Vaton S and Juva I (2006) Search for optimality in traffic matrix estimation: a rational approach by cram'er-rao lower bounds. In *Proceedings of the 2nd Conference on Next Generation Internet Design and Engineering (NGI 2006), Valencia, April 2006.*

Cao J, Davis D, Vander Wiel S and Yu B (2000) Time-varying network tomography: Router link data. *Journal of the American Statistical Association*, **95**, 1063–75.

Conti PL, De Giovanni L, Stoev S and Taqqu MS (2008) Confidence intervals for the long memory parameter based on wavelets and resampling. *Statistica Sinica*, **18**, 559–79.

Conti PL, De Giovanni L and Naldi M (2009) Blind maximum-likelihood estimation of traffic matrices in long range dependent traffic. In Valadas, R and Salvador, P (eds). *Proceedings of FITraMEn 2008—Lectures Notes in Computer Science* 5464, pp. 141–54, Berlin: Springer-Verlag.

Crovella ME and Bestavros A (1997) Self-similarity in world wide web taffic: evidence and possible causes. *IEEE/ACM Transactions on Networking*, **5**, 835–46.

Dahlhaus R (1997) Efficient parameter estimation for self-similar processes. *The Annals of Statistics*, **17**, 1749–66.

De Giovanni L and Naldi M (2008) Tests of correlation among wavelet-based estimates for long memory processes. *Communications in Statistics—Simulation and Computation*, **37**, 301–13.

Fay G, Moulines E, François R and Taqqu MS (2008) Estimators of long-memory: Fourier versus wavelets. Technical report, Eprint arXiv:0801.4329.

Fonseca N, Mayor G and Neto C (2000) On the equivalent bandwidth of self-similar sources. *ACM Transactions on Modeling and Computer Simulation*, **10**, 104–24.

Fox R and Taqqu MS (1986) Large sample properties of parameter estimates for strongly dependent stationary gaussian time series. *The Annals of Statistics*, **14**, 517–32.

Giraitis L and Surgailis D (1990) A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of whittle's estimate. *Probab. Theory Related Fields*, **86**, 87–104.

Goldschmidt O (2000) Isp backbone traffic inference methods to support traffic engineering. In *Internet Statistics and Metrics Analysis (ISMA) Workshop, San Diego (Ca), December 2000*.

Gong WB, Liu Y, Misra V and Towsley D (2005) Self-similarity and long range dependence on the internet: a second look at the evidence, origin and implications. *Computer Networks*, **48**, 377–99.

Gunnar A, Johansson M and Telkamp T (2004) Trafficc matrix estimation on a large ip backbone: a comparison on real data. In *4th ACM SIGCOMM conference on Internet measurement*, pp. 149–60.

Hall P, Jing B and Lahiri SN (1998a) On the sampling window method for long range-dependent data. *Statistica Sinica*, **8**, 1189–204.

Hall P, Jing B and Lahiri S (1998b) On the sampling window method for long-range dependent data. *Statistica Sinica*, **8**, 1189–204.

Hosoya Y (1997) A limit theory for long range dependence and statistical inference on related models. *The Annals of Statistics*, **25**, 105–37.

Hosoya Y and Taniguchi M (1982) A central limit theorem for stationary processes and the parameter estimation of linear processes. *The Annals of Statistics*, **10**, 132–53.

Huitema C (1999) *Routing in the Internet*. Upper Saddle River, New Jersey: Prentice-Hall.

Ibragimov IA and Rozanov YA (1978) *Gaussian random processes*. New York: Springer Verlag.

Juva I (2007) *Traffic matrix estimation in the Internet: measurement analysis, estimation methods and applications*. Ph.D. Thesis in Technology, Helsinki University of Technology.

Juva I, Susitaival R, Peuhkuri M and Aalto S (2005) Traffic characterization for traffic engineering purposes: analysis of funet data. In *Proceedings of the 1st EuroNGI Conference on Next Generation Internet Networks (NGI 2005), Rome, 2005*, pp. 404–22.

Juva I, Vaton S and Virtamo J (2006) Quick traffic matrix estimation based on link count covariances. In *Proceedings of the 2006 IEEE International Conference on Communications (ICC 2006), Istanbul, June 2006*, volume 2, pp. 603–8.

Juva I, Susitaival R, Peuhkuri M and Aalto S (2007) Effects of spatial aggregation on the characteristics of origin-destination pair traffic in funet. In *Proceedings of the 7th International Conference on Next Generation Teletraffic and Wired/Wireless Advanced Networking (NEW2AN 2007), St. Petersburg, September 2007*, pp. 1–12. Lectures Notes In Computer Science, 4712, Berlin: Springer-Verlag.

