Stochastic system identification for ATM network traffic models: a time domain approach[1]

Katrien De Cock, Tony Van Gestel and Bart De Moor[2]

October 1997

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.ac.be* in the directory *pub/SISTA/decock/reports/97-90.ps.gz*

[2]K.U.Leuven, Dept. of Electrical Engineering (ESAT), Research group SISTA, Kardinaal Mercierlaan 94, 3001 Leuven, Belgium, Tel. 32/16/32 17 09, Fax 32/16/32 19 70, WWW: *http://www.esat.kuleuven.ac.be/sista*. E-mail: *katrien.decock@esat.kuleuven.ac.be, tony.vangestel@esat.kuleuven.ac.be, bart.demoor@esat.kuleuven.ac.be*. Bart De Moor is a senior Research Associate with the F.W.O. (Fund for Scientific Research-Flanders). Katrien De Cock is a Research Assistant with the I.W.T. (Flemish Institute for Scientific and Technological Research in Industry). Tony Van Gestel is a Research Assistant with the F.W.O. (Fund for Scientific Research-Flanders). Work supported by the **Flemish Government** (Administration of Science and Innovation (Concerted Research Action MIPS: *Model-based Information Processing Systems*, Bilateral International Collaboration: *Modelling and Identification of nonlinear systems*, IWT-Eureka SINOPSYS: *Model-based structural monitoring using in-operation system identification*), FWO-Vlaanderen: *Analysis and design of matrix algorithms for adaptive signal processing, system identification and control, based on concepts from continuous time system theory and differential geometry*, *Numerical algorithms for subspace system identification: Extension towards specific applications*, FWO-Onderzoeksgemeenschappen: *Identification and Control of Complex Systems*, *Advanced Numerical Methods for Mathematical Modelling*); **Belgian Federal Government**(Interuniversity Attraction Pole IUAP IV/02: *Modelling, Identification, Simulation and Control of Complex Systems*, Interuniversity Attraction Pole IUAP IV/24: IMechS: Intelligent Mechatronic Systems); **European Commission:** (Human Capital and Mobility: *SIMONET: System Identification and Modelling Network*, *SCIENCE-ERNSI: European Research Network for System Identification.*)

Stochastic system identification for ATM network traffic models: a time domain approach

Katrien De Cock[1], Tony Van Gestel[2] and Bart De Moor[3]

K.U.Leuven, Dept. of Electrical Engineering (ESAT),
Research group SISTA,
Kardinaal Mercierlaan 94,
3001 Leuven, Belgium,
Tel. 32/16/32 19 25, Fax 32/16/32 19 70,
WWW: *http://www.esat.kuleuven.ac.be/sista.*
E-mail: *katrien.decock@esat.kuleuven.ac.be,*
tony.vangestel@esat.kuleuven.ac.be, bart.demoor@esat.kuleuven.ac.be.

Abstract

In this paper we discuss a time domain approach for traffic identification problems in ATM networks, for which we propose to identify a Markov modulated Poisson process and a circulant Markov modulated Poisson process, based on unconstrained optimisation algorithms. By applying a nonnegative least squares algorithm we obtain in a very fast way a description of the first order statistics of the data. This first order characterisation also includes an estimate for the model order. Consequently, we manage to identify a Markov modulated Poisson process without a priori knowledge of the model order.

Keywords: traffic identification, Markov modulated Poisson process, mixture distributions, ATM, subspace identification, parameter estimation.

---

[1]Research Assistant with the I.W.T. (Flemish Institute for Scientific and Technological Research in Industry).

[2]Research Assistant with the F.W.O. (Fund for Scientific Research-Flanders).

[3]Senior Research Associate with the F.W.O. (Fund for Scientific Research-Flanders).

Contents

1 Introduction **4**

2 Mathematical models in ATM networks **5**

3 The identification problem **6**
3.1 Model choice 6
3.2 The Markov modulated Poisson process 7
3.2.1 Model description 7
3.2.2 The first and second order statistics of arrivals 9
3.3 Identification of the MMPP by unconstrained optimisation . 10
3.4 The circulant Markov modulated Poisson process 12
3.4.1 Model structure 12
3.4.2 Properties of circulant stochastic matrices 12
3.5 Identification of the CMPP by unconstrained optimisation . . 14
3.6 Comparison of MMPP and CMPP identification 15

4 Characterisation of the first order statistics without a priori knowledge of the model order **16**
4.1 Order estimation of a Markov modulated Poisson process . . 17
4.2 Determination of the components of a mixture distribution . 19
4.3 Example 22
4.3.1 Determination of the model order based on the auto-correlation function [28] 24
4.3.2 Determination of the order based on the cumulative distribution function 28

5 Identification of an MMPP without a priori knowledge of the model order **29**

6 Identification of a CMPP without a priori knowledge of the model order **31**
6.1 Translation of the first order MMPP parameters into the first order CMPP parameters 31
6.2 Example 33
6.3 A more practical version for the translation 34
6.4 The complete identification method for CMPPs 35

7 Comparison between the different time domain identification methods **36**
7.1 Identification of an MMPP 36
7.1.1 The optimisation step 36
7.1.2 Operations preceding the optimisation 37
7.2 Identification of a CMPP 39
7.2.1 The optimisation step 39
7.2.2 Operations preceding the optimisation 40
7.3 Comparison between MMPP and CMPP identification 40

8 Conclusions **42**

1 Introduction

Mathematical models have had a rather passive role in communication networks. Li *et al.* [15] have indicated that mathematical models can be used to perform several tasks in control mechanisms of ATM networks. The models they propose are measurement based and exploit the time correlation of traffic. Whereas the approach of Li *et al.* is mainly based on the frequency domain [15, 16, 17], this paper is concerned with measurement based parameter estimation in the time domain: the models match the cumulative distribution function and the autocorrelation function of the measured data as described in [28]. An important innovation described in this paper is the fast characterisation of the first order statistics with automatic determination of the model order.

We begin with an introductory section where we briefly motivate the use of measurement based models in ATM networks and the role of identification in traffic modelling. We direct our attention to the modelling of the cell arrival process in one node of the network.

Section 3 describes the identification problem. First the choice of the model class is justified. Then the model, a Markov modulated Poisson process, and the corresponding identification algorithm are described. Next, a more restricted model, the circulant Markov modulated Poisson process, is introduced. Some of its interesting properties are discussed and an identification algorithm for this model is obtained. This section is concluded with a comparison of the identification with the two models.

In Section 4 the order estimation problem is tackled. We briefly discuss the method of Yi and De Moor [28] which is based on the autocorrelation of the data. Our new approach starts from the first order statistics: the number of components in the mixture distribution, the components themselves and their distribution are determined in one step. In an example we compare both techniques. This new method clearly outperforms that of Yi and De Moor [28].

In Section 5 and Section 6 we use this first order characterisation method to identify the Markov modulated Poisson process and its circulant version without a priori knowledge of the order.

All described time domain identification methods are compared in Section 7. Finally we end with some conclusions.

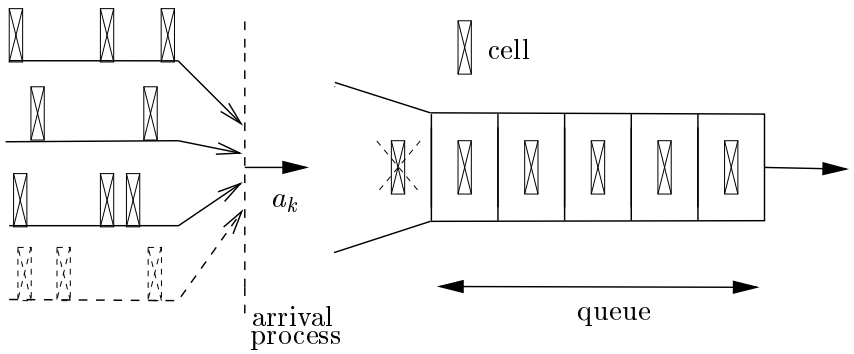

Figure 1: Buffer in a switch of an ATM network. The cells, coming from different users, enter the node according to a stochastic profile a_k ($k = 0, 1, 2, \ldots$). They have to wait in a queue before they can be served. As a consequence the cells encounter a delay. Some cells may even get lost if the buffer's capacity is exceeded.

2 Mathematical models in ATM networks

Asynchronous Transfer Mode (ATM) is a new standard for broadband telecommunication. Important features of ATM are the high transmission rates and the possibility to integrate all types of services (voice, video, data, ...). Mathematical models are used to facilitate the connection admission control (CAC): Whenever a new user wants to set up a connection, the server has to decide whether or not this user can be allowed. If a user is admitted, the network has to guarantee a certain negotiated level of quality of service (QoS) for the new user and for all present users. The traffic sent by the user, on the other hand, must conform to the features in the contract.
Besides this CAC, there are other problems in ATM networks in which mathematical models can play an important role: routing problems, detection of user violations, allocation of link capacity.

This paper is concerned with the modelling of the arrival process in one node of the network. In Figure 1 the different parts of interest are shown. The arrival process a_k ($k = 0, 1, 2, \ldots$) is the aggregated number of cells that arrive per time unit. To be able to deal with fluctuations in the traffic flows, a buffer is needed, in which the cells have to wait before they can be served. The queue introduces cell delays and it is even possible that cells get lost because of buffer overflow. The allowed delay, delay variation and cell loss

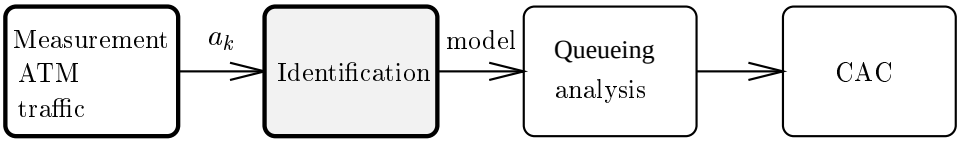

Figure 2: Different steps in traffic modelling. Our research is mainly concerned with the second step, the identification of a model that approximates the behaviour of the arrival process a_k $(k = 0, 1, 2, \ldots)$.

probability form part of the contracted QoS.
Traffic modelling consists of several steps (see Figure 2). First the arrival process a_k $(k = 0, 1, 2, \ldots)$ is measured. In the identification step a model that describes the process behaviour, is obtained. The model is then used in queueing analysis to get information on the state of the queues and finally the results are applied in the connection admission control.
The measurement based identification of a good model is the main subject of this paper.

3 The identification problem

In this section we consider the identification problem, i.e., the problem of finding a model for the arrival process, given the measurements a_k. In Section 3.1, we determine a suitable model class. The model (a Markov modulated Poisson process) is discussed in Section 3.2. Section 3.3 describes the identification approach, using this Markov modulated Poisson process. A more restricted model that makes the identification more efficient, is the circulant Markov modulated Poisson process. We discuss its structure and properties in Section 3.4 and the identification algorithm is given in Section 3.5. The two models are compared in Section 3.6.

3.1 Model choice

A multimedia input stream can be described by a stationary random process with measurable first order statistics (the probability distribution function p or cumulative distribution function F), second order statistics (the autocorrelation function R) and higher-order statistical functions.
The task of traffic identification is to construct stochastic models, the statistics of which match those of the data. This is difficult especially because

the models should fit into feasible analytical queueing solutions for network performance evaluation.
The research of Li and Hwang [16] has demonstrated that only the first order and second order statistics have a significant impact on queueing performance. Classic queueing theories have generally ignored the second order statistics entirely by assuming that the cell inter-arrival times form a renewal process. By using a Markov Modulated Poisson Process with first order and second order statistics matching those of the measured traffic, the correlation in multimedia traffic can be captured and exploited.

