Parameter estimation of Markov modulated Fluid Arrival Processes[☆]

Salah Al-Deen Almousa[a], Gábor Horváth[a], Miklós Telek[a,b,*]

[a] *Department of Networked Systems and Services, Budapest University of Technology and Economics, Budapest, Hungary*
[b] *ELKH-BME Information Systems Research Group, Budapest, Hungary*

Abstract

Markov modulated discrete arrival processes have a wide literature, including parameter estimation methods based on expectation-maximization (EM). In this paper, we investigate the adaptation of these EM based methods to Markov modulated fluid arrival processes (MMFAP), and conclude that only the generator matrix of the modulating Markov chain of MMFAPs can be approximated by EM based method. For the rest of the parameters, the fluid rates and the fluid variances, we investigate the efficiency of numerical likelihood maximization.

To reduce the computational complexity of the likelihood computation, we accelerate the numerical inverse Laplace transformation step of the procedure with function fitting.

Keywords: Markov modulated fluid arrival processes, expectation-maximization method, parameter estimation.

1. Introduction

Markovian queueing systems with discrete customers are widely used in stochastic modeling. Markovian Arrival Processes (MAPs, [1]), that are able to characterize a wide class of point processes, play important role in these systems. The properties of MAPs have been studied exhaustively, using queueing models involving MAPs nowadays has become common, queueing networks with MAP traffic have also been investigated. Several methods exist to create a MAP approximating real, empirical data. Some of them aim to match statistical quantities like marginal moments, joint moments, auto-correlation, etc. [2, 3]. An other approach to create MAPs from measurement data is based on likelihood maximization, which is often performed by Expectation-Maximization

[☆]This work is partially supported by the OTKA K-138208 project and the Artificial Intelligence National Laboratory Programme of Hungary.

[*]Corresponding author

Email addresses: almousa@hit.bme.hu (Salah Al-Deen Almousa), ghorvath@hit.bme.hu (Gábor Horváth), telek@hit.bme.hu (Miklós Telek)

(EM) [4]. Several EM-based fitting methods have been published for MAPs, based on randomization [5], based on special structures [6, 7], methods that support batch arrivals [8] and those that are able to work with group data [9].

In many systems, the workload can be represented easier with continuous, fluid-like models rather than discrete demands [10]. Basic Markovian fluid models have been introduced and analyzed in [11, 12], later on several model variants appeared and were investigated. Despite their practical relevance, the "ecosystem" around fluid models is far less complete than their discrete counterparts. In particular, fitting methods for Markov modulated fluid arrival processes (MMFAP) are available only to some very restricted cases like on-off models, motivated by telecommunication applications. To the best of our knowledge, fitting methods for the general class of MMFAPs based on likelihood maximization has not been investigated in the past. At first glance adapting the methods available for MAPs might seem feasible, since fluid models can be treated as a limit of a discrete model generating infinitesimally small fluid drops. In fact, adapting the algorithms for MAPs to MMFAPs is not straightforward at all, fitting MMFAPs is a qualitatively different problem.

In this paper we consider the likelihood maximization based fitting of MMFAPs. The input data is assumed to be given as a list of pairs, where each pair represents the duration of the measurement interval and the amount of fluid arrived during the measurement interval. The equivalent problem among discrete models is discussed e.g., in [9], where an EM algorithm is presented to fit MAPs based on group data. We show that this algorithm is not appropriate for MMFAPs, as it can not change the fluid rate and variance parameters. To overcome the difficulties, we introduce a hybrid algorithm, where the generator matrix of the modulating Markov chain of MMFAPs is approximated by an EM based method, and the fluid rates and the fluid variances are determined by numerical optimization of the likelihood function.

We also consider the numerical issues associated with introduced procedure and propose flexible numerical approximations which can be tuned to obtain the required accuracy – computational complexity trade-off applying a function fitting approach whose accuracy depends on the number of computed points.

This paper is an extended version of [13], in which we already recognized the inability of the EM method to approximate the fluid rate and variance parameters. In the current work, we propose an optimization procedure to fit those parameters as well.

The rest of the paper is organized as follows. Section 2 introduces the mathematical model and the parameter estimation problem. The next section discusses the applicability of the EM method for MMFAPs and shows that the rate and variance parameters cannot be optimized. A direct optimization approach for the rate and variance parameters is introduced and a combined fitting method is proposed in Section 4. Some implementation details associated with the proposed fitting method are provided in Section 5. Finally, Section 6 presents numerical experiments about the properties of the proposed method, and Section 7 concludes the paper.

2. Problem definition

2.1. Fluid arrival process

The fluid arrival process $\{\mathcal{Z}(t) = \{\mathcal{J}(t), \mathcal{X}(t)\}, t > 0\}$ consists of an irreducible background Markov chain $\{\mathcal{J}(t), t > 0\}$, with state space $\mathcal{S} = \{1, 2, \ldots, S\}$, which modulates the arrival process of the fluid $\{\mathcal{X}(t), t > 0\}$. When the Markov chain stays in state i for a Δ long interval a normal distributed amount of fluid arrives with mean $r_i\Delta$ and variance $\sigma_i^2\Delta$, that is, when $\mathcal{J}(\tau) = i, \forall \tau \in (t, t + \Delta)$

$$\frac{d}{dx} Pr(\mathcal{X}(t + \Delta) - \mathcal{X}(t) < x) = \mathcal{N}(r_i\Delta, \sigma_i^2\Delta, x), \tag{1}$$

where $\mathcal{N}(\mu, \sigma^2, x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ is the Gaussian density function. We note that our proposed analysis approach allows negative fluid rates as well. Since the normal distribution has infinite support also in case of strictly positive fluid rates Section 5.3 discusses a numerical approach to handle negative fluid samples.

The generator matrix and the initial probability vector of the N-state background continuous time Markov chain (CTMC) are $\mathbf{Q}$ and $\underline{\alpha}$, and the diagonal matrix of the fluid rates and the fluid variances are given by matrix $\mathbf{R}$ with $\mathbf{R}_{i,j} = \delta_{i,j} r_i$, and matrix $\mathbf{\Sigma}$ with $\mathbf{\Sigma}_{i,j} = \delta_{i,j}\sigma_i^2$, where $\delta_{i,j}$ denotes the Kronecker delta.

