An Empirical Comparison of MAP Fitting Algorithms

Jan Kriege and Peter Buchholz
Informatik IV, TU Dortmund
D-44221 Dortmund, Germany
{jan.kriege,peter.buchholz}@udo.edu

Author prepared version of a paper published in

Proc. of the 15th International GI/ITG Conference on Measurement, Modelling and Evaluation of Computing Systems and Dependability and Fault Tolerance (MMB & DFT 2010), Springer, 2010.

The original publication is available at www.springerlink.com

http://dx.doi.org/10.1007/978-3-642-12104-3_20

An Empirical Comparison of MAP Fitting Algorithms

Jan Kriege and Peter Buchholz

Informatik IV, TU Dortmund

D-44221 Dortmund, Germany

{jan.kriege,peter.buchholz}@udo.edu

Abstract

The paper presents an empirical comparison of different methods to fit the parameters of a MAP according to the quantities derived from three different real traces. The results indicate that for two of the three traces an adequate fitting with low order MAPs is possible whereas almost all approaches failed for the third trace. Apart form this the question for the best approach for fitting MAPs is still open although there seems to be a tendency that the most costly EM algorithms provide the best fitting results.

1 Introduction

In stochastic modeling, the appropriate representation of arrival and service processes is of major importance to build realistic models. It turns out that many real processes include some correlation which implies that random variables that are identically and independently distributed are not sufficient to describe real behavior, instead stochastic processes have to be used to model the distribution and the autocorrelation structure. Markovian arrival processes (MAPs) [17] are stochastic processes which can be applied to capture a wide range of different stochastic behaviors and can be used in queuing network models as arrival or service processes. Queuing networks with MAPs can be analyzed numerically by solving the global balance equations [23], if the state space is not too large, they can be analyzed with matrix analytical methods [18], if they are of the MAP/MAP/m type, they may as well be analyzed approximately [9] or by simulation.

To capture real behavior by MAPs, the parameters of a MAP have to be fitted according to some trace resulting from observations or measured behavior. The fitting problem of MAPs is a nonlinear optimization problem which becomes even more complex since the matrix representation of MAPs is redundant [24] and a canonical representation is only available for MAPs of order two [3]. Different fitting approaches have been proposed in the literature which all have their pros and cons. The most general approach is to find a MAP that maximizes the likelihood according to the available trace. The EM algorithm [2] can be used for this purpose and many specific variants of the algorithm for MAP fitting are available [4, 5, 14, 22]. However, EM algorithms have several disadvantages since they have a slow convergence, may converge towards local minima and require a huge effort that grows linearly in the length of a trace. Since for MAP fitting very long

traces are required to adequately match the autocorrelation structure, in practice, EM approaches are not sufficient to obtain good fitting result with an acceptable effort. Alternative approaches first derive some quantities from a trace, like higher order moments, joint moments or lag-k autocorrelations and then fit the parameters of a MAP according to these quantities. This implies that fitting becomes independent of the trace length. As shown in [24], a non redundant MAP of order n which is characterized by $2n^2 - n$ free parameters is completely determined by n^2 parameters, e.g., by the first $2n - 1$ moments and $(n-1)^2$ joint moments. Thus, one may fit a MAP according to the empirical moments and joint moments of a trace as done in [7]. Other approaches use the lag-k autocorrelation instead of the joint moments for fitting [10, 13].

However, all these approaches have their limitations since in practice n^2 parameters of a trace hardly define a MAP. In [7] we used least square fitting to obtain the *nearest* MAP of order n according to some measured moments and joint moments. It turns out that it is hard to fit even approximately in the range of n^2 parameters of a real trace with a MAP of order n. Another problem which is also considered in [7] is the reliability of quantities derived from a trace. In general, a trace is only a sample of the behavior of a system such that the quantities computed from the trace are only estimates. If one computes confidence intervals for these quantities, it turns out that confidence intervals become very wide for higher order moments or joint moments of traces from the Internet archive [1] which already contain more than a million entries. This observation implies that for MAP fitting long traces are required.

Although many approaches for MAP fitting are available, it is completely unclear which is the best approach and it is not even clear how to measure whether one approach is better than another. It seems that a lot of empirical work is necessary to find reliable and efficient fitting methods. In this paper we perform such empirical observations by comparing different fitting approaches and different quantities that are fitted. We apply a standard EM approach and two classes of fitting methods that fit first order quantities like joint moments and higher order quantities like lag-k autocorrelations. For this purpose we slightly extend available fitting methods that are based on a two step approach which first fits a phase type (PH) distribution, then possibly do some equivalence transformation on the representation and finally fit a MAP that leaves the distribution unchanged.

The paper is structured as follows. In the next section we introduce the basic notation and recall some basic results for PH distributions and MAPs. In chapter 3 we present several MAP fitting approaches. We start with a brief introduction of EM based MAP fitting and present afterwards two classes of approaches that expand an available PH distribution into a MAP. In the first case, the expansion is done by a least squares approach to fit joint moments. Then we consider the fitting of lag-k autocorrelations. In the following section we use the different fitting methods to fit MAPs according to specific quantities derived from real traces and we compare MAPs fitted with different approaches. The paper ends with the conclusions.

2 Background

We first introduce the basic notation and define PH distributions, then we briefly outline fitting methods for PH distributions and, finally, we present basic results for MAPs.

2.1 Basic Definitions and Results for PH Distributions

A PH distribution [18] of order n is defined by a non-singular $n \times n$ matrix $\mathbf{D}_0$ with $\mathbf{D}_0(i,j) \geq 0$ for $i \neq j$, $\mathbf{D}_0(i,i) \leq -\sum_{j=1, j\neq i}^{n} \mathbf{D}_0(i,j)$ and a row vector π with $\pi(i) \geq 0$ and $\pi\,\mathbb{1} = 1$ where $\mathbb{1}$ is the unit column vector of length n. Let $\mathbf{M} = (-\mathbf{D}_0)^{-1}$, the so called moment matrix. The distribution function, density and the moments of a random variable X with a PH distribution $(\mathbf{D}_0, \pi)$ are given by

$$F_X(t) = 1 - \pi e^{t\mathbf{D}_0}\,\mathbb{1} \tag{1}$$

$$f_X(t) = \pi e^{t\mathbf{D}_0}(-\mathbf{D}_0\,\mathbb{1}) \tag{2}$$

$$\mu_k = E(X^k) = k!\pi\,(\mathbf{M})^k\;\mathbb{1}\,. \tag{3}$$

It has been shown [19] that every non negative random variable with a continuous density that is non-zero in $(0, \infty)$ can be approximated arbitrarily close by a PH distribution.