3.2 The Markov modulated Poisson process

3.2.1 Model description

A Markov modulated Poisson process (MMPP) of order N consists of an N-state Markov chain in which each state i ($i = 1, 2, \ldots, N$) represents a Poisson process with rate γ_i. In other words, a Markov modulated Poisson process is a Poisson process, the rate of which is changed (modulated) according to the Markov chain. In this paper, we only consider discrete time Markov chains.
Let S_k ($k = 0, 1, 2, \ldots$) be the stochastic process that denotes the state moving at discrete time instants between a finite number of states 1, 2, ..., N. The trasition probabilities depend only on the previous state, so that S_k ($k = 0, 1, 2, \ldots$) is a Markov process.
Let the matrix $P = [p_{ij}]$ be the one-step state transition matrix:

$$p_{ij} = Pr\{S_{k+1} = j \mid S_k = i\} \ .$$

If $\theta(k)$ is the column vector with i-th entry $Pr\{S_k = i\}$, then

$$\theta(k+1)^T = \theta(k)^T P \ .$$

Throughout this paper, we will deal exclusively with the steady state case, where the state distribution θ is time independent:

$$\theta^T = \theta^T P \ .$$

Matrix P is a $N \times N$ stochastic matrix, that is $P \in \mathbb{R}^{N \times N}$, $p_{ij} \geq 0$ and $\sum_{j=1}^{N} p_{ij} = 1$. Vector θ is the left eigenvector of the stochastic matrix P corresponding to eigenvalue 1. For θ holds:

$$\theta_i \ \geq \ 0, \text{ for } i = 1, 2, \ldots, N \tag{1}$$

$$\sum_{i=1}^{N} \theta_i \ = \ 1 \ . \tag{2}$$

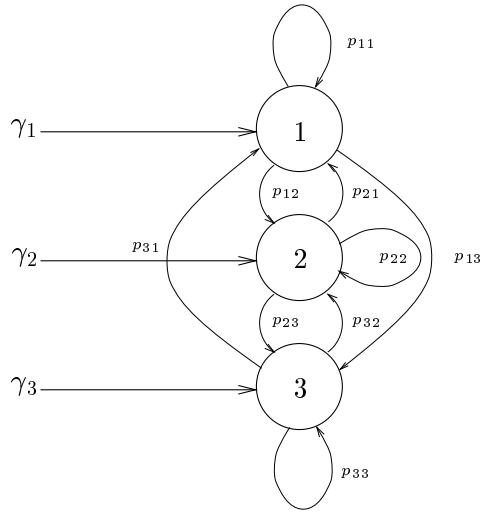

Figure 3: A third order MMPP. Each state i represents a Poisson process, characterised by the parameter γ_i. The transition probabilities p_{ij} between the states are indicated by arrows.

In addition to the (hidden) state process, there is the observation or measurement process a_k $(k = 0, 1, 2, \ldots)$, with values in the set $\{0, 1, \ldots, M\}$. This observation process a_k $(k = 0, 1, 2, \ldots)$ depends probabilistically on S_k $(k = 0, 1, 2, \ldots)$ via an $(M + 1) \times N$ matrix C which represents the Poisson distributions:

$$\begin{aligned} c_{ij} &= Pr\{a_k = i \mid S_k = j\} \\ &= \frac{e^{-\gamma_j}\gamma_j^i}{i!} \ . \end{aligned}$$

Let η be the vector with n-th entry $Pr\{a_k = n\}$, then

$$\eta = C\theta \ . \tag{3}$$

The Poisson parameters are arranged in a column vector $\gamma = [\gamma_i] \in \mathbb{R}^N$, where $\gamma_i \geq 0$ is the arrival rate of the Poisson process associated with state i. Throughout this paper we will use the notation $[x_i]$ to denote a column vector with elements x_i:

$$[x_i] = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \end{bmatrix} \ .$$

At time step k, the MMPP (P, γ) will emit a_k arrivals generated by the Poisson process with the arrival rate γ_i if the state of the Markov chain is i. Figure 3 shows an MMPP of order 3.
The aim of traffic identification is to find P and γ from arrivals a_k, so that they form an MMPP with first order and second order statistics matching those of the measured traffic a_k. P and γ will then be used in queueing analysis.

3.2.2 The first and second order statistics of arrivals

As already mentioned, only the first and second order statistics of arrivals a_k are considered. The first order statistics are described by the cumulative distribution function F, defined as

$$F(l) = Pr\{a_k \leq l\}.$$

For an MMPP, we have

$$F(l) = \sum_{i=1}^{N} \theta_i e^{-\gamma_i} \sum_{j=0}^{l} \frac{\gamma_i^j}{j!} \tag{4}$$

$$= \sum_{i=1}^{N} \theta_i F_{\gamma_i}(l) \tag{5}$$

where $F_{\gamma_i}(l)$ is the cumulative distribution function of the Poisson process with parameter γ_i.
The second order statistics of a_k are characterised by the autocorrelation function R, defined as

$$R(n) = E\{a_k \cdot a_{k+n}\} \ . \tag{6}$$

The autocorrelation of an MMPP is given by:

$$\begin{cases} R(0) = \theta^T(\Gamma + \Gamma^2)e \\ R(n) = \theta^T \Gamma P^n \Gamma e, \text{ for all } n \in \mathbb{N}_0 \ , \end{cases} \tag{7}$$

where

$$\Gamma = \mathtt{diag}(\gamma) = \begin{bmatrix} \gamma_1 & 0 & \cdots & 0 \\ 0 & \gamma_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & \gamma_N \end{bmatrix}$$

and

$$e = [\ 1\ 1\ \ldots\ 1\ ]^T$$

(see [24] for proof).
To compute the autocorrelation, the eigenvalue decomposition of P is used:

$$R(n) = \theta^T \Gamma V \Lambda^n V^{-1} \Gamma e, \text{ for all } n \in \mathbb{N}_0\ , \tag{8}$$

where $P = V\Lambda V^{-1}$, $\Lambda = \texttt{diag}([\lambda_i])$ (we assume that P is diagonalisable, which it is generically), where λ_i $(i = 1, 2, \ldots, N)$ are the eigenvalues of P.

3.3 Identification of the MMPP by unconstrained optimisation

This identification approach is mainly due to Yi and De Moor [28]. We will give an overview of the algorithm we have used (see Figure 4).
Assume F_d and R_d are respectively the cumulative distribution function and

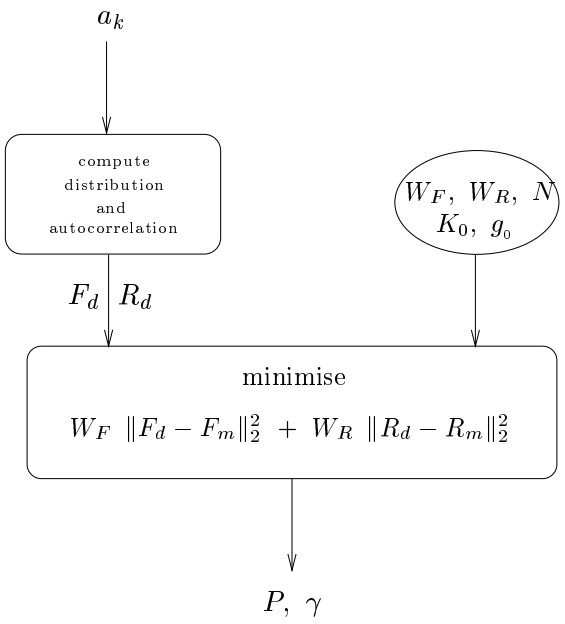

Figure 4: Identification of the MMPP by unconstrained optimisation. The user has to decide on the system order N and on the weighting factors W_F and W_R. In addition, a good initial estimate for g and K has to be chosen. The input of the user is indicated by the oval.

the autocorrelation function of the data a_k. The parameters P and γ of an MMPP (P, γ) must be found, so that the statistics of the model, F_m and R_m, match F_d and R_d as close as possible.
The problem can be cast as the following optimisation problem:

$$\min_{P \in \mathbb{P}_N,\ \gamma \geq 0} W_F\ \|F_d - F_m\|_2^2\ +\ W_R\ \|R_d - R_m\|_2^2\ , \tag{9}$$

where $\mathbb{P}_N$ stands for the set of all stochastic $N \times N$ matrices, $\|\cdot\|_2$ stands for the 2-norm and W_F and W_R are two weighting factors. One has to integrate both statistics in one cost function because both the autocorrelation function and the cumulative distribution function depend on P (see (5) and (7)). The cumulative distribution function depends indirectly on P, via θ, the left eigenvector of P corresponding to eigenvalue 1.
The following parameterisation

$$\Phi :\ \mathbb{R}^{N \times (N-1)} \to \mathbb{P} :\ K \to P = \Phi(K)$$

was used to reformulate the optimisation problem as an unconstrained optimisation problem:

$$P = \Phi(K) \quad = \quad \begin{bmatrix} \dfrac{k_{11}^2}{1 + \sum_{j=1}^{N-1} k_{1j}^2} & \cdots & \dfrac{k_{1(N-1)}^2}{1 + \sum_{j=1}^{N-1} k_{1j}^2} & \dfrac{1}{1 + \sum_{j=1}^{N-1} k_{1j}^2} \\ \vdots & & \vdots & \vdots \\ \dfrac{k_{N1}^2}{1 + \sum_{j=1}^{N-1} k_{Nj}^2} & \cdots & \dfrac{k_{N(N-1)}^2}{1 + \sum_{j=1}^{N-1} k_{Nj}^2} & \dfrac{1}{1 + \sum_{j=1}^{N-1} k_{Nj}^2} \end{bmatrix}$$

$$\gamma_i \quad = \quad g_i^2,\ \text{for } i = 1, \ldots, N\ .$$

The matrix $K = [k_{ij}] \in \mathbb{R}^{N \times (N-1)}$ and the vector $g = [g_i] \in \mathbb{R}^{N \times 1}$ are now the new variables in the optimisation algorithm. Define the enlarged matrix A

$$A = [K|g] \in \mathbb{R}^{N \times N}\ ,$$

which is the concatenation of K and g. It is obvious that the constrained nonlinear optimisation problem of (9) over the stochastic matrix set $\mathbb{P}_N$

for P and the nonnegative orthant for γ is changed to an unconstrained non-convex optimisation problem over $\mathbb{R}^{N\times N}$:

$$\min_{A\in\mathbb{R}^{N\times N}} W_F \ \|F_d - F_m\|_2^2 \ + \ W_R \ \|R_d - R_m\|_2^2 \ . \tag{10}$$

Some drawbacks of this method are:

- the user has to choose the model order N
- the user has to determine good weighting factors W_F and W_R
- the optimisation algorithm needs initial estimates for K and for g, called K_0 and g_0
- the method is not very efficient (see Section 3.6).

3.4 The circulant Markov modulated Poisson process

3.4.1 Model structure

A circulant stochastic matrix is a stochastic matrix which has the following structure:

$$Q = \begin{bmatrix} q_1 & q_2 & \cdots & q_N \\ q_N & q_1 & \cdots & q_{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ q_2 & q_3 & \cdots & q_1 \end{bmatrix} .$$

Circulant matrices have some important properties because of their special structure (see e.g. [5]). In the following paragraph the most important properties for the ATM traffic identification problem are discussed.
If one uses a circulant stochastic matrix Q instead of an ordinary stochastic matrix P, the MMPP becomes more restricted and is then called a circulant Markov modulated Poisson process or CMPP [16, 17].