Assuming $\mathcal{X}(0) = 0$, the amount of fluid arriving in the $(0, t)$ interval is $\mathcal{X}(t)$, with density matrix defined by

$$[\mathbf{N}(t, x)]_{i,j} = \frac{\partial}{\partial x} \Pr\left(\mathcal{X}(t) < x, \mathcal{J}(t) = j | \mathcal{J}(0) = i\right) \tag{2}$$

The double sided Laplace transform of this quantity regarding the amount of fluid arrived can be expressed as [14]

$$\mathbf{N}^*(t, v) = \int_{x=-\infty}^{\infty} e^{-xv} \mathbf{N}(t, x) dx = e^{(\mathbf{Q} - v\mathbf{R} + v^2\mathbf{\Sigma}/2)t}. \tag{3}$$

The stationary distribution of the CTMC is denoted by vector $\underline{\pi}$, which is the solution of the linear system of equations $\underline{\pi}\mathbf{Q} = \underline{0}, \underline{\pi}\mathbb{1} = \mathbb{1}$, where $\mathbb{1}$ is the column vector of ones. In this work, we are interested in the stationary fluid arrival process and assume that the initial probability vector of the background CTMC is $\underline{\alpha} = \underline{\pi}$. For notational convenience, we write $\mathbf{S}$ instead of $\mathbf{\Sigma}/2$ in the sequel.

2.2. Measurement data to fit

We assume that the data to fit is given by a series of pairs $\mathcal{D} = \{(t_k, x_k); k = 1, \ldots, K\}$, where t_k is the time since the last measurement instant and x_k is the amount of fluid arrived since the last measurement instant (which can be negative as well). That is, the measurement instances are $T_k = \sum_{\ell=1}^{k} t_\ell$ for $k \in \{1, \ldots, K\}$.

The likelihood of the data is defined as

$$\mathcal{L}_{\mathbf{Q},\mathbf{R},\mathbf{S}}(\mathcal{D}) = \underline{\alpha} \prod_{k=1}^{K} \mathbf{N}(t_k, x_k)\mathbb{1}. \tag{4}$$

Our goal is to find $\mathbf{Q}$, $\mathbf{R}$ and $\mathbf{S}$ which maximize the likelihood of the data set $\mathcal{D}$.

3. The EM algorithm

The EM algorithm is based on the observation that the likelihood would be easier to maximize when certain unobserved, hidden variables were known. In our case the hidden variables are related to the trajectory of the hidden Markov chain, specifically

- $J_n^{(k)}$ is the nth state visited by the Markov chain in the kth measurement interval,
- $\theta_n^{(k)}$ is the sojourn time of the nth sojourn of the Markov chain (which is in state $J_n^{(k)}$) in the kth measurement interval,
- $f_n^{(k)}$ is the fluid accumulated during the nth sojourn in the kth measurement interval,
- $n^{(k)}$ is the number of sojourns in the kth measurement interval.

Based on these hidden variables the logarithm of the likelihood is computed in the next section.

3.1. Log-likelihood as a function of the hidden variables

With the hidden variables defined above, the likelihood $\mathcal{L}$ can be expressed as

$$\begin{aligned}
\mathcal{L} = \prod_{k=1}^{K} & e^{-\theta_1^{(k)} q_{J_1^{(k)}}} \mathcal{N}\left(\theta_1^{(k)} r_{J_1^{(k)}}, \theta_1^{(k)} \sigma^2_{J_1^{(k)}}, f_1^{(k)}\right) q_{J_1^{(k)} J_2^{(k)}} \\
& \cdot e^{-\theta_2^{(k)} q_{J_2^{(k)}}} \mathcal{N}\left(\theta_2^{(k)} r_{J_2^{(k)}}, \theta_2^{(k)} \sigma^2_{J_2^{(k)}}, f_2^{(k)}\right) q_{J_2^{(k)} J_3^{(k)}} \cdot \ldots \\
& \cdot e^{-\theta_{n^{(k)}}^{(k)} q_{J_{n^{(k)}}^{(k)}}} \mathcal{N}\left(\theta_{n^{(k)}}^{(k)} r_{J_{n^{(k)}}^{(k)}}, \theta_{n^{(k)}}^{(k)} \sigma^2_{J_{n^{(k)}}^{(k)}}, f_{n^{(k)}}^{(k)}\right) \\
= \prod_{k=1}^{K} \prod_{n=1}^{n^{(k)}-1} & e^{-\theta_n^{(k)} q_{J_n^{(k)}}} \mathcal{N}\left(\theta_n^{(k)} r_{J_n^{(k)}}, \theta_n^{(k)} \sigma^2_{J_n^{(k)}}, f_n^{(k)}\right) q_{J_n^{(k)} J_{n+1}^{(k)}} \\
& \cdot e^{-\theta_{n^{(k)}}^{(k)} q_{J_{n^{(k)}}^{(k)}}} \mathcal{N}\left(\theta_{n^{(k)}}^{(k)} r_{J_{n^{(k)}}^{(k)}}, \theta_{n^{(k)}}^{(k)} \sigma^2_{J_{n^{(k)}}^{(k)}}, f_{n^{(k)}}^{(k)}\right),
\end{aligned}$$

where $\mathcal{N}(\mu, \sigma^2, x)$ is the Gaussian density function and $q_i = \sum_{j,j\neq i} q_{ij}$ is the departure rate of state i of the CTMC. Using $\log \mathcal{N}(\mu, \sigma^2, x) = -\frac{c}{2} - \frac{\log \sigma^2}{2} - \frac{(x-\mu)^2}{2\sigma^2}$ with $c = \log 2\pi$ we have

$$
\begin{aligned}
&\log\left(\mathcal{N}\left(\theta_n^{(k)} r_{J_n^{(k)}}, \theta_n^{(k)} \sigma^2_{J_n^{(k)}}, f_n^{(k)}\right)\right) \\
&= -\frac{c}{2} - \frac{\log(\theta_n^{(k)} \sigma^2_{J_n^{(k)}})}{2} - \frac{(f_n^{(k)} - \theta_n^{(k)} r_{J_n^{(k)}})^2}{2\theta_n^{(k)} \sigma^2_{J_n^{(k)}}} \\
&= -\frac{c}{2} - \frac{\log(\theta_n^{(k)}) + \log(\sigma^2_{J_n^{(k)}})}{2} - \frac{{f_n^{(k)}}^2 - 2 f_n^{(k)} \theta_n^{(k)} r_{J_n^{(k)}} + {\theta_n^{(k)}}^2 r^2_{J_n^{(k)}}}{2\theta_n^{(k)} \sigma^2_{J_n^{(k)}}} \\
&= -\frac{c}{2} - \frac{\log \theta_n^{(k)}}{2} - \frac{\log(\sigma^2_{J_n^{(k)}})}{2} - \frac{{f_n^{(k)}}^2}{2\theta_n^{(k)} \sigma^2_{J_n^{(k)}}} + \frac{f_n^{(k)} r_{J_n^{(k)}}}{\sigma^2_{J_n^{(k)}}} - \frac{\theta_n^{(k)} r^2_{J_n^{(k)}}}{2\sigma^2_{J_n^{(k)}}},
\end{aligned}
$$