2.2 Fitting Methods for PH distributions

The task of fitting PH distributions is to choose the parameters of a PH distribution in such a way that some measured quantities are matched. Usually these quantities result from a trace which is an observation of some real behavior. From a trace different quantities like moments, joint moments, lag-k coefficients of autocorrelation or values of the empirical distribution function or density can be computed. Since a trace is only a sample of some real behavior, all values are estimates. The goal of a fitting approach is to find a PH distribution that matches the quantities of the trace as good as possible. A large number of fitting methods for PH distributions exist, an overview can be found in [11].

We only outline a few approaches which we later use as a first step for MAP fitting. In general one can distinguish between fitting methods that work on the whole trace and those that try to match some quantities derived from the trace. Methods of the former type usually maximize the likelihood value which is defined for a trace $t_1, \ldots t_m$ as

$$L_{(\mathbf{D}_0,\pi)}(t_1, \ldots, t_m) = \prod_{k=1}^{m} \pi e^{t_k \mathbf{D}_0}\,(-\mathbf{D}_0\,\mathbb{1})\;. \tag{4}$$

Maximization is done with the so called EM algorithm [2]. However, the general variant of this algorithm is rather inefficient but if one restricts the class of PH distributions, much more efficient variants can be defined. In [25] an EM algorithm which fits the parameters of a generalized Erlang distribution is shown to be rather efficient. We will use this approach as a first step for MAP fitting.

Alternatively, one may fit the PH distribution according to the moments of the trace. In this case acyclic phase type distributions are used since for this subclass a canonical representation exists. Methods for moment fitting which we also apply as a first step for MAP fitting are proposed in [7, 12].

2.3 Basic Definitions and Results for MAPs

A MAP [17] of order n is a stochastic process defined by two $n \times n$ matrices $(\mathbf{D}_0, \mathbf{D}_1)$ where $\mathbf{D}_0$ is as defined for a PH distribution above and $\mathbf{D}_1 \geq 0$ such that $\mathbf{Q} = \mathbf{D}_0 + \mathbf{D}_1$ and $\mathbf{Q}\,\mathbb{1} = \mathbf{0}$. Matrix $\mathbf{D}_0$ contains the rates of internal transitions

without an arrival and matrix $\mathbf{D}_1$ contains the rates of transitions generating an arrival. We assume that $\mathbf{Q}$ is an irreducible generator matrix. Define $\mathbf{P} = -\mathbf{D}_0^{-1}\mathbf{D}_1$ as the transition matrix of the embedded discrete time Markov chain after an arrival. The stationary vector $\pi\mathbf{P} = \pi$, $\pi\,\mathbb{1} = 1$ includes the distribution just after an arrival. Consequently, $(\mathbf{D}_0, \pi)$ describes the interarrival time distribution of a MAP. Similarly each PH distribution $(\mathbf{D}_0, \pi)$ can be expanded into a MAP by defining $\mathbf{D}_1 = -\mathbf{D}_0\,\mathbb{1}\pi$.

The joint moments of consecutive arrivals of a MAP $(\mathbf{D}_0, \mathbf{D}_1)$ are given by

$$\mu_{k,l} = E(X_i^k X_{i+1}^l) = k!\; l!\pi\mathbf{M}^k\mathbf{P}\mathbf{M}^l\,\mathbb{1}\,, \tag{5}$$

the lag-k autocorrelation equals

$$\rho_k = \frac{\mu_1^{-2}\pi(-\mathbf{D}_0)^{-1}\mathbf{P}^k(-\mathbf{D}_0)^{-1}\,\mathbb{1} - 1}{2\mu_1^{-2}\pi(-\mathbf{D}_0)^{-1}(-\mathbf{D}_0)^{-1}\,\mathbb{1} - 1} \tag{6}$$

and the joint density of the first m interarrival times is defined as

$$f(\tau_1, \ldots, \tau_m) = \pi\left(\prod_{i=1}^{m} e^{\tau_i\mathbf{D}_0}\mathbf{D}_1\right)\mathbb{1}\,. \tag{7}$$

Fitting methods as introduced in the subsequent section try to approximate the empirical measures of a trace by a MAP. As for fitting PH distributions either the complete trace may be used resulting in the maximization of the likelihood

$$L_{(\mathbf{D}_0,\mathbf{D}_1)}(t_1, \ldots, t_m) = \pi\left(\prod_{k=1}^{m} e^{t_k\mathbf{D}_0}\mathbf{D}_1\right)\mathbb{1}\,. \tag{8}$$

or some derived quantities like joint moments or lag-k autocorrelations may be used for fitting.

One approach which has been applied successfully [7, 8, 13] is to separate distribution and dependency fitting. In a first step, a PH distribution is generated that captures the distribution of the elements in the trace and in a second step the distribution is expanded into a MAP by considering the dependencies in the trace. This expansion implies that matrix $\mathbf{D}_0$ remains unchanged and $\mathbf{D}_1$ is chosen such that $-\mathbf{D}_0\,\mathbb{1} = \mathbf{D}_1\,\mathbb{1}$ and $\pi\mathbf{M}\mathbf{D}_1 = \pi$ which puts $2n$ constraints for the elements of $\mathbf{D}_1$.

3 MAP Fitting Approaches

3.1 Expectation Maximization

We begin with a brief look on EM algorithms for MAP fitting and refer for the details of the approaches to the literature [2, 4, 5]. All EM algorithms perform an alternating sequence of expectation (E) and maximization (M) steps which improve the likelihood values in each step. Due to the structure of the M-step zero values in the matrices remain zero which implies that no fill in occurs if the algorithm is initially started with sparse matrices. The effort of a single iteration depends linearly on the length of the trace and the number of non-zero elements in the $\mathbf{D}_1$ matrix. Furthermore, it depends on the values in the trace in relation to the matrix entries since $e^{t_i\mathbf{Q}_0}$ has to be evaluated for each entry t_i in the trace and the effort depends on the number of non zero entries in $\mathbf{D}_0$ and the relation between transition rates and time steps. Unfortunately, the convergence of EM algorithms is very slow such that a large number of iterations is required.

If the likelihood is the measure to be maximized, then EM algorithms are currently the best alternative. However, one should start the EM algorithm with a MAP that has already been fitted to the trace using one of the approaches presented in the following two subsections. In this case, the EM algorithm improves the likelihood value but may reduce the fitting quality according to other measures like joint moments or lag-k autocorrelations which have been used to fit the initial MAP. The effort of EM algorithms applied to real traces is usually very high, e.g. for *LBL-TCP-3*, one of the traces we use later, the EM algorithm from [5] requires about 5 minutes per iteration with a MAP of order 5 and about 100 iterations are needed to reach convergence.