3.4.2 Properties of circulant stochastic matrices

The left eigenvector, corresponding to eigenvalue 1, θ, of a circulant stochastic matrix Q of order N is equal to $\left[\frac{1}{N} \ \frac{1}{N} \ \ldots \ \frac{1}{N}\right]^T$ because circulant stochastic matrices are double stochastic matrices. This vector is independent of the elements of Q.
A circulant stochastic matrix Q is always diagonalisable. Its eigenvalue decomposition can be computed very efficiently. If $Q = V\Lambda V^{-1}$ is the eigenvalue decomposition of the circulant stochastic matrix Q, then we have:

- The eigenvectors V_j are independent of Q. Moreover, V is a Fourier matrix, defined as:

$$\begin{aligned} v_{ij} &= \frac{1}{\sqrt{N}} \exp\Big(\frac{-2\pi\sqrt{-1}}{N}(i-1)(j-1)\Big) \\ V^{-1} &= V^H \ , \end{aligned} \tag{11}$$

  where H denotes the conjugate transposition. The product of a Fourier matrix V and a vector x is equal to the discrete Fourier transform of that vector x. The discrete Fourier transform can be computed using fast algorithms (fast Fourier transform, `fft`). This property is very advantageous in the computation of the autocorrelation of a CMPP (see (14)).

- The eigenvalues λ_i can be obtained as follows:

$$\begin{aligned} \lambda = [\lambda_i] &= \sqrt{N} V \, [q_1 \ q_2 \ \ldots \ q_N]^T \\ &= \texttt{fft}\Big([q_1 \ q_2 \ \ldots \ q_N]^T\Big) \ . \end{aligned} \tag{12}$$

These two properties considerably reduce the computational cost during the computation of the model statistics. The cumulative distribution function becomes

$$F_m(l) = \frac{1}{N} \sum_{j=1}^{N} \exp(-\gamma_j) \cdot \sum_{i=0}^{l} \frac{1}{i!} \ \gamma_j^i \ . \tag{13}$$

It does not depend on the transition matrix Q in contrast to the cumulative distribution function of an MMPP (4), which is dependent on the transition matrix P via the left eigenvector corresponding to eigenvalue 1, θ.
The autocorrelation function of a CMPP can be computed as follows:

$$\begin{aligned} R_m(n) &= \theta^T \Gamma Q^n \Gamma e \\ &= \frac{1}{N} \gamma^T V \Lambda^n V^H \gamma \\ &= \frac{1}{N^2} \texttt{fft}(\gamma^T) \texttt{diag}\,([\lambda_i^n])\, \overline{\texttt{fft}(\gamma)} \ (i = 1, 2, \ldots, N), \end{aligned} \tag{14}$$

where λ is obtained as described in (12) and $\overline{\texttt{fft}(\gamma)}$ denotes the complex conjugate of $\texttt{fft}(\gamma)$. Only two fast Fourier transforms are needed to obtain the autocorrelation function R_m of a CMPP: the fast Fourier transform of the first row of Q yields the eigenvalues of Q, and the product of the Fourier matrix V and the vector γ is also replaced by a fast Fourier transform. This makes the computation of the autocorrelation function much more efficient for CMPPs than for MMPPs (see also Section 3.6).

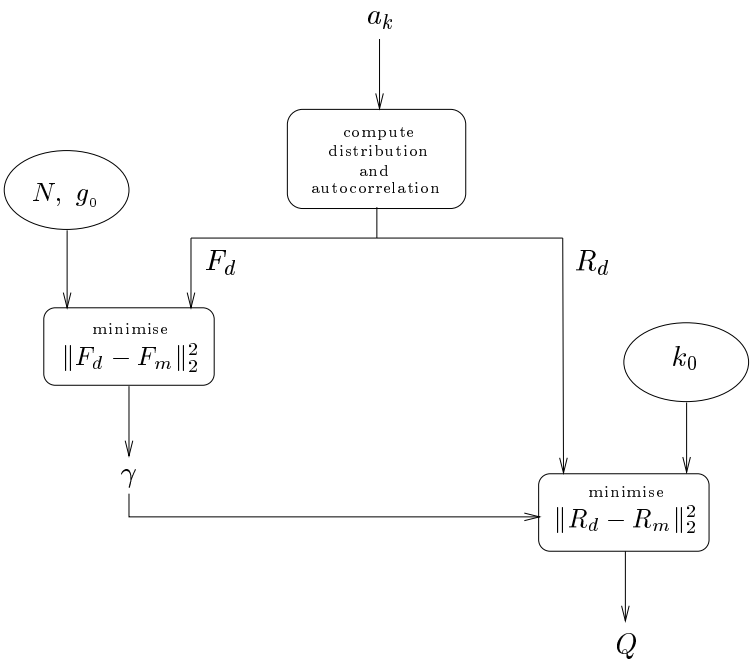

Figure 5: Identification of the CMPP by unconstrained optimisation. The user has to give initial estimates for g and k and the model order N has to be chosen. The optimisation of cumulative distribution and autocorrelation function are decoupled.

3.5 Identification of the CMPP by unconstrained optimisation

In Figure 5 the flow diagram of the identification algorithm is shown. The ovals represent the input of the user. The main difference between this method and the method described in Section 3.3 is that by using circulant stochastic matrices, the optimisation of autocorrelation and cumulative distribution can be decoupled.
The parameterisation of γ and Q to get an unconstrained optimisation prob-

lem is similar to the parameterisation of the ordinary MMPP:

$$\begin{aligned}
\gamma_i &= g_i^2, \text{ for } i = 1, \ldots, N \\
q_i &= \frac{k_i^2}{1 + \sum_{j=1}^{N-1} k_j^2}, \text{ for } i = 1, \ldots, N-1 \\
q_N &= \frac{1}{1 + \sum_{j=1}^{N-1} k_j^2} .
\end{aligned}$$

The variables used in the optimisation algorithms are $g = [g_i] \in \mathbb{R}^{N \times 1}$ and $k = [k_i] \in \mathbb{R}^{N \times 1}$.
Because θ does not depend on Q ($\theta = \left[\frac{1}{N} \ \frac{1}{N} \ \ldots \ \frac{1}{N}\right]^T$), it is possible to decouple the optimisation of the cumulative distribution function and the autocorrelation function. First, the γ that minimises $\|F_d - F_m\|_2^2$ is computed and then this optimal γ is used in the optimisation of $\|R_d - R_m\|_2^2$, which yields Q.
The decoupling of cumulative distribution and autocorrelation, which was not taken into account by Yi and De Moor [28], eliminates the need of weighting matrices. Yet, this method still needs a certain contribution of the user: the model order N and an initial estimate of g and k.

3.6 Comparison of MMPP and CMPP identification

In Table 1 a comparison is made between the traffic identification methods described in Section 3.3 and Section 3.5. The MMPP considered is an N_M-th order model with transition matrix P and Poisson rate vector γ. The CMPP has transition matrix Q and rate vector γ and is of order N_C.
As we have mentioned before, the cumulative distribution function of a CMPP does not depend on the transition matrix Q, whereas the cumulative distribution function of an MMPP is dependent on P. This is why the optimisation can be decoupled for a CMPP and not for an MMPP.
For an MMPP the computational cost per function evaluation is of the order N_M^3, whereas for a CMPP this is of the order $N_C \log(N_C)$ because of the efficient eigenvalue decomposition, cf. equations (11), (12) and (14). This is a very important numerical advantage of CMPPs.
The disadvantage of a CMPP is that the order must be chosen much higher

than for an MMPP (see Section 6.1 and Section 6.3). This is not only a drawback for the identification: in order to make the numerical queueing analysis feasible with CMPP input models, the order N must also be kept small. Li and Hwang [17] mention that the computational complexity of the existing queueing techniques is at least proportional to N^3, referring to [27].

4 Characterisation of the first order statistics without a priori knowledge of the model order

In this section we discuss a new efficient method for the identification of the first order statistics.
In Section 4.1 we give an overview of the model order estimation problem and we discuss how it is solved in the literature.
In Section 4.2 we describe our new method for the characterisation of the first order statistics. It determines the components of a mixture distribution. The order N of the MMPP, the rate vector γ and the state distribution vector θ are obtained by solving a nonnegative least squares problem.
In Section 4.3 we compare the model order estimation method of Yi and

	MMPP(P,γ) order $= N_M$	CMPP(Q,γ) order $= N_C$
F_m depends on	$P,\ \gamma$	γ
R_m depends on	$P,\ \gamma$	$Q,\ \gamma$
computational cost per function evaluation	$\mathcal{O}(N_M^3)$	$\mathcal{O}(N_C \log(N_C))$
number of parameters	N_M^2	$2N_C - 1$

Table 1: Comparison of traffic identification by unconstrained optimisation using MMPPs and CMPPs. The use of CMPPs reduces the numerical complexity and makes it possible to decouple the optimisation of the cumulative distribution function and the autocorrelation function, but for a CMPP the order must be chosen higher than for an MMPP.

De Moor [28], based on the autocorrelation function, with our new method. This is done by an example.

4.1 Order estimation of a Markov modulated Poisson process

In each method mentioned so far the user has to estimate the system order N.
All identification methods described in [20] assume the system order to be known. In [17] the order of the CMPPs is taken 'sufficiently large'. Li and Hwang [17] found that taking $N_C = 100$ is sufficient in most cases of their study.
We think it is more suitable to incorporate the determination of the model order in the identification algorithm. In this way the user is freed form the problem of choosing the model order. In contrast to Li and Hwang [17] we do not want to overestimate the model order of the MMPP. By computing the minimal sufficient model order we make the computational complexity as small as possible.
Yi and De Moor [28] base their approach for the order estimation on the second order statistics only. They determine the order of the associated stochastic realization problem. The autocorrelation function of an MMPP

$$R(n) = \theta^T \Gamma P^n \Gamma e, \text{ for } n > 0$$

(see (7)) can be considered as the covariance sequence Λ_n $(n = 1, 2, \ldots)$ of a linear time invariant stochastic system, characterised by the triplet of matrices (A, G, C), where

$$\Lambda_n = CA^{n-1}G, \text{ for } n = 1, 2, \ldots \tag{15}$$

For the meaning of these matrices in the context of stochastic realization we refer to [18, 25, 26]. By taking, e.g., $C = \theta^T \Gamma$ and $G = P\Gamma e$ the sequence of autocorrelation samples $R(1), R(2), \ldots$ can be considered as the covariance sequence of the linear system characterised by the triplet $(P, P\Gamma e, \theta^T \Gamma)$. The stochastic identification problem comprises more than the order estimation: the matrices P, C and G are estimated and used in the computation of the state space description of the linear system (see [1, 4, 11, 18, 25, 26]). The complete identification of the MMPP using this approach is difficult because of the model restrictions of the MMPP (see [23]). Yi and De Moor [28] only try to find the order of the MMPP, using the relevant parts of the stochastic realization algorithms.

The autocorrelation function R (see equation (6) for the definition) must be estimated from the data sequence a_k. Such estimates could be based on the ergodic limit

$$R(n) = \lim_{K\to\infty} \frac{1}{K+1} \sum_{k=0}^{K} a_k a_{k+n} \ , \tag{16}$$

which holds almost surely [6]. In practice we deal with truncated ergodic sums such as

$$R(n) \approx \frac{1}{K+1} \sum_{k=0}^{K} a_k a_{k+n}, \text{ for } n \leq L-K-1 \ , \tag{17}$$

where L is the total number of data points a_k $(k = 0, 1, 2, \ldots, L-1)$. If $K \leq L-1$ is sufficiently large, then these estimates will be good provided n is not too large. However, only a finite number of autocorrelation points

$$R(0), R(1), R(2), \ldots, R(\nu) \tag{18}$$

can be determined from the finite data sequence a_k $(k = 0, 1, 2, \ldots, L-1)$. In fact, we must have $\nu \ll L$.
Let H be a Hankel matrix formed from the autocorrelation sequence:

$$H = \begin{bmatrix} R(1) & R(2) & R(3) & \cdots & R(j) \\ R(2) & R(3) & R(4) & \cdots & R(j+1) \\ R(3) & R(4) & R(5) & \cdots & R(j+2) \\ \vdots & \vdots & \vdots & & \vdots \\ R(i) & R(i+1) & R(i+2) & \cdots & R(i+j-1) \end{bmatrix} \ , \tag{19}$$

where $i + j - 1 = \nu$ and the Hankel matrix H is chosen as close to square as possible by taking $|i - j| \leq 1$. Yi and De Moor [28] take the order of the MMPP equal to the rank of the Hankel matrix H. As the authors of [28] mention, this method only yields a lower bound for the order of the MMPP. There are several reasons for the underestimation of the model order using this method.