and the log-likelihood is

$$
\begin{aligned}
\log \mathcal{L} = \sum_{k=1}^{K} \sum_{n=1}^{n^{(k)}-1} & -\theta_n^{(k)} q_{J_n^{(k)}} + \log\left(\mathcal{N}\left(\theta_n^{(k)} r_{J_n^{(k)}}, \theta_n^{(k)} \sigma^2_{J_n^{(k)}}, f_n^{(k)}\right)\right) + \log q_{J_n^{(k)} J_{n+1}^{(k)}} \\
& - \theta^{(k)}_{n^{(k)}} q_{J^{(k)}_{n^{(k)}}} + \log\left(\mathcal{N}\left(\theta^{(k)}_{n^{(k)}} r_{J^{(k)}_{n^{(k)}}}, \theta^{(k)}_{n^{(k)}} \sigma^2_{J^{(k)}_{n^{(k)}}}, f^{(k)}_{n^{(k)}}\right)\right) \\
= \sum_{k=1}^{K} \sum_{n=1}^{n^{(k)}-1} & -\theta_n^{(k)} q_{J_n^{(k)}} - \frac{c}{2} - \frac{\log \theta_n^{(k)}}{2} - \frac{\log(\sigma^2_{J_n^{(k)}})}{2} - \frac{{f_n^{(k)}}^2}{2\theta_n^{(k)} \sigma^2_{J_n^{(k)}}} + \frac{f_n^{(k)} r_{J_n^{(k)}}}{\sigma^2_{J_n^{(k)}}} \\
& - \frac{\theta_n^{(k)} r^2_{J_n^{(k)}}}{2\sigma^2_{J_n^{(k)}}} + \log q_{J_n^{(k)} J_{n+1}^{(k)}} - \theta^{(k)}_{n^{(k)}} q_{J^{(k)}_{n^{(k)}}} - \frac{c}{2} - \frac{\log \theta^{(k)}_{n^{(k)}}}{2} - \frac{\log(\sigma^2_{J^{(k)}_{n^{(k)}}})}{2} \\
& - \frac{{f^{(k)}_{n^{(k)}}}^2}{2\theta^{(k)}_{n^{(k)}} \sigma^2_{J^{(k)}_{n^{(k)}}}} + \frac{f^{(k)}_{n^{(k)}} r_{J^{(k)}_{n^{(k)}}}}{\sigma^2_{J^{(k)}_{n^{(k)}}}} - \frac{\theta^{(k)}_{n^{(k)}} r^2_{J^{(k)}_{n^{(k)}}}}{2\sigma^2_{J^{(k)}_{n^{(k)}}}}.
\end{aligned}
$$

Observe that knowing each individual hidden variable is not necessary to express the log-likelihood. It is enough to introduce the following aggregated measures to fully characterize interval k:

- $\Theta_i^{(k)} = \sum_{n=1}^{n^{(k)}} \theta_n^{(k)} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}}$ is the total time spent in state i,
- $F_i^{(k)} = \sum_{n=1}^{n^{(k)}} f_n^{(k)} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}}$ is the total amount of fluid arriving during a visit in state i,
- $M_i^{(k)} = \sum_{n=1}^{n^{(k)}} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}}$ the number of visits to state i,

- $M_{i,j}^{(k)} = \sum_{n=1}^{n^{(k)}-1} \mathcal{I}_{\left\{J_n^{(k)}=i, J_{n+1}^{(k)}=j\right\}}$ the number of state transitions from state i to state j, additionally
- $L\Theta_i^{(k)} = \sum_{n=1}^{n^{(k)}} \log \theta_n^{(k)} \, \mathcal{I}_{\left\{J_n^{(k)}=i\right\}}$ is the sum of logarithms of the time spent in state i,
- $F\Theta_i^{(k)} = \sum_{n=1}^{n^{(k)}} \frac{{f_n^{(k)}}^2}{\theta_n^{(k)}} \, \mathcal{I}_{\left\{J_n^{(k)}=i\right\}}$ is the sum of square of arriving fluid over the elapsed time in state i.

With these aggregate measures the log-likelihood simplifies to

$$\begin{aligned}
\log \mathcal{L} = \sum_{k=1}^{K} \sum_{n=1}^{n^{(k)}-1} & -\theta_n^{(k)} q_{J_n^{(k)}} - \frac{c}{2} - \frac{\log \theta_n^{(k)}}{2} - \frac{\log(\sigma^2_{J_n^{(k)}})}{2} - \frac{{f_n^{(k)}}^2}{2\theta_n^{(k)} \sigma^2_{J_n^{(k)}}} + \frac{f_n^{(k)} r_{J_n^{(k)}}}{\sigma^2_{J_n^{(k)}}} \\
& - \frac{\theta_n^{(k)} r^2_{J_n^{(k)}}}{2\sigma^2_{J_n^{(k)}}} + \log q_{J_n^{(k)} J_{n+1}^{(k)}} - \theta_{n^{(k)}}^{(k)} q_{J_{n^{(k)}}^{(k)}} - \frac{c}{2} - \frac{\log \theta_{n^{(k)}}^{(k)}}{2} - \frac{\log(\sigma^2_{J_{n^{(k)}}^{(k)}})}{2} \\
& - \frac{{f_{n^{(k)}}^{(k)}}^2}{2\theta_{n^{(k)}}^{(k)} \sigma^2_{J_{n^{(k)}}^{(k)}}} + \frac{f_{n^{(k)}}^{(k)} r_{J_{n^{(k)}}^{(k)}}}{\sigma^2_{J_{n^{(k)}}^{(k)}}} - \frac{\theta_{n^{(k)}}^{(k)} r^2_{J_{n^{(k)}}^{(k)}}}{2\sigma^2_{J_{n^{(k)}}^{(k)}}} \\
= \sum_{k=1}^{K} & - \frac{cn^{(k)}}{2} + \sum_i \left(- \Theta_i^{(k)} \left(q_i + \frac{r_i^2}{2\sigma_i^2} \right) + F_i^{(k)} \frac{r_i}{\sigma_i^2} - \frac{L\Theta_i^{(k)}}{2} - \frac{F\Theta_i^{(k)}}{2\sigma_i^2} \right. \\
& \left. - M_i^{(k)} \frac{\log \sigma_i^2}{2} + \sum_{j, j \neq i} M_{i,j}^{(k)} \log q_{i,j} \right).
\end{aligned}$$