3.2 Fitting of Joint Moments

If fitting of the distribution and the autocorrelation structure are done separately, then the matrix $\mathbf{D}_0$ and vector π result from distribution fitting. Since we use the moment fitting approach from [7] or the EM algorithm of [25] for distribution fitting, the result is in both cases an acyclic PH distribution with an upper triangular matrix $\mathbf{D}_0$. Acyclic PH distributions of order n have $n(n+1)/2+(n-1)$ free parameters but only $2n-1$ parameters are necessary to characterize the distribution such that different representations of the same distribution exist. [6] summarizes three methods to perform equivalence transformations that generate different acyclic representations of the same distribution. For MAP fitting the number of non zero entries in π and $\mathbf{D}_0\,\mathbb{1}$ has to be maximized to maximize the number of possible non zero entries in $\mathbf{D}_1$. However, even with this goal the transformation is non unique and different approaches may be tried.

Define $\mathbf{v}^k = \pi\mathbf{M}^{k+1}$ and $\mathbf{w}^k = \mathbf{M}^k\,\mathbb{1}$, then

$$\mu_{k,l} = k!\; l!\; \mathbf{v}^k\mathbf{D}_1\mathbf{w}^l\;. \tag{9}$$

Now assume that $\mathcal{J}$ is a set of joint moments that should be matched by the MAP and let for $(k,l) \in \mathcal{J}$ $\nu_{k,l}$ be the joint moments of the trace. Then the following constrained non negative least squares problem has to be solved to find the nearest MAP.

$$\min_{\mathbf{D}_1:\mathbf{D}_1\geq\mathbf{0},\ \mathbf{D}_1\,\mathbb{1}=-\mathbf{D}_0\,\mathbb{1},\ \pi\mathbf{M}\mathbf{D}_1=\pi} \left(\sum_{(k,l)\in\mathcal{J}} \left(\beta_{k,l}\frac{\mu_{k,l}}{\nu_{k,l}} - \beta_{k,l}\right)^2\right) \tag{10}$$

$\beta_{k,l}$ are some weights which allow one to discriminate higher order joint moments. In our experiments we present later, all weights are set to 1. However, if the resulting MAP cannot match the required moments adequately, it is often appropriate to set the weights such that lower order joint moments get a higher weight, e.g., by choosing $\beta_{k,l} = 2^{-(k-1)(l-1)}$. In this case, lower order joint moments are often matched exactly or almost exactly with the price of a bad fit for higher order moments.

The least squares solution can be computed with available algorithms [15]. The major advantage of joint moment fitting is the efficiency. I.e., to fit the joint moments $\nu_{k,l}$ with $1 \leq k,l \leq 3$ for *LBL-TCP-3* with a MAP of order 5 requires less than 1 second which is negligible compared to the fitting times of EM-algorithms.

3.3 Fitting of Autocorrelations

The approach for the fitting of autocorrelation works similarly to joint moment fitting. In a first step, the initial probability vector π and the matrix $\mathbf{D}_0$ are determined by a PH fitting algorithm like [7] or [25] and are transformed such that the number

of non zero entries is maximized. Then matrix $\mathbf{D}_1$ is generated such that the autocorrelations $\rho = (\rho_1, \cdots, \rho_n)$ of the MAP $(\mathbf{D}_0, \mathbf{D}_1)$ (cf. Eq. 6) approximate the autocorrelations $\hat{\rho} = (\hat{\rho}_1, \cdots, \hat{\rho}_n)$ that have been estimated from the trace, i.e. we have to solve the following minimization problem:

$$\min_{\mathbf{D}_1:\mathbf{D}_1 \geq \mathbf{0},\ \mathbf{D}_1 \mathbb{1} = -\mathbf{D}_0 \mathbb{1},\ \pi \mathbf{M} \mathbf{D}_1 = \pi} \left(\sum_{i=1}^{n} (\beta_i |\rho_i - \hat{\rho}_i|) \right) \tag{11}$$

where the β_i are weights which again may be used to privilege lower lag autocorrelations.

In this paper, we use a slightly modified approach of the two step algorithm presented in [13]. For minimizing Eq. 11 we use the Nelder-Mead algorithm [16]. An implementation can for example be found in [21]. For a MAP of order n we have n^2 variables from matrix $\mathbf{D}_1$ and Nelder-Mead requires $n^2 + 1$ initial solutions $D_1^{(i)}, i = 1, \cdots, n+1$. The first initial solution is the MAP representation of the given PH distribution $(\pi, \mathbf{D}_0)$, i.e. $\mathbf{D}_1^{(1)} = (-\mathbf{D}_0 \mathbb{1})\pi$. The possible range for other valid initial solutions is bounded by the constraints on the row sums ($-\mathbf{D}_0 \mathbb{1} = \mathbf{D}_1 \mathbb{1}$) and on the steady-state vector ($\pi \mathbf{P} = \pi$). Let $\mathbf{x}$ be a vector that contains the first row of matrix $\mathbf{D}_1$ in positions $1, \cdots, n$, the second row in positions $n+1, \cdots, 2n$ etc. Then we can define a linear system of equations using the conditions on row sums and steady-state vector (cf. [13]):

$$\underbrace{\begin{bmatrix} 1 & 1 & \cdots & 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \ddots & & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 & 1 & 1 & \cdots & 1 \\ \pi'(1) & 0 & \cdots & 0 & & \cdots & & \pi'(n) & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \ddots & & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \pi'(1) & & \cdots & & 0 & 0 & \cdots & \pi'(n) \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} \mathbf{D}_1(1,1) \\ \mathbf{D}_1(1,2) \\ \vdots \\ \mathbf{D}_1(1,n) \\ \mathbf{D}_1(2,1) \\ \mathbf{D}_1(2,2) \\ \vdots \\ \mathbf{D}_1(n,n) \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} \begin{bmatrix} -\mathbf{D}_0 \mathbb{1} \end{bmatrix} \\ \pi(1) \\ \pi(2) \\ \vdots \\ \pi(n) \end{bmatrix}}_{\mathbf{b}} \tag{12}$$

where $\pi' = \pi(\mathbf{M})$.

If initial solutions of the Nelder-Mead algorithm differ only slightly, the algorithm gets stuck at a local minimum close to one of the starting points. Therefore we apply the simplex algorithm to find the initial solutions for Nelder-Mead: For each x_i of Eq. 12 we solve

$$\max x_i \ , Ax = b \ , \ \text{ and } x \geq 0$$

which ensures that the Nelder-Mead algorithm has initial solutions with a large stepwidth for each x_i.

Fitting according to autocorrelations is not as efficient as fitting according to joint moments, since the minimization problem is not a simple least squares problem. Depending on the number of lag-k autocorrelations that are considered the approach takes between some seconds and few minutes, e.g. to fit the first 30 lags for *LBL-TCP-3* with a MAP of order 5 the algorithm required less than 10 seconds, for the first 100 lags it took 2 minutes.