1. The minimal order of the linear stochastic model, characterised by (A, G, C) is not always equal to the rank of the Hankel matrix H. In order to have a positive real extension of the covariance sequence $R(1), R(2), \ldots R(\nu)$ the model order should be taken equal to the positive degree of the covariance sequence. The relation between the rank

of the Hankel matrix H (19), the algebraic degree r and the positive degree p of the covariance sequence is as follows:

$$\texttt{rank}(H) \leq r \leq p \ , \tag{20}$$

where each inequality could be strict. Only the first equality in (20) is a generic property [9]. It is shown in [8] that the positive degree has no generic value. In fact, there is no easy way to determine the positive degree [7].

2. A necessary condition for the realized triplet (A, G, C) to represent an MMPP is element-wise positivity of the matrices A, G and C. The model order should be sufficiently large in order to have a positive realization of the given autocorrelation sequence. Up to now the positive realization problem is unsolved. There exists a necessary and sufficient condition for the existence of a positive realization, but the characterisation of minimality is an unsolved question [22].

3. Besides this positivity constraint, an MMPP has additional restrictions compared with a positive system. This implies that an MMPP that matches the autocorrelation function of the data, could require a larger model order than a positive linear system that models the autocorrelation function.

Because of the underestimation of the model order, Yi and De Moor [28] propose to choose the order obtained by their method, as a first try. One could then increase the order until the value of the cost function (10) is small enough. This is of course very time consuming.
In the next section we propose a method to determine the order N (and γ and θ) of an MMPP starting from the first order statistics.

4.2 Determination of the components of a mixture distribution

The probability distribution of a Markov modulated Poisson process is the following discrete function:

$$\begin{aligned} p(a_k = l) &= Pr\{a_k = l\} \\ &= \sum_{i=1}^{N} Pr\{S_k = i\} \cdot Pr\{a_k = l \mid S_k = i\} \end{aligned}$$

$$= \sum_{i=1}^{N} \theta_i \cdot p_{\gamma_i}(a_k = l) \tag{21}$$

$$= \sum_{i=1}^{N} \theta_i \cdot \frac{e^{-\gamma_i} \gamma_i^l}{l!} \ , \tag{22}$$

where $p_{\gamma_i}(a_k = l)$ is the probability distribution function of the Poisson process with parameter γ_i. The probability distribution of an MMPP is a mixture (linear combination) of Poisson distributions. The order of the MMPP is equal to the number of components in this mixture distribution. In [21] several methods are described to determine the components of mixture distributions. Most of these methods assume a given number of components. A graphical method can be used to determine the number of components, provided they are well separated. Our method does not have this restriction.

Recall equation (3):

$$\eta = C\theta \ , \tag{23}$$

in which θ satisfies (1) and (2).
Equation (23) is (22) in matrix notation, where

$$\eta = [p(a_k = l)] \ , \ \text{for } l = 0, 1, 2, \ldots, M \ ,$$

and M is equal to the maximum number of arrivals of the given set of data:

$$M = \max(a_k) \ .$$

The j-th column of C, C_j, is the Poisson probability distribution function with parameter γ_j, truncated to $l = M$:

$$C_j = \left[\frac{e^{-\gamma_j} \gamma_j^l}{l!} \right] , \ \text{for } l = 0, 1, 2, \ldots, M \ .$$

The truncated distribution function η is computed as follows:

$$\eta(l) = \frac{1}{L} \sum_{k=0}^{L-1} \delta(a_k, l), \ \text{for } l = 0, 1, 2, \ldots, M \ ,$$

where $\delta(i, j)$ is the Kronecker delta and L is the number of data.
Assume that the states (i.e., the γ_i's) of the MMPP which models a given dataset, are given. The matrix C is then fixed and θ can be determined by solving the system of equations (23).

Without knowledge of the states or the number of states, the same approach yields N, γ and θ if in (23) C is replaced by an enlarged version $\overline{C}$. This matrix $\overline{C}$ has much more columns than C and the same number of rows. Each column $\overline{C_j}$ of $\overline{C}$ represents a *possible* state or *possible* Poisson distribution. In fact, the domain of possible γ_i's, e.g., $[0.001, M]$, is discretized with a step h in order to get a broad choice of states: $\gamma_j = \gamma_1 + (j-1)h$. The first Poisson parameter, γ_1, has to be chosen small enough in order to incorporate possible zero-valued a_k's. The computed distribution function η must be reconstructed as a nonnegative linear combination of the columns of $\overline{C}$. Therefore one solves $\eta = \overline{C}x$ in a nonnegative least squares sense:

$$\min_x \|\eta - \overline{C}x\|_2, \text{ subject to } x_i \geq 0 \ ,$$

which is a convex optimisation problem as it is a positive definite quadratic programming problem. Let x^* be the optimal value of x. The number of non-zero components of x^* gives the model order N. The indices of the non-zero components of x^* denote the γ_i's. The θ_i's are given by the values of the non-zero components of x^*.

By this new approach we are not only able to determine the model order, we also obtain the state parameters γ_i and the state distribution vector θ. The first order statistics are identified very efficiently and without any contribution of the user, because the matrix $\overline{C}$ can be pre-constructed, before one starts identifying. If the matrix $\overline{C}$ is formed over a large range of γ_i's, $\overline{C}$ can be adjusted by the program to a given dataset. This adjustment is done by taking a submatrix of $\overline{C}$. The columns that represent Poisson distributions $p_{\gamma_i}(a_k = l)$ (see (21)) with rate parameter $\gamma_i > \max(a_k)$, are left out of $\overline{C}$. The rows that correspond to samples of the probability distributions $p_{\gamma_i}(a_k = l)$ where $l > \max(a_k)$ are eliminated too.

Instead of dealing with the distribution of the data (η) and the Poisson distribution functions as columns of $\overline{C}$, a better alternative is to use the *cumulative* distribution function. If F_d is the cumulative distribution function of the data, then F_d must be reconstructed as a positive linear combination of cumulative distributions of Poisson processes $F_{\gamma_i}(l)$ (see (5)). An analogous derivation as for the probability distribution can be made for this situation. Consider the matrix $\overline{D}$, the columns of which are the cumulative Poisson distribution functions $F_{\gamma_i}(l)$, according to the discretized γ-domain. The j-th column of $\overline{D}$, $\overline{D_j}$, is the Poisson cumulative distribution function with rate γ_j, truncated to $l = M$:

$$\overline{D_j} = \left[e^{-\gamma_j} \sum_{m=0}^{l} \frac{\gamma_j^m}{m!} \right], \text{ for } l = 0, 1, 2, \ldots, M \ .$$

The nonnegative least squares problem is now:

$$\min_x \|F_d - \overline{D}x\|_2, \text{ subject to } x_i \geq 0 \ . \tag{24}$$

Let x^* be the optimal value of x, then again the following applies: the number of non-zero components of x^* gives the model order N. The indices of the non-zero components of x^* denote the γ_i's. The θ_i's are given by the values of the non-zero components of x^*.
The advantage of the cumulative distribution function over the probability distribution function is that it is less sensitive to noise because the cumulative distribution function is an integrated version of the probability distribution function. A second benefit is the possibility to sample the cumulative distribution function without much loss of information. By sampling the cumulative distribution function the number of rows in $\overline{D}$ and F_d decreases. As a consequence the computation of F_d can be accelerated. It is also possible to adjust the matrix $\overline{D}$ to the given set of data. This is explained in Figure 6.

4.3 Example

In this example we compare the order estimation method of Yi and De Moor [28], discussed in Section 4.1, with the one described in the previous section. The first method is based on the autocorrelation function of the

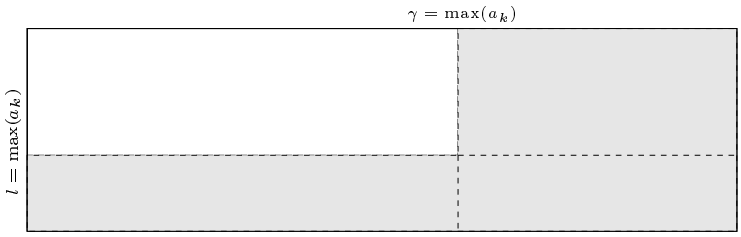

Figure 6: A representation of the adaptation of the original matrix $\overline{D}$ (the large rectangle) to the given set of data. The grey parts are left out because they do not contain valuable information. All the columns which represent cumulative Poisson distributions with parameter γ larger than the maximum value of the dataset are eliminated. The rows which correspond to samples of the cumulative distribution $F_m(l)$, where $l > \max(a_k)$ are also left out. The nonnegative least squares problem (24) is solved for the resulting smaller matrix (the upper left hand white part).

data, the second method is based on the cumulative distribution function. Although it is clear from Section 4.1 that the method of Yi and De Moor [28] might yield an underestimation of the model order, we want to stress another aspect, namely that it is important to consider the first order statistics in order to get a good estimation of the model order.
We will show that

1. It is not straightforward to determine the rank of the Hankel matrix (19), made up of the autocorrelation function of the data, estimated as in (17).

2. It is essential to take the first order statistics into account in order to obtain a good estimate of the model order.

3. Starting from the first order statistics of the data (the cumulative distribution function) we find the parameters of the MMPP that describe the first order statistics (N, γ and θ), the model order included, by solving a nonnegative least squares problem.

For this example the data are obtained by simulating a sixth order MMPP. The Poisson parameter vector γ is chosen in such a way that there is one non-observable mode in the linear system $(P, P\Gamma e, \theta^T\Gamma)$, which makes the triplet $(P, P\Gamma e, \theta^T\Gamma)$ non-minimal. As a consequence, the model order will be smaller than six, although we need six states in order to describe the first order statistics properly.
The simulated MMPP is characterised by the following transition matrix P and Poisson vector γ. The vector θ is the left eigenvector of P, corresponding to eigenvalue 1.

$$
\begin{aligned}
P &= \begin{bmatrix}
0.1804 & 0.0111 & 0.3652 & 0.2352 & 0.2081 & 0.0000 \\
0.2626 & 0.0059 & 0.0160 & 0.6325 & 0.0107 & 0.0722 \\
0.1111 & 0.1759 & 0.2290 & 0.2127 & 0.0646 & 0.2067 \\
0.2788 & 0.0273 & 0.0062 & 0.1602 & 0.4627 & 0.0648 \\
0.0780 & 0.5642 & 0.2235 & 0.0153 & 0.0708 & 0.0483 \\
0.2776 & 0.0233 & 0.1436 & 0.2304 & 0.1681 & 0.1570
\end{bmatrix} \qquad (25) \\
\gamma &= \begin{bmatrix} 10 & 25 & 70 & 120 & 46.0825 & 180 \end{bmatrix}^T \\
\theta &= \begin{bmatrix} 0.1925 & 0.1450 & 0.1650 & 0.2306 & 0.1858 & 0.0812 \end{bmatrix}^T .
\end{aligned}
$$

The obtained data sequence is 100000 time steps long[4].

4.3.1 Determination of the model order based on the autocorrelation function [28]

We will consider two cases:
First we construct the Hankel matrix with the truncated autocorrelation estimates (see equation (17)), based on the finite dataset. We call this Hankel matrix H_1 and the autocorrelation function used for the construction of H_1 is R_1. In the second case we consider the Hankel matrix constructed with the ideal autocorrelation function, obtained as described in (7), using the system characteristics P, γ and θ. This autocorrelation function, R_2, would be the result of (16), the asymptotic case where an infinite dataset is available. The second (ideal) Hankel matrix is called H_2.
For the computation of the autocorrelation sequence $R_1(0), R_1(1), \ldots R_1(99)$ from the dataset (see (17)), we have taken the following parameter values

$$\begin{aligned} L &= 100000 \\ K &= 99900 \\ \nu &= 99\ . \end{aligned} \tag{26}$$

The estimated autocorrelation function R_1 and the ideal autocorrelation function R_2 are compared in Figure 7. The estimation approximates the ideal autocorrelation function, but they are not equal.