3.2. The maximization step of the EM method

The maximization step of the EM method aims to find the optimal value of the model parameters based on the hidden variables. They are obtained from the partial derivatives of the log-likelihood as detailed in Appendix Appendix A.

Summarizing the results, the model parameter value which maximizes the log-likelihood based on the hidden variables are

$$q_{i,j} = \frac{\sum_{k=1}^{K} M_{i,j}^{(k)}}{\sum_{k=1}^{K} \Theta_i^{(k)}}, r_i = \frac{\sum_{k=1}^{K} F_i^{(k)}}{\sum_{k=1}^{K} \Theta_i^{(k)}}, \text{ and } \sigma_i^2 = \frac{\sum_{k=1}^{K} \Theta_i^{(k)} r_i^2 - 2F_i^{(k)} r_i + F\Theta_i^{(k)}}{\sum_{k=1}^{K} M_i^{(k)}}.$$

That is, $\sum_{k=1}^{K} \Theta_i^{(k)}$, and $\sum_{k=1}^{K} M_{i,j}^{(k)}$ are needed for computing the optimal $q_{i,j}$ parameters and additionally, $\sum_{k=1}^{K} F_i^{(k)}$, $\sum_{k=1}^{K} M_i$ and $\sum_{k=1}^{K} F\Theta_i^{(k)}$ are needed for the optimal r_i and σ_i^2 parameters.

3.3. The expectation step of the EM method

In the expectation step of the EM method the expected values of the hidden variables has to be evaluated based on the samples. Appendix Appendix B provides the analysis of those expectations, resulting $E(F_i^{(k)}) = r_i E(\Theta_i^{(k)})$ and $\mathrm{E}\left(F\Theta_i^{(k)}\right) = \mathrm{E}\left(M_i^{(k)}\right)\sigma_i^2 + \mathrm{E}\left(\Theta_i^{(k)}\right) r_i^2$, from which the zth iteration of the EM method updates the fluid rate and variance parameters as

$$r_i(z+1) = \frac{\sum_{k=1}^{K} \mathrm{E}\left(F_i^{(k)}\right)}{\sum_{k=1}^{K} \mathrm{E}\left(\Theta_i^{(k)}\right)} = r_i(z) \tag{5}$$

and

$$\begin{aligned}
\sigma_i^2(z+1) &= \frac{\sum_{k=1}^{K} \mathrm{E}\left(\Theta_i^{(k)} r_i^2(z) - 2F_i^{(k)} r_i(z) + F\Theta_i^{(k)}\right)}{\sum_{k=1}^{K} \mathrm{E}\left(M_i^{(k)}\right)} \\
&= \frac{\sum_{k=1}^{K} \mathrm{E}\left(\Theta_i^{(k)} r_i^2(z)\right) - 2\mathrm{E}\left(F_i^{(k)}\right) r_i(z) + \mathrm{E}\left(F\Theta_i^{(k)}\right)}{\sum_{k=1}^{K} \mathrm{E}\left(M_i^{(k)}\right)} \\
&= \frac{\sum_{k=1}^{K} \mathrm{E}\left(M_i^{(k)}\right)\sigma_i^2(z)}{\sum_{k=1}^{K} \mathrm{E}\left(M_i^{(k)}\right)} = \sigma_i^2(z).
\end{aligned} \tag{6}$$

Consequently, the fluid rate and variance parameters remain untouched by the EM method.

Remark 1. *This result is in line with the results obtained for discrete arrival processes in [9] considering the special features of the fluid model. That is, we consider the MMPP arrival process, since there is no state transition at the fluid drop arrival, and fluid drops are assumed to be infinitesimal, hence for a finite amount of time there is an unbounded number of fluid drop arrivals. Using these features, equations (21) and (23) of [9] take the form*

$$\begin{aligned}
\mathrm{E}\left(Z_i^{[k]}\right) &= \sum_{l=0}^{x_k} \int_0^{t_k} [f_k(l,\tau)]_i [b_k(x_k - l, t_k - \tau)]_i d\tau \\
\mathrm{E}\left(Y_{ii}^{[k]}\right) &= \lambda_{ii} \sum_{l=0}^{x_k - 1} \int_0^{t_k} [f_k(l,\tau)]_i [b_k(x_k - l, t_k - \tau)]_i d\tau.
\end{aligned}$$

Assuming x_k is large, the update of λ_{ii} in the zth step of the iteration is ((12) of [9])

$$\lambda_{ii}(z+1) = \frac{\sum_{k=1}^{K} \mathrm{E}\left(Y_{ii}^{[k]}\right)}{\sum_{k=1}^{K} \mathrm{E}\left(Z_i^{[k]}\right)}$$

$$= \lambda_{ii}(z) \frac{\sum_{k=1}^{K} \sum_{l=0}^{x_k-1} \int_0^{t_k} [f_k(l,\tau)]_i [b_k(x_k-l, t_k-\tau)]_i d\tau}{\sum_{k=1}^{K} \sum_{l=0}^{x_k} \int_0^{t_k} [f_k(l,\tau)]_i [b_k(x_k-l, t_k-\tau)]_i d\tau} \approx \lambda_{ii}(z).$$

The transition rate parameters are updated by the EM method as

$$q_{i,j} = \frac{\sum_{k=1}^{K} \mathrm{E}\left(M_{i,j}^{(k)}\right)}{\sum_{k=1}^{K} \mathrm{E}\left(\Theta_i^{(k)}\right)}. \tag{7}$$

The computation of $\mathrm{E}\left(M_{i,j}^{(k)}\right)$ and $\mathrm{E}\left(\Theta_i^{(k)}\right)$ are detailed in Appendix Appendix C and the results are summarized in (C.3) and (C.4).