4 Experimental Results

To compare the different fitting algorithms we use three different traces. The trace *BC-pAug89* contains a million packet arrivals observed at the Bellcore Morristown Research and Engineering facility in August 1989. The trace *LBL-TCP-3* [20] contains two hours of TCP traffic from the Lawrence Berkeley Laboratory and was recorded in January 1994. Both traces are taken from the Internet Traffic Archive [1]. The third trace *TUDo* contains the interarrival times of one million packets that have been measured from the Squid proxy server at the Computer Science Department of TU Dortmund in 2006.

We fitted MAPs of different order (from $n = 2$ to $n = 6$) with the three fitting approaches from Sec. 3. Since fitting according to joint moments and autocorrelations both require a given distribution that is fitted in the first step, we used Gfit [25] and a moment matching approach [7]. In a first step of our empirical evaluation we will compare the joint moments, lag-k autocorrelations and the likelihood of MAPs that have been fitted according to one of the characteristics with MAPs for which other properties have been used in the fitting process. The second part of our empirical evaluation compares the queueing behavior. In the following we will present the results for some of the fitted MAPs.

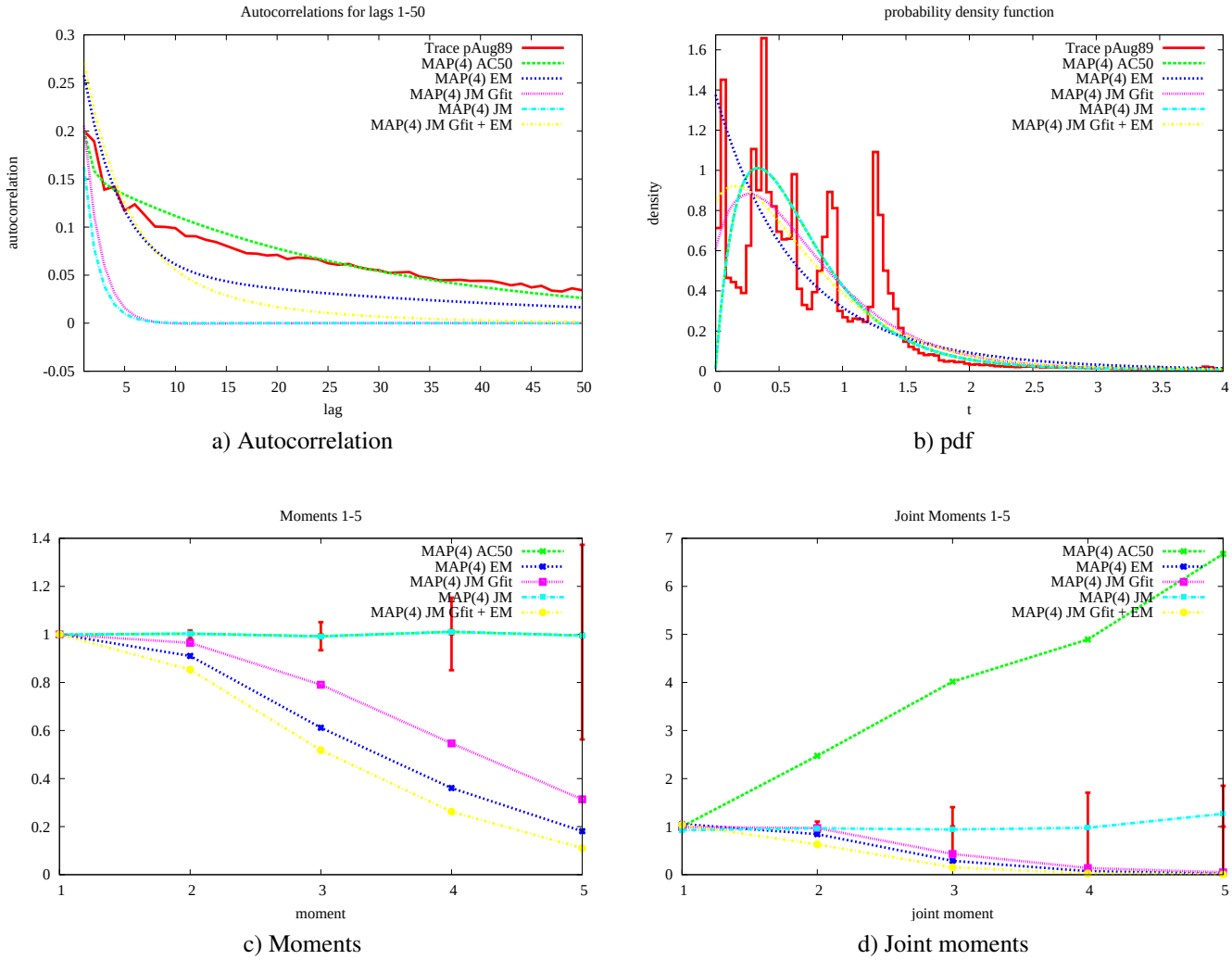
4.1 Comparison of Quantities

We start with the comparison of the fitted MAPs for the trace *BC-pAug89*. Figs. 1 and 2 show the results for MAPs of order 4 and 6 that we obtained for the different fitting algorithms. The curves resulting from autocorrelation fitting are labeled with `AC` and the number of lags that have been considered for fitting. Curves from joint moment fitting and expectation maximization are labeled with `JM` and `EM`, respectively. For moment matching we used the first five moments $\nu_k, (k = 1...5)$ and for joint moment fitting 25 joint moments $\nu_{k,l}, (k, l = 1...5)$ and set all weights β_k and $\beta_{k,l}$ to 1. Usually we used moment matching [7] to obtain the distribution for joint moment and autocorrelation fitting. In cases where Gfit [25] was used this is denoted in the plots. In addition to the pure EM algorithm that starts with a random MAP to improve the likelihood we used the MAPs resulting from AC and JM fitting as initial solutions for the EM algorithm as mentioned in Sec. 3.1. These MAPs are labeled with `JM + EM` or `AC + EM`. The likelihood values for the MAPs are shown in Table 1.