Case 1: MMPP order estimation, based on the estimated autocorrelation sequence R_1.
The matrix H_1 of dimension 50×50 is not rank-deficient. The numerical rank of H_1 is determined by looking at its singular values. Table 2 and the left picture of Figure 8 show the largest 14 singular values of H_1. There is no clear gap between the 6th and the 7th singular value. Looking at these singular values, we would rather decide for order 1 or 3.

Case 2: MMPP order estimation, based on the ideal autocorrelation function R_2.

[4]This specific dataset is available from DAISY (Database for the Identification of Systems) at URL http://www.esat.kuleuven.ac.be/sista/daisy/ [10]. At this site we propose to collect all datasets that are used in publications on system identification and parameter estimation, so that the results in these papers really become reproducible, which at present is more often not the case

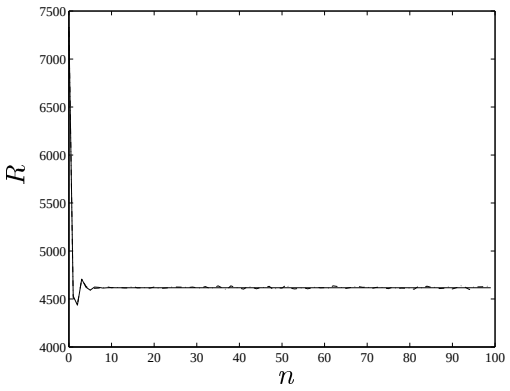

Figure 7: Comparison of the estimated autocorrelation function R_1 (dash-dotted line) to the ideal autocorrelation function R_2 (full line). The estimated autocorrelation function R_1 is computed as in equation (17) with parameters (26) and the ideal one is obtained by computation (7) using the system matrices (25)

σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7
$2.3092 \cdot 10^5$	257.0885	225.6831	145.8831	143.1864	121.7564	120.2108
σ_8	σ_9	σ_{10}	σ_{11}	σ_{12}	σ_{13}	σ_{14}
116.2721	107.4374	95.2405	94.6723	93.4165	92.9683	90.3395

Table 2: The first 14 singular values of H_1. There is no clear gap between the 6th and the 7th singular value. It is difficult to decide which order should be chosen.

Now we consider the ideal case, where the autocorrelation function is computed using the system matrices P, θ and γ from (25). This is the asymptotic case for an infinitely large set of data. The computation is done according to (14). The largest singular values of the obtained Hankel matrix H_2 are shown in Table 3 and in the right picture of Figure 8.
From the 6th one on, all singular values are practically equal to 0. The model order should be chosen equal to 5. However, the system is of order 6. In Figure 9 the probability distribution function of the data sequence is shown. It is clear that it is a mixture of at least 6 components. It is impossible to model the first order statistics with a 5th order MMPP. This

σ_1	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7
$2.3086 \cdot 10^5$	229.0966	216.9958	2.5769	2.1636	0.0000	0.0000

Table 3: The first 7 singular values of H_2. It is now clear that the order of the model should be 5. However, this is not the order of the original MMPP.

indicates that it is more natural to base the model order estimation on the first order statistics instead of on the second order statistics.
The reason for the undermodelling by looking at the rank of the Hankel matrix is the non-minimality of the system $(P, P\Gamma e, \theta^T \Gamma)$. The Popov-Belevitch-Hautus test for observability (see, e.g., [13]) shows that the sixth mode is not observable:
Let V be the matrix whose columns V_i are the eigenvectors of P and let W be the matrix consisting of the left eigenvectors of P. The eigenvalues λ_i, corresponding to the eigenvectors V_i are the elements of the main diagonal

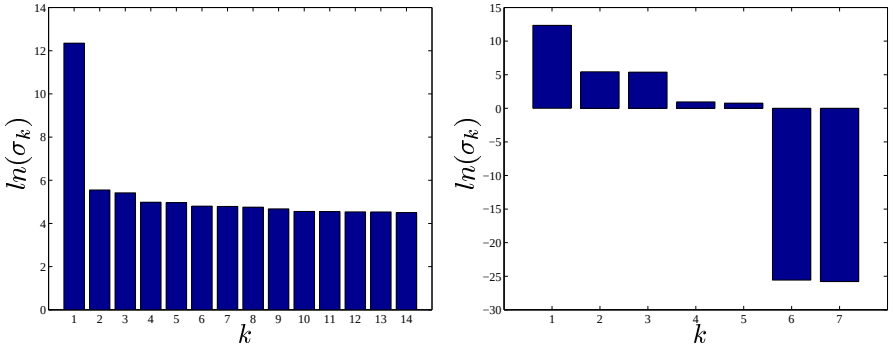

Figure 8: In the left picture the largest 14 singular values of H_1 are shown. The Hankel matrix $H_1 \in \mathbb{R}^{50\times 50}$ is constructed with the autocorrelation samples from data obtained by simulating the MMPP described in (25). The logarithm of the singular values is plotted against their index. The right picture shows the largest 7 singular values of H_2. This Hankel matrix $H_2 \in \mathbb{R}^{50\times 50}$ is obtained by computing the autocorrelation of the MMPP (25) according to (7). This autocorrelation would be obtained if one would have an infinitely long data sequence. The scale in the right picture is also logarithmic. From the left picture it is difficult to decide which order should be chosen. The right picture indicates clearly model order 5.

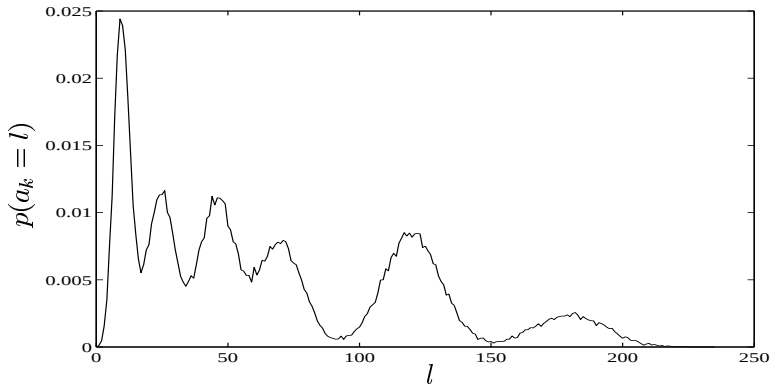

Figure 9: The probability distribution function of the data, obtained by simulating the CMPP(P, γ). One can distinguish 6 components.

of Λ:

$$\Lambda = \texttt{diag}(\begin{bmatrix} 1 \\ -0.1692 + 0.4825j \\ -0.1692 - 0.4825j \\ -0.0366 + 0.1464j \\ -0.0366 - 0.1464j \\ 0.2149 \end{bmatrix}) \ , \tag{27}$$

where j is equal to $\sqrt{-1}$.
The PBH test for controllability yields:

$$(P\Gamma e)^T W = \begin{bmatrix} -160.3624 \\ 23.3319 - 15.6230j \\ 23.3319 + 15.6230j \\ 9.5875 - 4.9857j \\ 9.5875 + 4.9857j \\ 16.7156 \end{bmatrix}^T \tag{28}$$

The PBH test for observability yields:

$$\theta^T \Gamma V = \begin{bmatrix} 27.7407 \\ -4.8977 - 3.4080j \\ -4.8977 + 3.4080j \\ 6.0204 - 9.9090j \\ 6.0204 + 9.9090j \\ 0 \end{bmatrix}^T \tag{29}$$

From (28) and (29) we conclude that mode 6 is not observable, but it is controllable. All the other modes are controllable and observable. Indeed, the

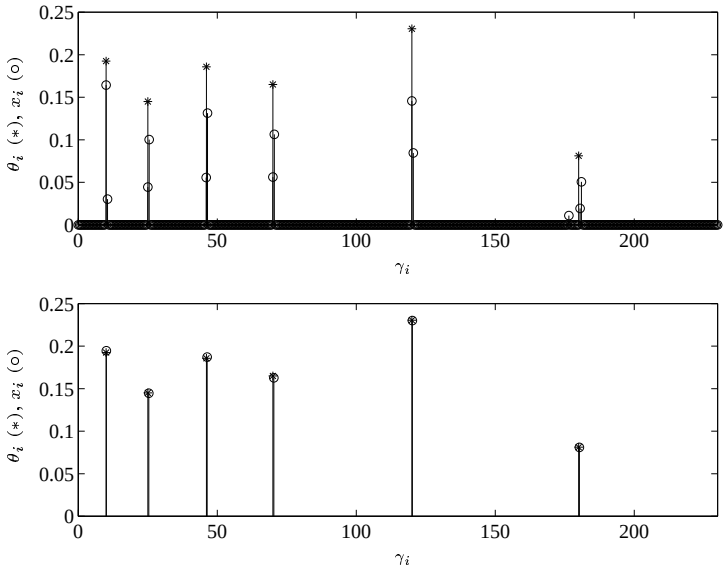

Figure 10: Estimation of the model order and of the parameters γ and θ based on the cumulative distribution of the simulation data. In both figures the θ_i's of the original MMPP are indicated by a star ('$*$') above the corresponding γ_i's. The identified parameter values are indicated by a circle ('$\circ$'). The top figure shows the result of the algorithm described in Section 4.2. In the bottom figure the model is reduced by eliminating the zero-states and by taking together very similar states. It is already clear from the top figure that the model order should be 6. We do not only recover the system order, but we also have a very good estimate for γ and θ.

rank of the controllability matrix is equal to 6, whereas the rank of the observability matrix equals 5. As a consequence the rank of H_2 is theoretically equal to 5.

4.3.2 Determination of the order based on the cumulative distribution function

We now use the same dataset to determine the order as described in Section 4.2. Instead of the autocorrelation function, the cumulative distribution

function $F_d(l)$ $(l = 0, 10, 20, \ldots, 230)$ of the data is computed:

$$F_d(l) = \frac{1}{L} \sum_{j=0}^{l} \sum_{k=1}^{L} \delta(a_k, j) \ , \tag{30}$$

where $\delta(i, j)$ is the Kronecker delta and L is the number of data. The matrix $\overline{D}$ that consists of possible Poisson cumulative distributions, is constructed with discretisation step $h = 0.5$. The cumulative distribution functions, which are the columns of $\overline{D}$, are discretised with step $\Delta l = 10$. The nonnegative least squares problem (24) is solved, where $\overline{D} \in \mathbb{R}^{24 \times 471}$. The result of this algorithm is shown in the top picture of Figure 10. The stars indicate the parameters γ and θ of the original 6th order MMPP. The circles indicate the results of (24), the x_i values. There are 13 non-zero valued x_i's. However, after reduction of the model by taking neighbouring states together, the resulting model is of order 6. The parameters of the reduced model are shown in the bottom picture of Figure 10. We do not only recover the system order, but we also have a very good estimate for γ and θ:

$$\begin{aligned} \gamma_m &= \begin{bmatrix} 10.0778 & 25.3465 & 70.3272 & 120.1836 & 46.3512 & 180.2739 \end{bmatrix}^T \\ \theta_m &= \begin{bmatrix} 0.1945 & 0.1447 & 0.1627 & 0.2301 & 0.1871 & 0.0809 \end{bmatrix}^T \ . \end{aligned}$$