4. Combined fitting method

As a result of the inability of the EM method to optimize the fluid rate and variance parameters of MMFAPs, we propose a combined fitting method, which is composed of consecutive optimization of the background Markov chain using the EM method and the fluid rate and variance parameters using a direct likelihood optimization as it is summarized in Algorithm 1.

Algorithm 1 MMFAP-fit($\mathcal{D}, \mathbf{Q}_{\text{init}}, \mathbf{R}_{\text{init}}, \mathbf{S}_{\text{init}}$)

$\mathbf{Q} \leftarrow \mathbf{Q}_{\text{init}}$; $\mathbf{R} \leftarrow \mathbf{R}_{\text{init}}$;$\mathbf{S} \leftarrow \mathbf{S}_{\text{init}}$;$L \leftarrow -\infty$; $\epsilon \leftarrow 10^{-2}$;
while $|L - \mathcal{L}_{\mathbf{Q},\mathbf{R},\mathbf{S}}(\mathcal{D})|/|L| > \epsilon$ **do**
  $L \leftarrow \mathcal{L}_{\mathbf{Q},\mathbf{R},\mathbf{S}}(\mathcal{D})$
  $\mathbf{Q} \leftarrow$ EM-fit($\mathcal{D}, \mathbf{Q}, \mathbf{R}, \mathbf{S}$)
  $\{\mathbf{R}, \mathbf{S}\} \leftarrow$ Grad-fit($\mathcal{D}, \mathbf{Q}, \mathbf{R}, \mathbf{S}$)
end while

4.1. Direct optimization of the rate and variance parameters

For a direct optimization of the rate and variance parameters, we investigated the behaviour of the likelihood as a function of the system parameters for a given MMFAP, based on 200 samples generated from the MMFAP with parameters

$$\mathbf{Q} = \begin{bmatrix} -0.5 & 0.4 & 0.1 \\ 0.6 & -0.7 & 0.1 \\ 0.2 & 0.3 & -0.5 \end{bmatrix}, \ \mathbf{R} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 8 \end{bmatrix}, \ \mathbf{S} = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.2 & 0 \\ 0 & 0 & 0.4 \end{bmatrix}. \tag{8}$$

Figure 1 and 2 plot the dependence of the likelihood when a single model parameter is adjusted, and all other model parameters were kept at the original MMFAP parameter values from which the samples are generated. The figures indicate that some parameters strongly effect the likelihood while others have less dominant effect. Typically the effect of the lowest fluid rate (r_1) has the most prominent effect.

It is also interesting to notice that the maximum of the likelihood according to s_1 is significantly different from the original s_1 value, and the likelihood at the optimal s_1 value is significantly better than the likelihood computed with the original model parameters.

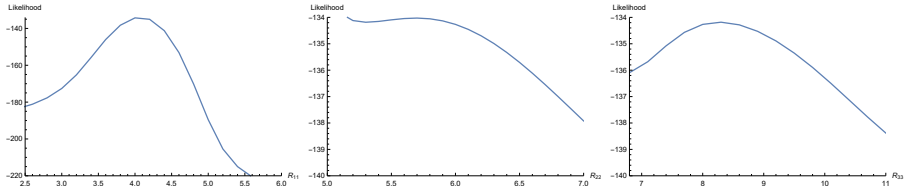

Figure 1: Likelihood as a function of the fluid rate parameters

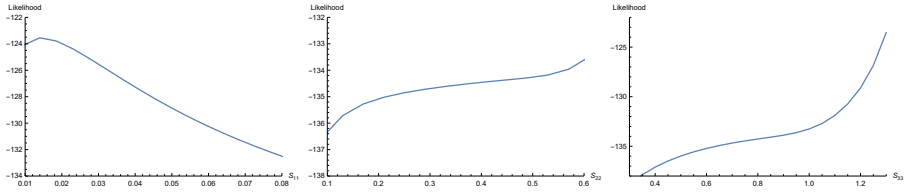

Figure 2: Likelihood as a function of the fluid variance parameters

Motivated by this experiment, we apply a gradient ascent method for finding the parameters which optimize likelihood.

5. Implementation details

The implementation of the EM based parameter estimation method contains some intricate elements which influence the computational complexity and the accuracy of the computations. This section summarizes our proposals for those elements.

5.1. Structural restrictions of MMFAP models

In the case of many discrete Markov modulated arrival processes (e.g. MAP, BMAP), the representation is not unique, and starting from a given representation of an arrival process infinitely many different, but stochastically equivalent representations of the same process can be generated with similarity transformation. From the related literature, it is known that optimizing non-unique representations should be avoided, because the optimization procedure might go back and forth between almost equivalent models having significantly different parameters. The usual solution to address this issue is to apply some structural restrictions (e.g., the Jordan representation of some of the matrices), which can make the representation unique [3].

In this work, we also apply a structural restriction to make the optimization of MMFAPs more efficient (by making the path to the optimum more straight): We restrict matrix $\mathbf{R}$ to be diagonal such that the diagonal elements of $\mathbf{R}$ are non-decreasing, which makes the representation of an MMFAP unique except for the ordering of states with identical fluid arrival rates.

5.2. Initial guesses for the fluid rates

If the variance parameters of the MMFAP were zero, then the amounts of the fluid arrival in a unit of time would be between the r_1 and r_S (the minimal and the maximal) fluid rates. When the variance parameters are positive, the samples might be smaller than r_1 and larger than r_S, but we assume that the smallest and the largest samples, denoted as $x_{\min}$ and $x_{\max}$, carry information on the fluid rates, which we utilize to obtain an initial guess for the fluid rates as

$$r_i = x_{\min} + \frac{i-1}{S-1}(x_{\max} - x_{\min}). \tag{9}$$

For the initial guesses of the rest of the parameters we have much less support from the measured data set. The main problem is that the random effects of the background Markov chain and the variance of the fluid rates can not be easily distinguished.

In our experiments, we assume "small" variances; intuitively it means that the randomness of the samples are dominated by the state transitions of the background Markov chain. Under this assumption, one can gain information on the "speed" of the background Markov chain as it is exemplified in Figure 3 in Section 6.1. Still, more precise information on the transition rates is hard to obtain.