Likelihood	**Trace pAug89**	**Likelihood**	**Trace pAug89**
MAP(4) AC50	-891757.644354	MAP(6) AC100	-1138434.529730
MAP(4) EM	-857368.761951	MAP(6) EM	-850032.779585
MAP(4) JM Gfit	-833728.156640		
MAP(4) JM	-879217.986678	MAP(6) JM	-1142582.505733
MAP(4) JM Gfit + EM	-806247.089218	MAP(6) AC100 + EM	-804071.232488

Table 1: Likelihood for the MAPs of order 4 and 6 for the trace *BC-pAug89*

The MAPs resulting from joint moment fitting failed to capture the autocorrelations, while both autocorrelation fitting and EM algorithm resulted in a much better approximation of the lag-k autocorrelations, although the latter tends to underestimate the autocorrelation. In contrast EM and AC fitting do not capture the joint moments of the trace, while, of course, JM fitting

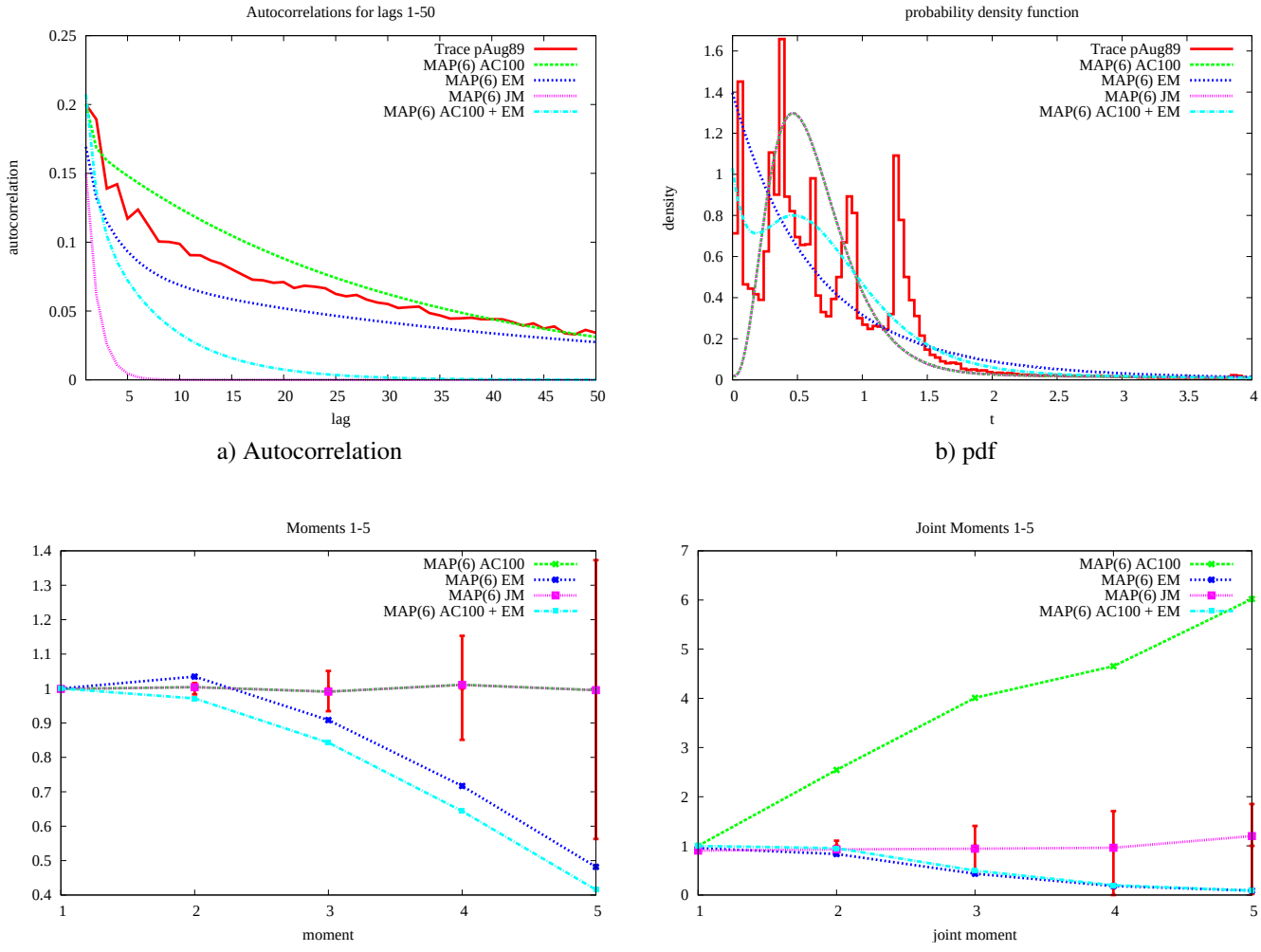


a) Autocorrelation b) pdf

c) Moments d) Joint moments

Figure 1: Fitting results for MAPs of order 4 for the trace *BC-pAug89*

provides a good approximation as one can see from Figs. 1 d) and 2 d). The curves show the joint moments $\mu_{k,k}$ of the MAPs relative to the joint moments of the trace. Joint moments $\mu_{k,l}(k \neq l)$ are not shown but are similar. Additionally the confidence intervals of the joint moments are printed in red. For all the MAPs we fitted with the EM algorithm the joint moments are smaller than the ones of the trace, while AC fitting resulted in larger joint moments. Similarly the MAPs resulting from the EM algorithm underestimated the higher moments of the trace (cf. Figs. 1 c) and 2 c)). The distributions of the MAPs are shown in Figs. 1 b) and 2 b). Note, that AC fitting and JM fitting used the same PH distribution. Regarding the likelihood EM fitting provides a larger likelihood than AC fitting and JM fitting that used a PH distribution obtained from the moment matching algorithm (cf. Table 1). Interestingly, using a PH distribution that has been fitted by Gfit as basis for MAP fitting resulted in a very high likelihood. As a drawback those PH distributions showed to be less flexible for the subsequent AC fitting and JM fitting compared to PH distributions resulting from the moment matching approach.



a) Autocorrelation b) pdf

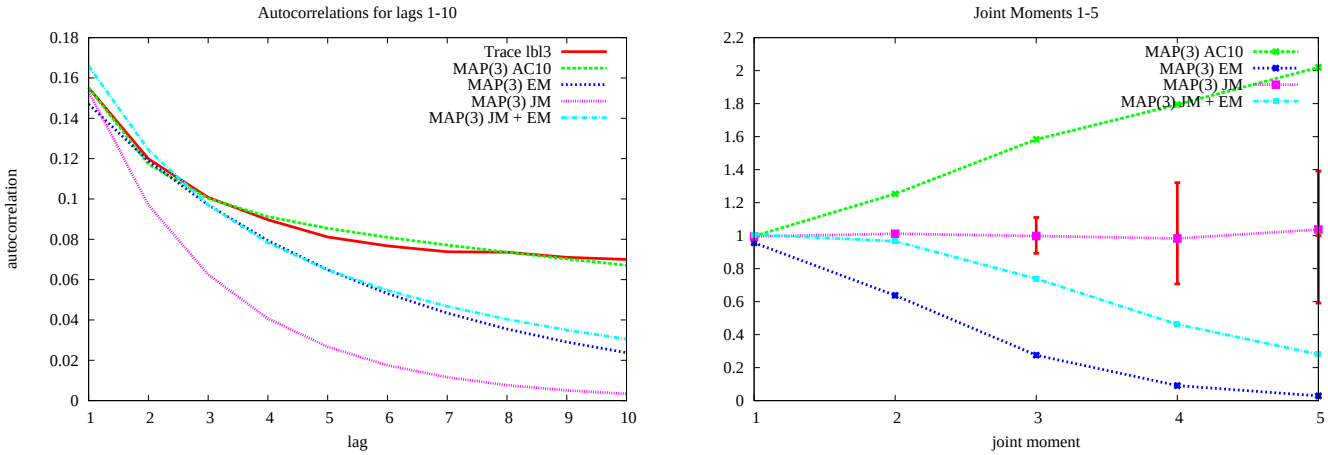
c) Moments d) Joint moments

Figure 2: Fitting results for MAPs of order 6 for the trace *BC-pAug89*

As already mentioned we used the MAPs resulting from AC and JM fitting as initial solution for the EM algorithm. In these cases EM fitting was able to improve the likelihood significantly, although the fitting quality according to other measures was reduced as one can see from Figs. 1 and 2.