Karatzas I and Shreve SE (1991) *Brownian motion and stochastic calculus*, 2nd edition. New York: Springer Verlag.

Lahiri SN (2003) *Reasampling methods for dependent data*. New York: Springer Verlag.

Leland WE, Taqqu MS, Willinger W and Wilson WD (1994) On the self-similar nature of ethernet traffic extended version. *IEEE/ACM Transactions on Networking*, **2**, 1–15.

Magnus JR and Neudecker H (1985) Matrix differential calculus with applications to simple, hadamard and kronecker products. *Journal of Mathematical Psychology*, **29**, 474–92.

Magnus JR and Neudecker H (1999) *Matrix differential calculus with applications in statistics and econometrics,* 3rd edition. New York: Wiley. Also available at http://cdata4.uvt.nl/websitefiles/magnus/mdc.pdf.

Medina A, Taft N, Salamatian K, Bhattacharyya S and Diot C (2002) Traffic matrix estimation: existing techniques and new directions. In *SIGCOMM02*, Pittsburgh, Pennsylvania, USA.

Naldi M and Salaris C (2006) Rank-size distribution of teletraffic and customers over a wide area network. *European Transactions on Telecommunications*, **17**, 415–21.

Norros I and Kilpi J (2002) Testing the gaussian approximation of aggregate traffic. In *Internet Measurement Workshop IMW2002, Marseille, 2002*.

Norros I and Pruthi P (1996) On the applicability of gaussian traffic models. In Emstad PJ, Helvik BE and MAH (eds), *The Thirteenth Nordic Teletraffic Seminar, Trondheim, August 1996*.

Park C, Hernandez-Campos F, Marron JS and Donelson Smith F (2005a) Long range dependence in a changing internet traffic mix. *Computer Networks*, **48**, 401–22.

Park C, Hernandez-Campos F, Marron JS and Donelson Smith F (2005b) Long range dependence in a changing internet traffic mix. *Computer Networks*, **48**, 401–22.

Park K and Willinger W (2000) *Selfsimilar network traffic and performance evaluation*. New York: Wiley.

Paxson V and Floyd S (1994) Widearea traffic: the failure of poisson modeling. In *Proceedings of the ACM/SIGCOMM '94*, pp. 257–68.

Resnick S and Samorodnitsky G (2001) Steady-state distribution of the buffer content for m/g/∞ input fluid queues. *Bernoulli*, **7**, 191–210.

Susitaival R, Juva I, Peuhkuri M and Aalto S (2006) Characteristics of origin-destination pair traffic in funet. *Telecommunication Systems*, **33**, 67–88.

Taqqu MS, Willinger W and Sherman R (1997) Proof of a fundamental result in selfsimilar traffic modeling. *Computer Communications Review*, **27**, 5–23.

Tebaldi C and West M (1988) Bayesian inference on network traffic uning link count data. *Journal of the American Statistical Association*, **83**, 557–73.

Vardi Y (1996) Network tomography: estimating source-destination traffic intensities from link data. *Journal of the American Statistical Association*, **91**, 365–77.

Vaton S and Gravey A (2002) Iterative bayesian estimation of network traffic matrices in the case of bursty flows. In *Internet Measurement Workshop IMW2002, Marseille, 2002*. Available at http://public.enst-bretagne.fr/vaton/extended-abstract.ps.

Veitch D and Abry P (1999) A wavelet based joint estimator of the parameters of long range dependence traffic. *IEEE Transactions on Information Theory,* special issue on 'Multiscale Statistical Signal Analysis and its Applications', **45**, 878–97.

Whittle P (1952) Some results in time series analysis. *Skand. Aktuarietidskr.*, **35**, 48–60.

Willinger W, Taqqu MS, Leland WE and Wilson DV (1995) Self simlarity in high-speed traffic: analysis and modelling of ethernet traffic measurements. *Statistical Science*, **10**, 67–85.

Willinger W, Taqqu MS, Leland WE and Wilson DV (1998) Self similarity and heavy tails: structural modeling of network traffic. In Adler RJ, Feldman R and Taqqu MS (eds) *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Birkhäuser, Boston, 27–53.

Yajima Y (1985) On estimation of long-memory time series models. *Australian Journal of Statistics*, **27**, 303–20.

Zipf G (1949) *Human behavior and the principle of least effort*. New York: Addison-Wesley.