5.3. Computation of the double sided inverse Laplace transform

A crucial step of the algorithm, both in terms of execution speed and numerical accuracy, is to compute the numerical inverse Laplace transformation (NILT) of the expression in (3). There are many efficient numerical inverse transformation methods for single sided functions [15]. However, in our case, the function is defined in double sided Laplace transformation domain (as Gauss distributions can be negative, too), and numerical inverse transformation of double sided Laplace transforms are rather limited.

If $f(t)$ is the density of a positive random variable, then $\int_{-\infty}^{\infty} e^{-st} f(t)dt = \int_0^{\infty} e^{-st} f(t)dt$ and the single and double sided Laplace transforms of $f(t)$ are identical. If $f(t)$ is the density of a random variable which is positive with a high probability, then $\int_{-\infty}^{\infty} e^{-st} f(t)dt \approx \int_0^{\infty} e^{-st} f(t)dt$. Based on this approximation, one can apply single sided numerical inverse Laplace transformation for density functions of dominantly positive random variables.

For a general MMFAP the non-negativity of the fluid increase samples in $T = \{(t_k, x_k); k = 1, \ldots, K\}$ can not be assumed. To make the single sided

numerical inverse Laplace transformation appropriately accurate also in this case, we apply the following model transformation

$$\mathcal{L}_{\mathbf{Q},\mathbf{R},\mathbf{S}}(\{(t_k,x_k);k=1,\ldots,K\})=\mathcal{L}_{\mathbf{Q},\mathbf{R}+c\mathbf{I},\mathbf{S}}(\{(t_k,x_k+ct_k);k=1,\ldots,K\}), \tag{10}$$

where $\mathcal{L}_{\mathbf{Q},\mathbf{R},\mathbf{S}}(\{(t_1,x_1);(t_2,x_2);\ldots;(t_K,x_K)\})=\underline{\alpha}\prod_{k=1}^{K}\mathbf{N}(t_k,x_k)\mathbb{1}$ is the likelihood of the samples when $\mathbf{N}(t_k,x_k)$ is computed with $\mathbf{Q},\mathbf{R},\mathbf{S}$ according to (3) and c is an appropriate constant. If c is too large, the relative difference of the fluid increase samples reduces, and the likelihood function gets less sensitive to the changes of the model parameters. If c is small, fluid increase samples might become negative, and the single sided numerical inverse Laplace transformation might cause numerical issues.

According to (2),

$$[\mathbf{H}(t,y)]_{i,j}=\int_{x=-\infty}^{y}[\mathbf{N}(t,x)]_{i,j}dx=\Pr\left(\mathcal{X}(t)<y,\mathcal{J}(t)=j|\mathcal{J}(0)=i\right) \tag{11}$$

is a probability, which is lower bounded by 0 and upper bounded by 1. Its double sided Laplace transform is

$$\mathbf{H}^*(t,v)=\int_{y=-\infty}^{\infty}e^{-yv}\mathbf{H}(t,y)dy=\int_{y=-\infty}^{\infty}e^{-yv}\int_{x=-\infty}^{y}\mathbf{N}(t,x)dxdy=\frac{1}{v}\mathbf{N}^*(t,v).$$

We set parameter c to the smallest value for which the (single-sided) NILT of $\mathbf{N}^*(t,v)/v$ at point $r_{max}t_{max}$ is between 0 and 1 element wise, where $r_{max}=\max(r_1,\ldots,r_S)$ and $t_{max}=\max(t_1,\ldots,t_K)$.

5.4. Reducing computational cost for equidistant measurement intervals

For computing the likelihood function, the NILT of matrix $\mathbf{N}^*(t,v)$ needs to be performed once for each data point, i.e., K times, which might be computationally expensive.

In the special case when the samples are from identical time intervals, that is $t_1=\ldots=t_K=\bar{t}$, we apply the following approximate approach to reduce the computational complexity to M ($M<<K$) numerical inverse Laplace transformation of matrix $\mathbf{N}^*(t,v)$.

- Let $x_{min}=\min(x_1,\ldots,x_K)$, $x_{max}=\max(x_1,\ldots,x_K)$ and $\Delta=(x_{max}-x_{min})/(M-1)$.
- Compute $\mathbf{N}(\bar{t},x_{min}+(m-1)\Delta)$ for $m=1,\ldots,M$ by NILT of $\mathbf{N}^*(t,v)$.
- For the i,j element of the matrix, approximate the set of points $(x_{min}+(m-1)\Delta,\log\left([\mathbf{N}(\bar{t},x_{min}+(m-1)\Delta)]\right)$ for $m=1,\ldots,M$ with a polynomial of order $\approx M/2$ such that the square error of the approximation in points $x_{min}+(m-1)\Delta$ (with $m=1,\ldots,M$) is minimal.
- Use the exponent of the obtained polynomial function to approximate the NILT.

The higher the parameter M, the higher the accuracy and also the computational complexity of the procedure. In practice we used $M = 20$. An example of $\mathbf{N}(t, x)$ approximation is provided in Section 6.3.

5.5. Computation of $E(\Theta_i^{(k)})$ and $E(M_{i,j}^{(k)})$

Let us introduce the forward and backward likelihood vectors for the beginning and the end of the kth observation period

$$\hat{\underline{f}}_k = \underline{\alpha} \left(\prod_{\ell=1}^{k-1} \mathbf{N}(t_\ell, x_\ell) \right) = \underline{\alpha} \prod_{\ell=1}^{k-1} \mathrm{ILT}_{v \to x_\ell} \mathbf{N}^*(t_\ell, v), \tag{12}$$

$$\hat{\underline{b}}_k = \left(\prod_{\ell=k+1}^{K} \mathbf{N}(t_\ell, x_\ell) \right) \mathbb{1} = \prod_{\ell=k+1}^{K} \mathrm{ILT}_{v \to x_\ell} \mathbf{N}^*(t_\ell, v) \mathbb{1}. \tag{13}$$

and the forward and backward likelihood vectors for an internal point in the kth observation period as

$$\underline{f}_k(t, x) = \underline{\alpha} \left(\prod_{\ell=1}^{k-1} \mathbf{N}(t_\ell, x_\ell) \right) \mathbf{N}(t, x), \tag{14}$$

$$\underline{b}_k(t, x) = \mathbf{N}(t, x) \left(\prod_{\ell=k+1}^{K} \mathbf{N}(t_\ell, x_\ell) \right) \mathbb{1}. \tag{15}$$

We note that, using $\hat{\underline{f}}_k$ and $\hat{\underline{b}}_k$, the likelihood can be expressed as

$$\begin{aligned} \mathcal{L} &= \underline{\alpha} \cdot \underline{b}_1(t_1, x_1) = \underline{f}_{k-1}(t_{k-1}, x_{k-1}) \cdot \underline{b}_k(t_k, x_k) = \underline{f}_K(t_K, x_K) \mathbb{1} \\ &= \underline{\alpha} \cdot \hat{\underline{b}}_0 = \hat{\underline{f}}_\ell \cdot \hat{\underline{b}}_{\ell-1} = \hat{\underline{f}}_{K+1} \mathbb{1}, \end{aligned}$$

for any $k = 2, \ldots, K-1$ and $\ell = 1, \ldots, K$.