Figs. 3 and 4 and Table 2 show fitting results for the trace *LBL-TCP-3*. The results are similar to the previous trace: JM fitting and to a lesser degree EM fitting underestimate the autocorrelations, while AC fitting over- and EM fitting underestimate the joint moments.

The last trace we used for our comparison was observed at a proxy server at TU Dortmund. It contains various bursts with very small interarrival times followed by a larger break until the next burst and therefore has high autocorrelations. As one can see from Figs. 5 and 6 the trace was difficult to fit for all algorithms. Again we used the MAPs resulting from AC and JM fitting as initial solutions for the EM algorithm. The $MAP(4)$ resulting from AC fitting using a PH distribution obtained



a) Autocorrelation b) Joint moments

Figure 3: Fitting results for MAPs of order 3 for the trace *LBL-TCP-3*

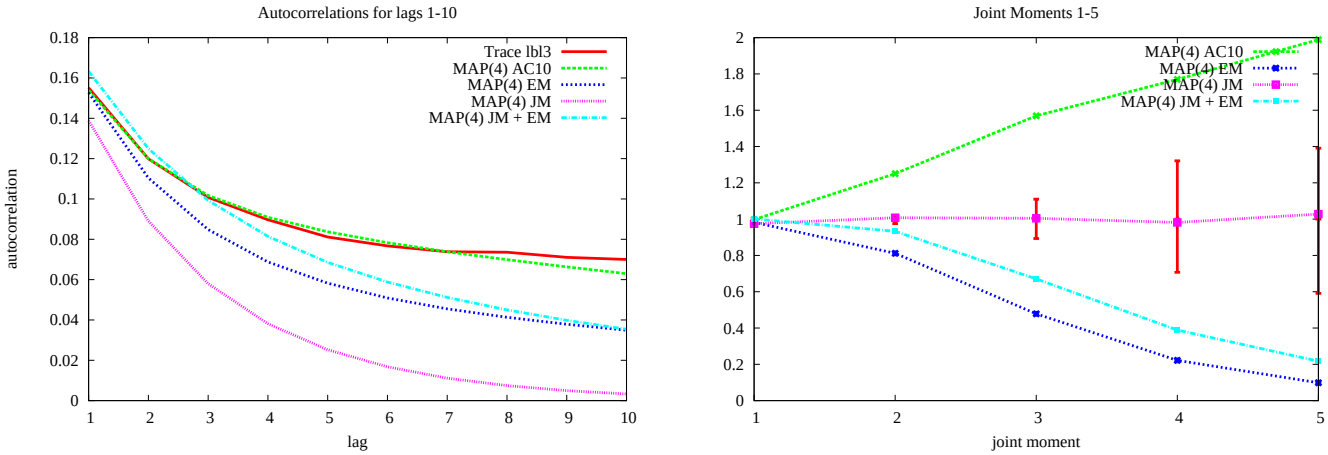
Likelihood	**Trace lbl3**	**Likelihood**	**Trace lbl3**
MAP(3) AC10	-1652039.018594	MAP(4) AC10	-1654146.788752
MAP(3) EM	-1637813.770617	MAP(4) EM	-1627420.100721
MAP(3) JM	-1639440.176623	MAP(4) JM	-1647272.298004
MAP(3) JM + EM	-1626938.629808	MAP(4) JM + EM	-1626267.526219

Table 2: Likelihood for the MAPs of order 3 and 4 for the trace *LBL-TCP-3*

from Gfit showed to be unsuitable for this task, since the structure of the MAP caused a very poor runtime performance of the EM algorithm. Hence, we only used the MAPs that resulted from moment matching and subsequent AC or JM fitting as initial solutions.

4.2 Comparison of Queueing Behavior

Table 4 shows the queueing results for the different traces. The original traces and the fitted MAPs are used as arrival processes, for the service process we used an exponential distribution with different rates between $\mu = 0.6$ and $\mu = 1.2$. We use a single server system with capacity 10. The system was simulated for each of the traces until all of the interarrival times from the trace have been used. After that we simulated the model with the fitted MAPs for the same amounts of time. Table 4 contains the mean queue length and the probability that the queue is completely filled for all combinations of arrival and service processes. For each trace and order of the MAPs the results that are closest to the results of the trace are emphasized. For the Trace *BC-pAug89* one can see that the MAPs resulting from the EM algorithm result in a mean queue length that is closest to the one of the trace. Regarding the probability that the queue is completely filled all the fitted MAPs provide an appropriate approximation of the results from the trace, although in almost all cases either the pure EM or the