5 Identification of an MMPP without a priori knowledge of the model order

The fact that we already have a first order characterisation (N, γ, θ) from (24) accelerates the identification of the MMPP model significantly. Only the optimisation of the autocorrelation function must be carried out. In this optimisation the state parameters γ remain constant. The state distribution vector θ, which is known from (24), should be the left eigenvector of P corresponding to eigenvalue 1. The transition matrix P however, must be found by optimisation. To ensure that the resulting matrix P has θ (or a vector close to θ) as its left eigenvector corresponding to eigenvalue 1, the following unconstrained optimisation problem is solved:

$$\min_{K \in \mathbb{R}^{N \times (N-1)}} W_v \ \|\theta - v\|_2 + W_R \ \|R_d - R_m\|_2 \ ,$$

where v is the left eigenvector of $P = \Phi(K)$ corresponding to eigenvalue 1 and W_v and W_R are weighting factors. The essence of this method is that we keep the first order parameters from (24) and impose θ to be the left

eigenvector of P corresponding to the eigenvalue 1, where P together with the constant γ models the autocorrelation function of the data. The resulting method is summarised in Figure 11.
The fact that we have a good model for the first order statistics and that we keep this model during the optimisation is not only important for the matching of the first order statistics. The performance of the autocorrelation optimisation also depends on the first order model (γ, θ). This is because $R(0)$ and $R(\infty)$ of an MMPP only depend on the first order parameters γ

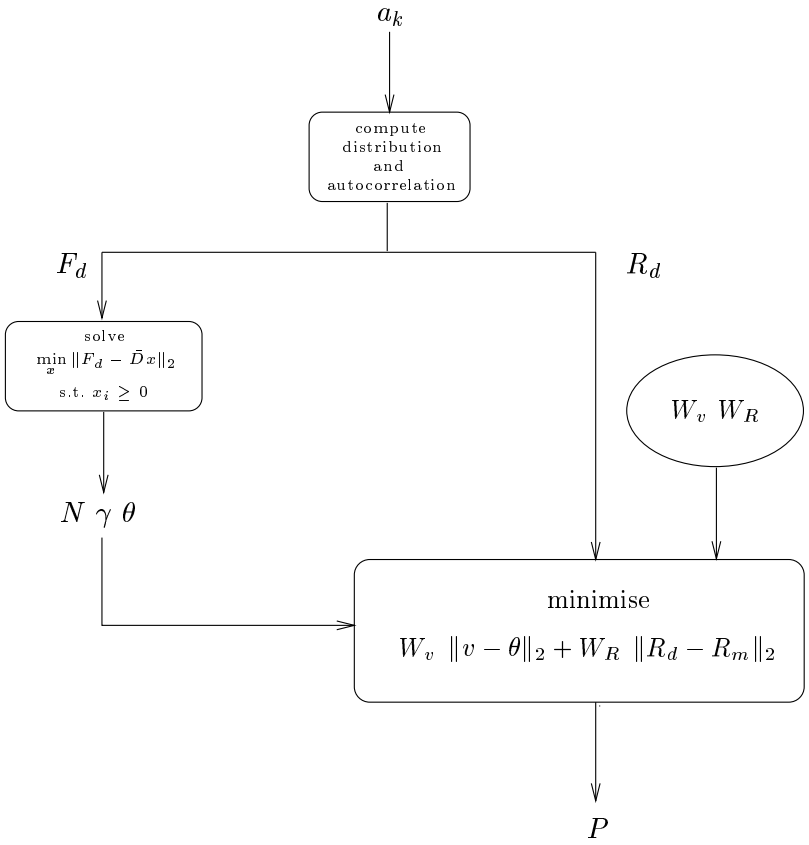

Figure 11: Identification of the MMPP by unconstrained optimisation without a priori knowledge of the model order N. The contribution of the user is reduced to providing the weighting factors W_R and W_v.

and θ:

$$\begin{aligned} R(0) &= \sum_{i=1}^{N} \theta_i(\gamma_i + \gamma_i^2) \\ R(\infty) &= \left(\sum_{i=1}^{N} \theta_i \gamma_i \right)^2 \\ &= (E\{a_k\})^2 \ . \end{aligned} \tag{31}$$

If there is an error on γ and/or θ, it is impossible to find a transition matrix P which models the autocorrelation function ($R(0)$ and $R(\infty)$ included) in a satisfactory way. It is important to have a good estimate of $R(\infty)$ because this value is equal to the square of the mean cell arrival rate (see (31)).
In Section 7 a comparison is made between the time domain identification method for MMPPs, proposed by Yi and De Moor [28] and this new method.

6 Identification of a CMPP without a priori knowledge of the model order

In Section 4 was shown how to obtain the order, the states and the state distribution of an MMPP from measurements a_k. In this section we discuss how this information is used in the identification of CMPPs.
We show how the MMPP parameters N_M, γ_M and θ_M which characterise the first order statistics, are translated into the corresponding CMPP parameters. In Section 6.1 a theoretical way to translate the parameters is described. This is illustrated with an example in Section 6.2. A more practical translation is described in Section 6.3.
The complete identification method for CMPPs in which the characterisation of the first order statistics is used, is given in Section 6.4.

6.1 Translation of the first order MMPP parameters into the first order CMPP parameters

Assume that the data a_k can be modelled by an MMPP. The order N_M, the states γ_M and the state distribution vector θ_M of the MMPP model are then obtained by solving (24). To conform these results to the state distribution vector θ_C of a CMPP, $\left[\begin{array}{cccc} \frac{1}{N_C} & \frac{1}{N_C} & \dots & \frac{1}{N_C} \end{array} \right]^T$, a translation of N_M, γ_M

and θ_M into N_C, γ_C and θ_C is necessary. This is done by taking replicas of each MMPP state. Each state contributes as many times as needed to match the first order statistics by θ_C and the obtained γ_C:

$$\gamma_C = \left[\underbrace{\gamma_{M_1}\ \gamma_{M_1}\ \cdots\ \gamma_{M_1}}_{t_1 \text{ times}}\ \underbrace{\gamma_{M_2}\ \cdots\ \gamma_{M_2}}_{t_2 \text{ times}}\ \cdots\ \underbrace{\gamma_{M_{N_M}}\ \cdots\ \gamma_{M_{N_M}}}_{t_{N_M} \text{ times}} \right] .$$

The number of copies of MMPP state i is denoted by t_i $(i = 1, 2, \ldots, N_M)$. Consequently, the order of the CMPP is equal to:

$$N_C = t_1 + t_2 + \ldots + t_{N_M} . \tag{32}$$

The value of each t_i is determined as follows:

- Assume that all elements of θ_M are rational numbers $\frac{n_i}{d_i}$, where n_i and d_i are integers. If not, approximate the irrational number θ_{M_i} by a rational one.
- Compute the least common multiple of all the denominators d_i:

$$\delta = \texttt{lcm}(d_1, d_2, \ldots, d_{N_M}) .$$

- Define $n_i^* = \frac{n_i \delta}{d_i}$, for $i = 1, 2, \ldots, N_M$.
- Compute the greatest common divisor of all n_i^*'s:

$$\nu = \texttt{gcd}(n_1^*, n_2^*, \ldots, n_{N_M}^*) .$$

- The number of times that one MMPP state is repeated in the CMPP is then equal to:

$$t_i = \frac{\theta_{M_j} \delta}{\nu} . \tag{33}$$

From (32), (33) and (2) we obtain:

$$N_C = \frac{\delta}{\nu} .$$

An example will make this more clear.

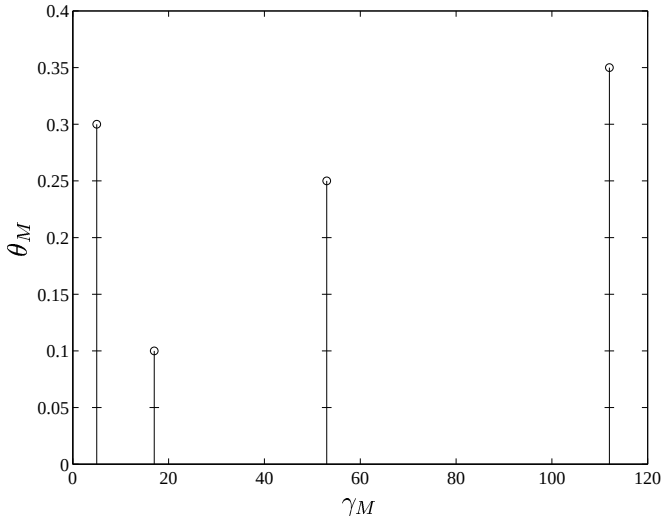

Figure 12: The states and the state distribution for the MMPP described in (34) and (35). The circles indicate the θ_{M_i} values. The abscissas of the circles are the corresponding γ_{M_i} values. The little horizontal lines divide all θ_{M_i} in equal parts of size 0.05. The number of parts of each θ_{M_i} is equal to t_i. The total number of parts is equal to the order of the CMPP, $N_C = 20$.

6.2 Example

Assume one has a 4th order MMPP model in which the states are determined by

$$\gamma_M = \begin{bmatrix} 5 & 17 & 53 & 112 \end{bmatrix}^T \tag{34}$$

and the distribution of the states is equal to

$$\theta_M = \begin{bmatrix} \frac{3}{10} & \frac{1}{10} & \frac{1}{4} & \frac{7}{20} \end{bmatrix}^T . \tag{35}$$

These two model features are shown in Figure 12.
In order to describe this first order characterisation by a CMPP, each θ_{M_i} must be written as a multiple of one 'unit' $\frac{1}{N_C}$. This unit is found by dividing each θ_{M_i} into as few as possible equal parts, as shown in Figure 12 by the little horizontal lines. The size of one such a part is then equal to $\frac{1}{N_C}$, i.e., $\frac{\nu}{\delta}$. In this example $\delta = 20$ and $\nu = 1$, consequently, $N_C = 20$. This is also clear from Figure 12: the size of one part is equal to 0.05 . The number of

parts per γ_{M_i} is equal to t_i:

$$\left[\begin{array}{cccc} t_1 & t_2 & t_3 & t_4 \end{array}\right] = \left[\begin{array}{cccc} 6 & 2 & 5 & 7 \end{array}\right]$$

and the total number of pieces is equal to the model order of the CMPP, N_C. The resulting first order parameters of the CMPP model are:

$$\begin{aligned} \gamma_C &= [5\ \ 5\ \ 5\ \ 5\ \ 5\ \ 5\ \ 17\ \ 17\ \ 53\ \ 53\ \ 53\ \ 53\ \ 53\ \ 112\ \ 112\ \ 112\ \ 112\ \ 112\ \ 112\ \ 112]^T \\ \theta_C &= [\tfrac{1}{20}\ \tfrac{1}{20}]^T \ . \end{aligned}$$

6.3 A more practical version for the translation

If the translation of the first order MMPP parameters into the first order CMPP parameters is carried out as explained in Section 6.1, the resulting order of the CMPP can become very large. A more practical solution is to impose a certain order N_C and divide each θ_{M_i} in pieces of size $\dfrac{1}{N_C}$. To obtain the t_i's, i.e., the number of times each state is duplicated, a rounding off is performed:

$$\begin{aligned} t_i &= \texttt{floor}\left(\frac{\theta_{M_i}}{\frac{1}{N_C}}\right), \text{ for } i = 1, 2, \ldots, N_M \\ &= \texttt{floor}(\theta_{M_i} N_C), \text{ for } i = 1, 2, \ldots, N_M \ , \end{aligned}$$

where $\texttt{floor}(\theta_{M_i} N_C)$ is the largest integer that is less than or equal to $\theta_{M_i} N_C$. If $\sum_{i=1}^{N_M} t_i < N_C$ (by round off errors) we add units to the states for which the error was the largest.

Values that are very suitable for N_C are powers of 2. This is because in the optimisation step `fft`'s are used to evaluate the autocorrelation of the CMPPs (see (14)). If N_C is a power of 2, a fast radix-2 fast Fourier transform algorithm is used. As mentioned before, Hwang and Li [17] found that taking $N_C = 100$ is sufficient in most cases of their study. Therefore we take $N_C = 128 = 2^7$.