To compute the expected value of $\Theta_i^{(k)}$, the integrals of the forward and backward likelihood vectors have to be evaluated. The special form of the integrals allows for simplifications as

$$\begin{aligned} E(\Theta_i^{(k)}) &= \int_{x=0}^{x_k} \int_{t=0}^{t_k} [\underline{f}_k(t, x)]_i \cdot [\underline{b}_k(t_k - t, x_k - x)]_i \, dt \, dx \\ &= \hat{\underline{f}}_k \left(\int_{x=0}^{x_k} \int_{t=0}^{t_k} \mathbf{N}(t, x) \underline{e_i} \cdot \underline{e_i}^T \mathbf{N}(t_k - t, x_k - x) \, dt \, dx \right) \hat{\underline{b}}_k \\ &= \hat{\underline{f}}_k \, \mathrm{ILT}_{v \to x_k} \left(\int_{t=0}^{t_k} \mathbf{N}^*(t, v) \underline{e_i} \cdot \underline{e_i}^T \mathbf{N}^*(t_k - t, v) \, dt \right) \hat{\underline{b}}_k \\ &= \hat{\underline{f}}_k \, \mathrm{ILT}_{v \to x_k} \left(\int_{t=0}^{t_k} e^{(\mathbf{Q} - v\mathbf{R} + v^2 \boldsymbol{\Sigma}/2)t} \underline{e_i} \cdot \underline{e_i}^T e^{(\mathbf{Q} - v\mathbf{R} + v^2 \boldsymbol{\Sigma}/2)(t_k - t)} \, dt \right) \hat{\underline{b}}_k \\ &= \hat{\underline{f}}_k \, \mathrm{ILT}_{v \to x_k} \left(\begin{bmatrix} \mathbf{0} & \mathbf{I} \end{bmatrix} e^{\begin{bmatrix} \mathbf{Q} - v\mathbf{R} + v^2 \boldsymbol{\Sigma}/2 & \underline{e_i} \cdot \underline{e_i}^T \\ \mathbf{0} & \mathbf{Q} - v\mathbf{R} + v^2 \boldsymbol{\Sigma}/2 \end{bmatrix} t_k} \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \right) \hat{\underline{b}}_k. \end{aligned}$$

That is, the convolution integral is replaced by the evaluation of a matrix exponential of double size [16]. In a similar manner, the expected value of $M_{i,j}^{(k)}$ is

$$E(M_{i,j}^{(k)}) = \int_{x=0}^{x_k} \int_{t=0}^{t_k} [\underline{f}_k(t,x)]_i \cdot q_{ij} \cdot [\underline{b}_k(t_k - t, x_k - x)]_j \, dt \, dx$$
$$= q_{ij} \; \hat{\underline{f}}_k \; \mathrm{ILT}_{v \to x_k} \left(\begin{bmatrix} \mathbf{0} & \mathbf{I} \end{bmatrix} e^{\begin{bmatrix} \mathbf{Q} - v\mathbf{R} + v^2\mathbf{\Sigma}/2 & \underline{e_i} \cdot \underline{e_j}^T \\ \mathbf{0} & \mathbf{Q} - v\mathbf{R} + v^2\mathbf{\Sigma}/2 \end{bmatrix} t_k} \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \right) \hat{\underline{b}}_k.$$

We note that these expressions give an interpretation for the zth iteration of the EM method for $q_{i,j}$

$$q_{i,j}(z+1) = \frac{\sum_{k=1}^{K} \mathrm{E}\left(M_{i,j}^{(k)}\right)}{\sum_{k=1}^{K} \mathrm{E}\left(\Theta_i^{(k)}\right)} =$$
$$q_{i,j}(z) \; \frac{\hat{\underline{f}}_k \; \mathrm{ILT}_{v \to x_k} \left(\begin{bmatrix} \mathbf{0} & \mathbf{I} \end{bmatrix} e^{\begin{bmatrix} \mathbf{Q} - v\mathbf{R} + v^2\mathbf{\Sigma}/2 & \underline{e_i} \cdot \underline{e_j}^T \\ \mathbf{0} & \mathbf{Q} - v\mathbf{R} + v^2\mathbf{\Sigma}/2 \end{bmatrix} t_k} \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \right) \hat{\underline{b}}_k}{\hat{\underline{f}}_k \; \mathrm{ILT}_{v \to x_k} \left(\begin{bmatrix} \mathbf{0} & \mathbf{I} \end{bmatrix} e^{\begin{bmatrix} \mathbf{Q} - v\mathbf{R} + v^2\mathbf{\Sigma}/2 & \underline{e_i} \cdot \underline{e_i}^T \\ \mathbf{0} & \mathbf{Q} - v\mathbf{R} + v^2\mathbf{\Sigma}/2 \end{bmatrix} t_k} \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \right) \hat{\underline{b}}_k},$$

that is, $q_{i,j}(z+1)$ is the product of $q_{i,j}(z)$ and an actual guess dependent value.

5.6. Computation of $\hat{\underline{f}}_k$ and $\hat{\underline{b}}_k$

The computation of $\hat{\underline{f}}_k$ and $\hat{\underline{b}}_k$ follows a similar pattern and contains the same difficulties, except that $\hat{\underline{f}}_k$ is computed from $k = 0$ onward and $\hat{\underline{b}}_k$ is computed from $k = K$ downward. The main implementation issue with the computation of $\hat{\underline{f}}_k$ and $\hat{\underline{b}}_k$, is to avoid under-/overflow during the computation. We adopted the under-/overflow avoiding method proposed in [17].