a) Autocorrelation b) Joint moments

Figure 4: Fitting results for MAPs of order 4 for the trace *LBL-TCP-3*

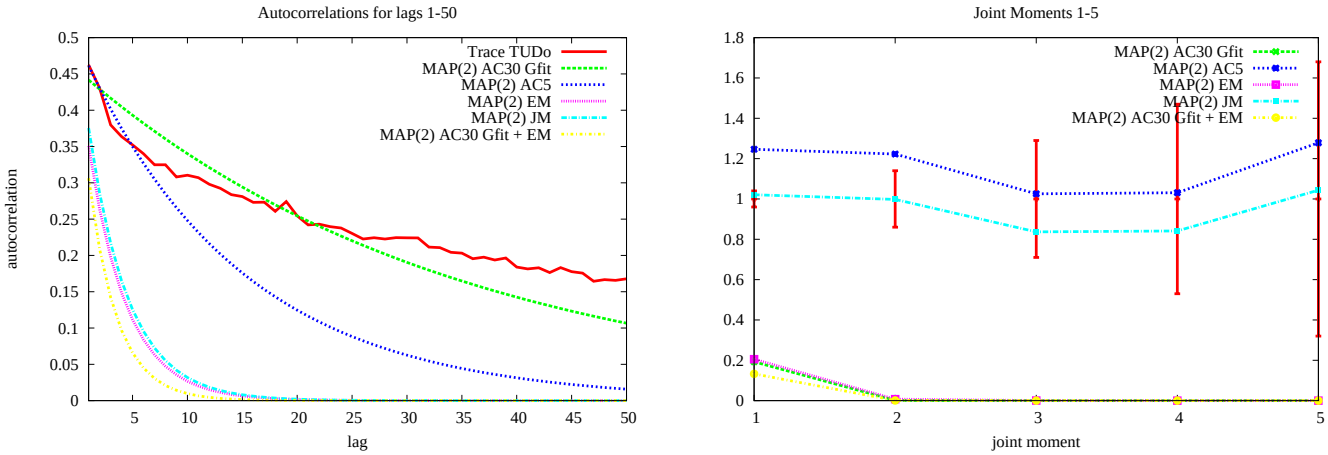
Likelihood	Trace TUDo	Likelihood	Trace TUDo
MAP(2) AC30 Gfit	409209.247942	MAP(4) AC30 Gfit	450455.635351
MAP(2) AC5	17874.468130	MAP(4) AC10	297622.747212
MAP(2) EM	34339.840111	MAP(4) EM	134878.147095
MAP(2) JM	33220.591429	MAP(4) JM	285171.814999
MAP(2) AC30 Gfit+EM	487457.907053	MAP(4) AC10 + EM	385977.643978

Table 3: Likelihood for the MAPs of order 2 and 4 for the trace *TUDo*

EM algorithm combined with one of the other approaches provided the closest approximation. For the trace *LBL-TCP-3* we obtained similar results. As already mentioned all fitting algorithms had problems with the MAP *TUDo*. This becomes also visible in Table 4. Only the MAP(4) that has been fitted with Gfit and a subsequent fitting of the autocorrelations provided a sufficient approximation of the queueing behavior.

5 Conclusions

This paper presents a comparison of different MAP fitting approaches applied to three different traces. Two of the traces have been taken from the Internet archive and have been used several times as benchmarks for MAP fitting approaches. However, these traces are also very old. The third trace is much newer and shows different characteristics. Our results indicate that the older traces can be fitted adequately with most approaches whereas the new trace exhibits a much stronger autocorrelation and is much harder to fit. It is an interesting question whether current network traffic, which probably differs from the traffic analyzed twenty years ago, really contains higher autocorrelations or whether this is an artifact in our measurements.



a) Autocorrelation b) Joint moments

Figure 5: Fitting results for MAPs of order 2 for the trace *TUDo*

However, to answer this question, more measurements are necessary.

The comparison of the different fitting methods gives a mixed picture. Obviously, using a method that fits a MAP according to one quantity, like the joint moments or the autocorrelation, gives good results according to this quantity but usually results in a bad fitting according to other quantities that are not used for fitting. Thus, no approach is superior to all others according to all quantities. However, our results indicate that fitting according to the likelihood using the EM algorithm gives the best results but is, unfortunately, also the by far most costly method. Furthermore, it should be mentioned that the fitting quality and the effort of the EM algorithm depends on the initial MAP and might be poor for badly chosen initial MAPs.

References

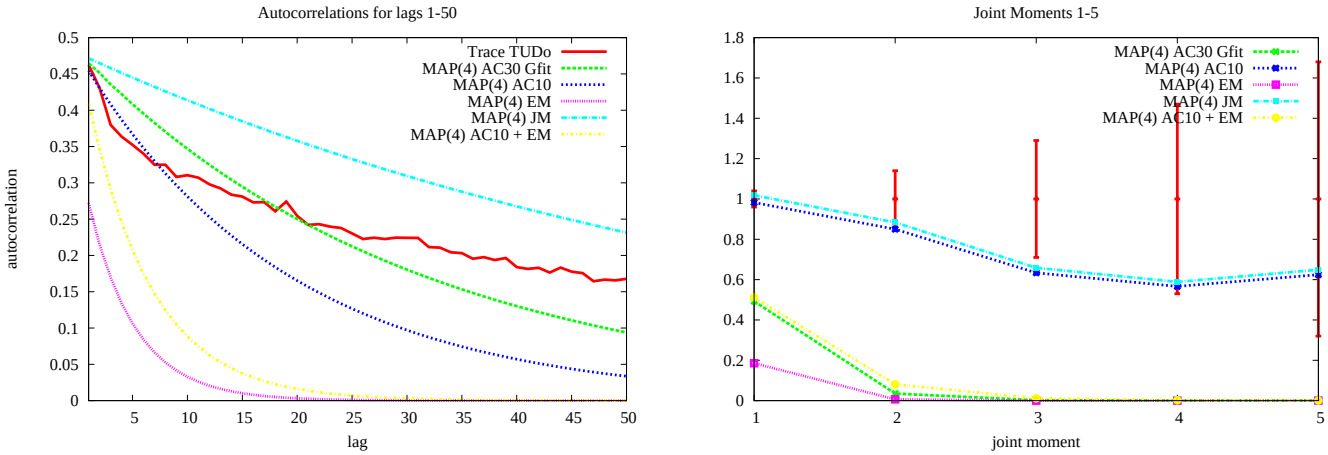
[1] The internet traffic archive. http://ita.ee.lbl.gov/.

[2] S. Asmussen, O. Nerman, and M. Olsson. Fitting phase-type distributions via the EM-algorithm. *Scand. J. Stat.*, 23(4):419–441, 1996.

[3] L. Bodrog, A. Heindl, G. Horváth, and M. Telek. A Markovian canonical form of second-order matrix-exponential processes. *European Journal of Operational Research*, 190(2):459–477, 2008.

[4] L. Breuer. An EM algorithm for batch Markovian arrival processes and its comparison to a simpler estimation procedure. *Annals of Operations Research*, 112:123–138, 2002.

[5] P. Buchholz. An EM-algorithm for MAP fitting from real traffic data. In P. Kemper and W. H. Sanders, editors, *Computer Performance Evaluation / TOOLS*, volume 2794 of *Lecture Notes in Computer Science*, pages 218–236. Springer, 2003.



a) Autocorrelation b) Joint moments

Figure 6: Fitting results for MAPs of order 4 for the trace *TUDo*

[6] P. Buchholz and J. Kriege. Equivalence transformations for acyclic phase type distributions. Technical Report 827, Dep. of Informatics, TU Dortmund, 2009.

[7] P. Buchholz and J. Kriege. A heuristic approach for fitting MAPs to moments and joint moments. In *Proc. of 6th International Conference on Quantitative Evaluation of SysTems (QEST09)*. IEEE, 2009.