Let ν_t be the greatest common divisor of all computed t_i's ($i = 1, 2, \ldots, N_M$):

$$\nu_t = \texttt{gcd}(t_1, t_2, \ldots, t_{N_M}) \ .$$

If ν_t is not equal to 1, the CMPP model can be reduced: divide the obtained t_i's by ν_t and the model order becomes

$$N_C^* = \frac{N_C}{\nu_t} \ .$$

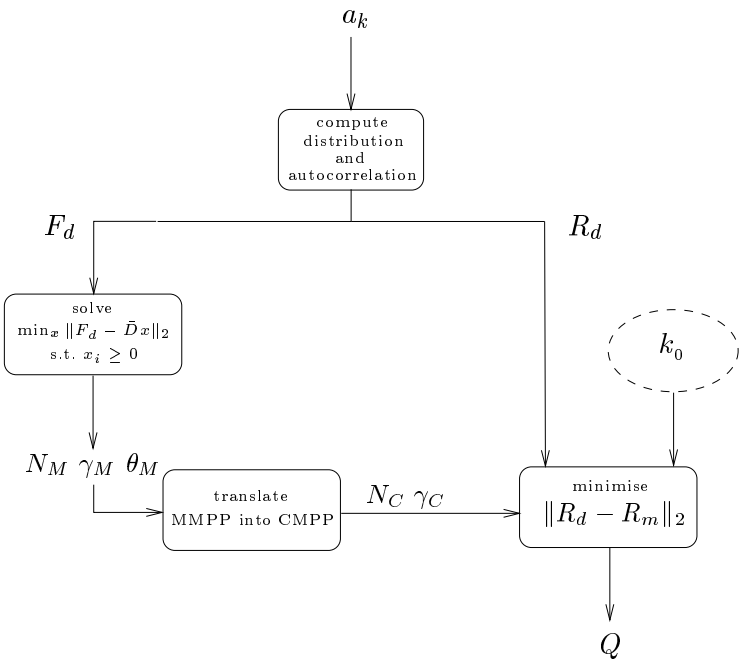

Figure 13: Identification of the CMPP where the cumulative distribution function is matched by solving the nonnegative least squares problem (24) followed by a translation of MMPP parameters to CMPP parameters. The autocorrelation function is matched by unconstrained optimisation. If the program decides for k_0 by taking a random vector, the contribution of the user is avoided completely.

6.4 The complete identification method for CMPPs

In Figure 13 the complete identification method for CMPPs is shown. The first order statistics are matched in two steps: first, the MMPP parameters which describe the first order statistics (N_M, γ_M and θ_M) are found by solving the nonnegative least squares problem (24) and then these MMPP parameters are translated into the CMPP parameters N_C, γ_C and θ_C. The circulant stochastic matrix Q is obtained by unconstrained minimisation of $\|R_d - R_m\|_2$ in which γ_C is kept constant. The optimisation is executed efficiently by using fast Fourier transforms in the evaluation of the cost function (see (14)). The user has to provide the optimisation algorithm with vector k_0. The choice of k_0 is easily incorporated in the program by taking a random vector for k_0. As a consequence, the contribution of the user could be

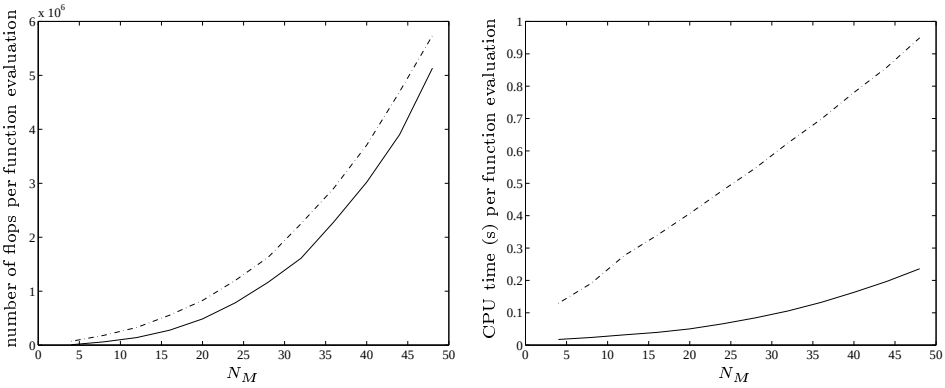

Figure 14: The number of floating point operations (left) and CPU time (right) in seconds per function evaluation in the optimisation step of two time domain MMPP identification methods. The solid line shows the result for the new method described in Section 5. The dash-dotted line represents the identification method of Yi and De Moor [28] (see Section 3.3).

eliminated completely.

7 Comparison between the different time domain identification methods

In this section we first compare the MMPP identification method by Yi and De Moor [28] with the method described in Section 5.
In Section 7.2 we make the comparison for the CMPP identification methods. In both Section 7.1 and Section 7.2 we first compare the optimisation step in the algorithms to be compared and then the operations preceding the optimisation in our new methods of sections 5 and 6 are discussed.
Finally both model structures, MMPP and CMPP, are compared in Section 7.3.

7.1 Identification of an MMPP

7.1.1 The optimisation step

In Figure 14 we compare the optimisation step of Yi *et al.*'s method described in Section 3.3 and the optimisation step of our new MMPP identification method without a priori knowledge of the model order (Section 5). The left

picture shows the number of floating point operations ('flops') per function evaluation and the right picture shows the computation time per function evaluation. Both properties are plotted in function of the MMPP model order. The solid line represents the identification method described in Section 5 whereas the dash-dotted line is the MMPP identification method of Yi and De Moor [28]. For both methods one needs $\mathcal{O}(N_M^3)$ operations per function evaluation (left picture of Figure 14). This is because the eigenvalue decomposition which is carried out to compute the autocorrelation function of the MMPP model (see (8)) is dominating. However, the CPU time per function evaluation is strongly reduced by eliminating the optimisation of the cumulative distribution function (see right picture of Figure 14). The cost function in (9):

$$W_F \ \|F_d - F_m\|_2 \ + \ W_R \ \|R_d - R_m\|_2 \ , \tag{36}$$

is replaced by

$$W_v \ \|\theta - v\|_2 + W_R \ \|R_d - R_m\|_2 \tag{37}$$

in our new method. To compute the first term of (37) no extra information is needed. The left eigenvector of P corresponding to eigenvalue 1, v, is known from the eigenvalue decomposition in the computation of R_m. On the other hand, the computation of the first term of (36) takes a lot of time. For each Poisson component of the model the cumulative distribution function must be computed.

7.1.2 Operations preceding the optimisation

Yi and De Moor [28] presented a method in which all MMPP parameters are determined in one step, the optimisation algorithm (see, e.g., Figure 4). No preceding operations are needed. The most important disadvantage of this approach is already discussed in 7.1.1: the computation of the cumulative distribution function of the model in each function evaluation, which is time consuming. Another disadvantage is the fact that the user has to decide for the model order. This is avoided in our new approach: the model order is determined automatically.
However, we need to solve the nonnegative least squares problem (24). In Figure 15 the CPU time needed for the steps preceding the optimisation in our new MMPP identification algorithm (see, e.g., Figure 11) is plotted. This cost includes the adaptation of matrix $\bar{D}$ in (24) to the set of data, computing the solution of the nonnegative least squares problem (24) and the reduction of the model by taking similar states together. The CPU

time is plotted in function of the maximum value of the data, $\max(a_k)$. This maximum value determines the size of matrix $\bar{D}$ in (24). The figure is obtained for a matrix $\bar{D}$ with $\gamma = \begin{bmatrix} 0.0001 & 0.5 & 1 & 1.5 & \dots & 1000 \end{bmatrix}$ (discretisation step $h = 0.5$) and sampled cumulative distribution functions with discretisation step $\Delta l = 10$. The matrix $\bar{D}$ has an original size of 101 rows and 2001 columns. The adaptation to $\max(a_k)$ reduces this to $\texttt{floor}\left(\frac{\max(a_k)}{10}\right)$ rows and $(2 \cdot \max(a_k) + 1)$ columns, where $\texttt{floor}(x)$ is the largest integer less than or equal to the real number x. In Figure 6 we show how a sub-matrix is taken from the original matrix $\bar{D}$, according to the maximum value of the dataset.
The CPU time needed to perform the operations preceding the optimisation is small compared to what the optimisation algorithm needs. If the optimisation algorithm is stopped after, e.g., 1000 function evaluations, it still takes 17.5229 to 236.0496 seconds, depending on the model order (see Figure 14). To perform the nonnegative least squares problem, the adaptation and the reduction one needs less than 4 seconds.

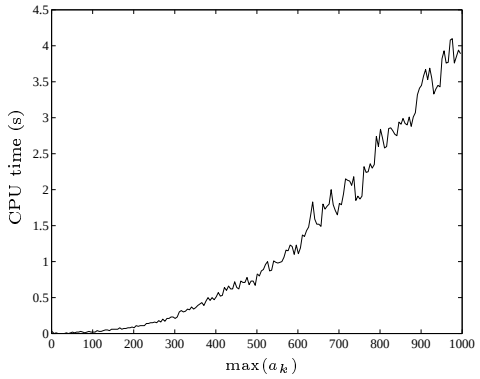

Figure 15: The CPU time in seconds needed to process the operations preceding the optimisation step of the MMPP identification algorithm described in Section 5. It includes the adaptation of matrix $\bar{D}$ to the set of data, the solution of the nonnegative least squares problem (24) and the reduction of the model. The CPU time is plotted in function of the maximum value of the dataset. This value determines the size of the sub-matrix of $\bar{D}$ (see Figure 6).

7.2 Identification of a CMPP

7.2.1 The optimisation step

Figure 16 shows the results of the cost calculations for the optimisation step of the CMPP identification methods of Section 3.5 (dash-dotted line) and of

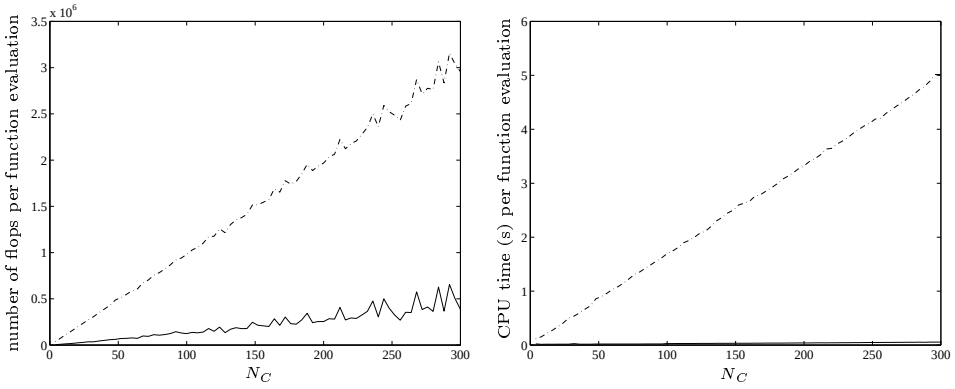

Figure 16: The number of floating point operations (left) and CPU time in seconds (right) per function evaluation in the optimisation step of two time domain CMPP identification methods. The solid line shows the result for the new method described in Section 6. The dash-dotted line represents the identification method of Yi and De Moor [28] (see Section 3.5).

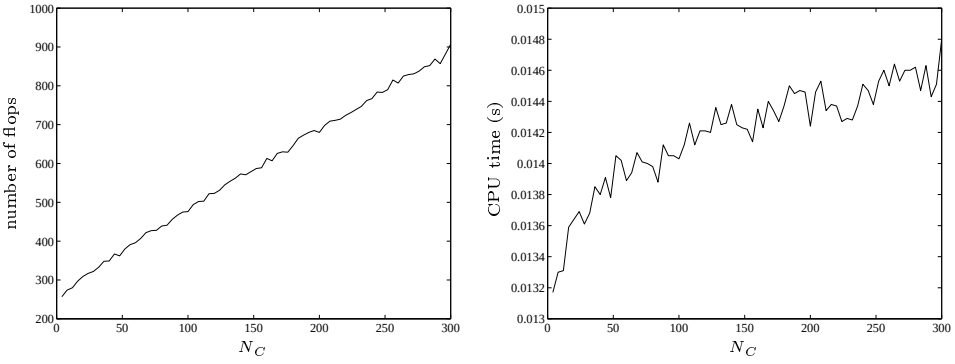

Figure 17: The number of floating point operations (left) and CPU time in seconds (right) needed to translate the first order parameters γ_M and θ_M of an MMPP to γ_C and θ_C of a CMPP in function of the CMPP order, N_C. The practical translation described in 6.3 was used.

Section 6 (solid line). The left picture represents the number of floating point operations (flops) per function evaluation. This exhibits a $\mathcal{O}(N_C \log(N_C))$ path for both methods. The dominating eigenvalue decomposition is computed with a fast `fft` algorithm (see equation 14). In the right picture it is again clear that the elimination of the computation of the cumulative distribution function reduces the CPU time per function evaluation.

7.2.2 Operations preceding the optimisation

In contrast with the method of Yi and De Moor [28] where the CMPP model is found in one (optimisation) step, we have two stages preliminary to the optimisation (see, e.g., Figure 13):

- The nonnegative least squares problem to obtain the first order MMPP parameters γ_M and θ_M and the model order N_M.
- The translation of the first order MMPP parameters to the first order CMPP parameters γ_C, θ_C.

The properties of the nonnegative least squares problem and related problems concerning their computational cost are discussed in Section 7.1.2. In Figure 17 the number of operations and the CPU time needed to translate the first order MMPP parameters to CMPP parameters are plotted in function of the CMPP model order.
The time needed to perform both steps preceding the optimisation is only a fraction of the time needed to optimise the autocorrelation. Their contribution to the global cost is negligible.

7.3 Comparison between MMPP and CMPP identification

In Figure 18 we compare the number of floating point operations per function evaluation of the optimisation step of all the mentioned time domain identification methods. The solid lines represent the algorithms of sections 5 and 6. The dash-dotted lines show the results of the algorithms presented by Yi and De Moor [28]. CMPP models are clearly numerically more advantageous than MMPPs. It is practically impossible to use high order MMPP models to optimise the autocorrelation function.
In Table 4 we compare the optimisation steps of the identification methods with regard to their computational cost per function evaluation and the number of parameters to be optimised. For our new method the computational cost is of the same order as for the method of Yi and De Moor [28]. The number of parameters is different. In our approach the Poisson rate

	MMPP(P,γ) order $= N_M$		CMPP(Q,γ) order $= N_C$	
computational cost per function evaluation	$\mathcal{O}(N_M^3)$		$\mathcal{O}(N_C \log(N_C))$	
	old [28]	new (Section 5)	old [28]	new (Section 6)
number of parameters	N_M^2	$N_M(N_M-1)$	$2N_C-1$	N_C-1

Table 4: Comparison of traffic identification by unconstrained optimisation using MMPPs and CMPPs. The new methods of sections 5 and 6 have less parameters to be optimised than the algorithms described in [28]. This is because γ remains constant during the optimisation.

vector γ remains constant during the optimisation. This means a reduction of the number of parameters to be optimised with N. As a consequence the number of function evaluations needed for the computation of the gradients

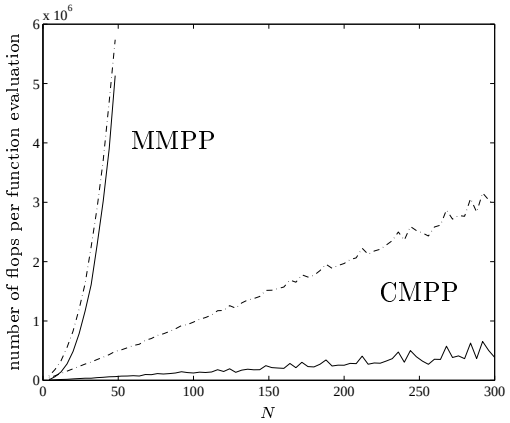

Figure 18: The number of floating point operations per function evaluation in the optimisation step of four time domain MMPP identification methods. The solid lines are the results for the new method described in Section 5 for the MMPPs and Section 6 for the CMPPs. The dash-dotted lines represent the identification methods of Yi and De Moor [28] (see Section 3.3).

in the optimisation algorithm decreases.

8 Conclusions

In this report we have presented two new time domain identification algorithms for ATM network traffic. One algorithm identifies the Markov modulated Poisson process (MMPP), the other algorithm identifies the circulant Markov modulated Poisson process (CMPP). Only the first order and second order statistics are matched, respectively the cumulative distribution function and the autocorrelation function. Both algorithms are based on the approach of Yi and De Moor [28], who use a one-step unconstrained optimisation algorithm.
Our new approach consists of two steps. The first order statistics are identified very efficiently by solving a nonnegative least squares problem. As a consequence the optimisation step needs less computation time. Another advantage of our approach is the fact that the order of the MMPP is determined automatically during the estimation of the first order parameters.

Acknowledgements

This work was supported by the **Flemish Government** (Administration of Science and Innovation (Concerted Research Action MIPS: *Model-based Information Processing Systems*, Bilateral International Collaboration: *Modelling and Identification of nonlinear systems*, IWT-Eureka SINOPSYS: *Model-based structural monitoring using in-operation system identification*), FWO-Vlaanderen: *Analysis and design of matrix algorithms for adaptive signal processing, system identification and control, based on concepts from continuous time system theory and differential geometry*, *Numerical algorithms for subspace system identification: Extension towards specific applications*, FWO-Onderzoeksgemeenschappen: *Identification and Control of Complex Systems*, *Advanced Numerical Methods for Mathematical Modelling*); **Belgian Federal Government**(Interuniversity Attraction Pole IUAP IV/02: *Modelling, Identification, Simulation and Control of Complex Systems*, Interuniversity Attraction Pole IUAP IV/24: IMechS: Intelligent Mechatronic Systems); **European Commission:** (Human Capital and Mobility: *SIMONET: System Identification and Modelling Network*, *SCIENCE-ERNSI: European Research Network for System Identification.*)

References

[1] H. Akaike. *Markovian Representation of Stochastic Processes by Canonical Variables.* SIAM Journal on Control, vol. 13, 1975, pp. 162-173.

[2] B.D.O. Anderson. *New Developments in the Theory of Positive Systems.* In C.I. Byrnes, B.N. Datta, D.S. Gilliam and C.F. Martin (Eds), Systems & Control in the 21st Century, Proceedings of the Conference on Mathematical Theory of Networks and Systems, Saint Louis, Missouri, USA, June 1996, pp. 17-36.

[3] B.D.O. Anderson, M. Deistler, L. Farina and L. Benvenuti. *Nonnegative Realization of a Linear System with Nonnegative Impulse Response.* IEEE Transactions on Circuits and Systems – I: Funcamental Theory and Applications, vol. 43, no. 2, February 1996, pp. 134-142.

[4] M. Aoki. *State Space Modeling of Time Series.* Springer Verlag, Berlin, 1987.

[5] A. Berman and R.J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences.* Classics in Applied Mathematics series, vol. 9. SIAM, Philadelphia, 1994.

[6] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods.* Springer-Verlag, New York, 1993.

[7] C.I. Byrnes and A. Lindquist. *On the Partial Stochastic Realization problem.* IEEE Transactions on Automatic Control, vol. 42, no. 8, August 1997, pp. 1049-1070.

[8] C.I. Byrnes and A. Lindquist. *The Stability and Instability of Partial Realizations.* Systems & Control Letters, vol. 2, 1982, pp. 99-105.

[9] A. Dahlén, A. Lindquist and J. Mari. *Experimental Evidence Showing that Stochastic Subspace Identification Methods May Fail.* Technical Report TRITA/MAT-97-OS10, Dept. of Optimization and System Theory, Royal Institute of Technology, Stockholm, 1997.

[10] B. De Moor, P. De Gersem, B. De Schutter and W. Favoreel. *DAISY: A Database for Identification of Systems.* Journal A (Benelux publication of the Belgian Federation of Automatic Control), Special Issue on CACSD, vol. 38, no. 3, pp. 4-5, September 1997.

[11] P. Faurre. *Stochastic Realization Algorithms.* In R. Mehra and D. Lainiotis (Eds), System Identification: Advances and Case Studies. Academic Press, NY, 1976.

[12] Internet address: *http://town.hall.org/archives/pub/ita/html/contrib/lbl-pkt.html*, lbl-pkt-4.

[13] T. Kailath. *Linear Systems.* Prentice-Hall, 1980.

[14] R.E. Kalman. *Realization of Covariance Sequences.* Proceedings of the Toeplitz Memorial Conference, Tel Aviv, Israel, 1981.

[15] S.Q. Li and C.L. Hwang. *Queue Response to Input Correlation Functions: Discrete Spectral Analysis.* IEEE/ACM Transactions on Networking, vol. 1, no. 5, October 1993, pp. 522-533.

[16] S.Q. Li and C.L. Hwang. *Queue Response to Input Correlation Functions: Continuous Spectral Analysis.* IEEE/ACM Transactions on Networking, vol. 1, no. 6, December 1993, pp. 678-692.

[17] S.Q. Li. and C.L. Hwang. *On the Convergence of Traffic Measurement and Queueing Analysis: A Statistical-Matching and Queueing (SMAQ) Tool.* IEEE/ACM Transactions on Networking, vol. 5, no. 1, Feb. 1997, pp. 95-110.

[18] A. Lindquist and G. Picci. *Canonical Correlation Analysis, Approximate Covariance Extension, and Identification of Stationary Time Series.* Automatica, vol. 32, no. 5, 1996, pp. 709-733.

[19] opzoeken of vragen aan PVO!! (Kalman-ontbinding)

[20] T. Rydén. *Parameter Estimation for Markov Modulated Poisson Processes.* Stochastic Models, no. 10, 1994, pp. 795-829.

[21] D.M. Titterington, A.F.M. Smith and U.E. Makov. *Statistical Analysis of Finite Mixture Distributions.* Wiley series in probability and mathematical statistics. Applied probability and statistics. Wiley, Chichester, 1985.

[22] J.M. van den Hof. *Realizations of Positive Linear Systems.* Linear Algebra and its Applications, no. 256, pp. 287-308.

[23] T. Van Gestel. *Identification of ATM traffic.* Masters's thesis (in Dutch), Katholieke Universiteit Leuven, ESAT-SISTA, 1997.

[24] P. Van Mieghem. *The Mapping of Traffic Parameters into a MMP(N).* Technical report, Alcatel Corporate Research Center, Antwerpen, Belgium, 1995.

[25] P. Van Overschee and B. De Moor. *Subspace Algorithms for the Stochastic Identification Problem.* Automatica, vol. 29, no. 3, 1993, pp. 649-660.

[26] P. Van Overschee and B. De Moor. *Subspace Identification for Linear Systems. Theory, implementation, applications.* Kluwer Academic Publishers, Boston/London/Dordrecht, 1996.

[27] J. Ye and S.Q. Li. *Folding Algorithm: a Computational Method for Finite QBD Processes with Level-Dependent Transitions.* IEEE Transactions on Communications, vol. 42, no. 2, February 1994, pp. 625-639.

[28] C. Yi and B. De Moor. *Traffic Identification of ATM Networks with Optimization Algorithms.* Proceedings of the 35th IEEE Conference on Decision and Control, Kobe, Japan, December 1996, pp. 277-282.