[8] P. Buchholz and A. Panchenko. A Two-Step EM Algorithm for MAP Fitting. In C. Aykanat, T. Dayar, and I. Korpeoglu, editors, *ISCIS*, volume 3280 of *Lecture Notes in Computer Science*, pages 217–227. Springer, 2004.

[9] G. Casale and E. Smirni. MPA-AMVA: Approximate mean value analysis of bursty system. In *Proc. DSN 09*, 2009.

[10] G. Casale, E. Z. Zhang, and E. Smirni. KPC-Toolbox: Simple Yet Effective Trace Fitting Using Markovian Arrival Processes. In *QEST*, pages 83–92. IEEE Computer Society, 2008.

[11] A. Horváth and M. Telek. Markovian modeling of real data traffic: Heuristic phase type and MAP fitting of heavy tailed and fractal like samples. In M. Calzarossa and S. Tucci, editors, *Performance*, volume 2459 of *Lecture Notes in Computer Science*, pages 405–434. Springer, 2002.

[12] A. Horváth and M. Telek. PhFit: A general phase-type fitting tool. In T. Field, P. G. Harrison, J. T. Bradley, and U. Harder, editors, *Computer Performance Evaluation / TOOLS*, volume 2324 of *Lecture Notes in Computer Science*, pages 82–91. Springer, 2002.

[13] G. Horvath, M. Telek, and P. Buchholz. A MAP fitting approach with independent approximation of the inter-arrival time distribution and the lag-correlation. In *Proc. of 2nd Int. Conf. on the Quantitative Analysis of Systems*. IEEE CS Press, 2005.

[14] A. Klemm, C. Lindemann, and M. Lohmann. Modeling IP traffic using the batch markovian arrival process. *Perform. Eval.*, 54(2):149–173, 2003.

[15] C. L. Lawson and B. J. Hanson. *Solving Least Squares Problems*. Prentice-Hall, 1974.

[16] J.A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7:308–313, 1965.

[17] M. F. Neuts. A versatile Markovian point process. *Journal of Applied Probability*, 16:764–779, 1979.

[18] M. F. Neuts. *Matrix-geometric solutions in stochastic models*. Johns Hopkins University Press, 1981.

[19] C. A. O'Cinneide. Characterization of phase-type distributions. *Stochastic Models*, 6:1–57, 1990.

[20] V. Paxson and S. Floyd. Wide-area traffic: The failure of Poisson modeling. *IEEE/ACM Transactions in Networking*, 3:226–244, 1995.

[21] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C - The Art of Scientific Computing*. Cambridge University Press, 2nd edition, 1993.

[22] T. Ryden. An EM algorithm for estimation in Markov-modulated Poisson processes. *Comput. Statist. Data Anal.*, 21(4):431–447, 1996.

[23] W. J. Stewart. *Introduction to the numerical solution of Markov chains*. Princeton University Press, 1994.

[24] M. Telek and G. Horváth. A minimal representation of Markov arrival processes and a moments matching method. *Perform. Eval.*, 64(9-12):1153–1168, 2007.

[25] A. Thümmler, P. Buchholz, and M. Telek. A Novel Approach for Phase-Type Fitting with the EM Algorithm. *IEEE Trans. Dependable Sec. Comput.*, 3(3):245–258, 2006.

Model	**Mean queue length**				**Probability for full queue**			
	$\mu = 0.6$	0.8	1.0	1.2	$\mu = 0.6$	0.8	1.0	1.2
Trace pAug89	6.688	5.530	4.333	3.282	0.347	0.234	0.150	0.094
MAP(4) AC50	7.466	6.315	4.955	3.540	0.3779	0.246	0.139	0.066
MAP(4) EM	**6.749**	**5.556**	**4.352**	**3.321**	0.328	0.219	0.140	**0.089**
MAP(4) JM Gfit	7.363	6.290	5.027	3.627	0.374	0.249	0.143	0.068
MAP(4) JM	7.316	6.156	4.909	3.626	0.367	0.243	0.145	0.072
MAP(4) JM Gfit+EM	6.752	5.645	4.521	3.428	**0.341**	**0.231**	**0.146**	0.087
MAP(6) AC100	7.462	6.298	4.941	3.532	0.383	0.251	0.143	0.068
MAP(6) EM	**6.736**	**5.617**	**4.459**	**3.392**	**0.333**	**0.226**	0.145	0.091
MAP(6) JM	7.367	6.151	4.899	3.609	0.370	0.246	**0.147**	0.075
MAP(6) AC100+EM	6.926	5.861	4.868	3.797	**0.361**	0.253	0.164	**0.093**
Trace lbl3	7.081	5.407	4.009	2.969	0.304	0.183	0.110	0.067
MAP(3) AC10	7.309	6.015	4.590	3.312	0.333	0.206	0.115	0.059
MAP(3) EM	7.467	5.641	**4.117**	**3.019**	**0.309**	0.178	0.103	0.061
MAP(3) JM	7.461	5.838	4.325	3.163	0.313	0.184	0.105	0.059
MAP(3) JM+EM	**7.302**	**5.591**	4.121	3.020	0.310	**0.183**	**0.107**	**0.064**
MAP(4) AC10	7.311	6.016	4.593	3.313	0.335	0.208	0.116	0.059
MAP(4) EM	7.376	5.580	**4.029**	**2.939**	**0.308**	0.178	0.103	0.063
MAP(4) JM	7.528	5.941	4.432	3.212	0.323	0.193	**0.108**	0.058
MAP(4) JM+EM	**7.295**	**5.565**	4.092	3.012	0.309	**0.183**	**0.108**	**0.065**
Trace TUDo	2.510	2.226	1.999	1.816	0.137	0.114	0.098	0.086
MAP(2) AC30 Gfit	1.317	1.163	1.088	1.038	0.087	0.083	0.080	**0.078**
MAP(2) AC5	4.202	4.110	3.997	3.849	0.317	0.279	0.241	0.204
MAP(2) EM	4.894	4.552	4.245	3.934	0.308	0.260	0.214	0.172
MAP(2) JM	4.383	4.254	4.116	3.945	0.322	0.283	0.244	0.204
MAP(2) AC30 Gfit+EM	**3.645**	**2.886**	**2.417**	**2.100**	**0.119**	**0.096**	**0.083**	0.074
MAP(4) AC30 Gfit	**2.118**	**2.051**	**1.999**	**1.949**	**0.156**	**0.142**	**0.129**	0.118
MAP(4) AC10	4.343	3.930	3.547	3.186	0.236	0.193	0.158	0.131
MAP(4) EM	4.283	3.949	3.688	3.457	0.267	0.232	0.200	0.169
MAP(4) JM	3.862	3.507	3.161	2.847	0.211	0.172	0.141	**0.117**
MAP(4) AC10+EM	3.541	3.214	2.951	2.732	0.215	0.185	0.162	0.144

Table 4: Queueing results