6. Numerical examples

6.1. MMFAP simulator

For the numerical evaluation of the proposed method, we developed a simulator which generates the required number (K) of traffic samples based on matrices $\mathbf{Q}$, $\mathbf{R}$ and $\mathbf{S}$. In each simulation step, the program samples the next state transition of the Markov chain and checks if it occurs before or after the next measurement instance. In the first case, it samples the accumulated fluid

until the next state transition and performs the state transition. In the second case, it samples the accumulated fluid until the next measurement instance and maintains the state of the Markov chain.

Our simulator assumes equidistant time intervals, such that $t_1 = \ldots = t_K = 1$, which allows to utilize the computationally efficient approximate approach introduced in Section 5.4.

For the MMFAP with

$$\mathbf{Q}_{slow} = \begin{bmatrix} -0.8 & 0.5 & 0.3 \\ 0.6 & -0.7 & 0.1 \\ 0.2 & 0.3 & -0.5 \end{bmatrix}, \mathbf{R} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 8 \end{bmatrix}, \mathbf{S} = \begin{bmatrix} 0.01 & 0 & 0 \\ 0 & 0.02 & 0 \\ 0 & 0 & 0.04 \end{bmatrix}, \tag{16}$$

the histogram of the samples is presented in Figure 3a. The histogram indicates that the Markov chain is "slow" in this case, i.e., it stays in a single state (e.g., state i) during the measurement interval of length 1 with high probability and accumulates $\mathcal{N}(r_i, \sigma_i^2)$ distributed amount of fluid during this interval. That is the explanation of the peaks at around $r_1 = 2$, $r_2 = 4$ and $r_3 = 8$. It is also visible that the transitions between state 1 and 2 are faster than the transitions to and from state 3, and consequently, the histogram indicates fluid samples in the $x \in (2, 4)$ interval. These samples might come from measurement intervals starting from state 1 with $r_1 = 2$ and moving to the state with $r_2 = 4$, or vice versa.

To indicate the effect of the "speed" of the Markov chain on the histogram of the generated samples, Figure 3b depicts the histogram when the Markov chain is "fast", namely 10 times faster, $\mathbf{Q}_{fast} = 10\mathbf{Q}_{slow}$. The "fast" Markov chain experiences state transitions during the measurement interval with very high probability, and the amount of fluid accumulated during the interval gets to be less dependent on the state of the Markov chain at the beginning of the measurement interval.

When the variance is low, as it is in this example, one can easily predict the values of the $\mathbf{R}$ matrix with the "slow" Markov chain, while for the "fast" Markov chain, the values of the $\mathbf{R}$ matrix is not possible to guess based on the histogram. Still, the minimal and the maximal sample values allow estimating the minimal and the maximal fluid rates of matrix $\mathbf{R}$.

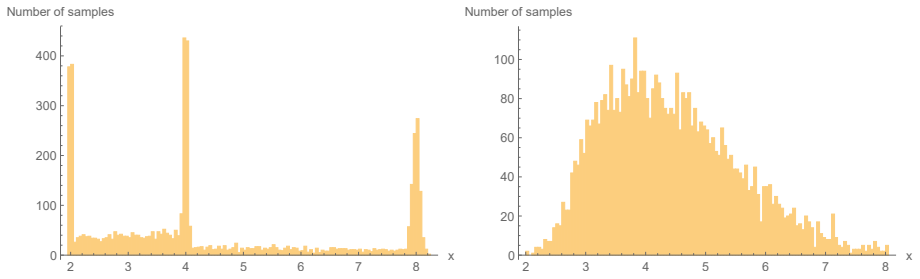

Figure 3: Histogram of 5000 generated samples with $\mathbf{Q}_{slow}$, $\mathbf{R}$ and $\mathbf{S}$ and $\mathbf{Q}_{fast}$, $\mathbf{R}$ and $\mathbf{S}$ defined in (16).

6.2. Approximating $\mathbf{Q}$ *with the EM method*

Based on 300 samples of the MMFAP with

$$\bar{\mathbf{Q}} = \begin{bmatrix} -2 & 2 \\ 4 & -4 \end{bmatrix}, \ \bar{\mathbf{R}} = \begin{bmatrix} 4 & 0 \\ 0 & 8 \end{bmatrix}, \ \bar{\mathbf{S}} = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.02 \end{bmatrix}, \tag{17}$$

we approximate the MMFAP starting from $\bar{\mathbf{Q}}_{\mathbf{0}} = \begin{bmatrix} -1 & 1 \\ 2.5 & -2.5 \end{bmatrix}$, $\bar{\mathbf{R}}_{\mathbf{0}} = \bar{\mathbf{R}}$, $\bar{\mathbf{S}}_{\mathbf{0}} = \bar{\mathbf{S}}$ with the EM method. The log-likelihood value and the transition rates of the Markov chain are depicted in Figure 4a and 4b, respectively, where the dotted horizontal lines refer to the MMFAP according to (18), which was used to generate the samples. The figure indicates that the obtained transition rates provide a bit higher log-likelihood than the ones in (18). Additionally, the figures report convergence after ~ 25 iterations of the EM method.

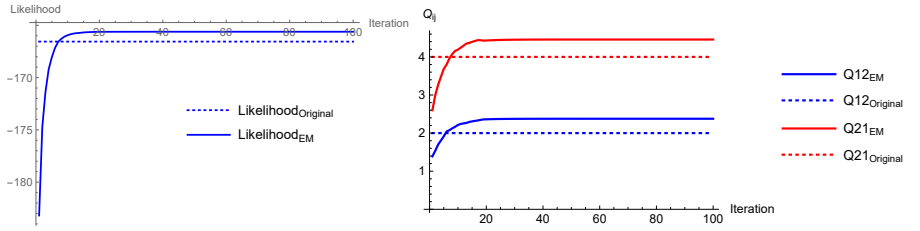

Figure 4: Behaviour of the EM method based on 300 samples generated from $\bar{\mathbf{Q}}$, $\bar{\mathbf{R}}$ and $\bar{\mathbf{S}}$ in (18) with initial guess $\bar{\mathbf{Q}}_{\mathbf{0}}$, $\bar{\mathbf{R}}_{\mathbf{0}}$ and $\bar{\mathbf{S}}_{\mathbf{0}}$. According to (5) and (6), $\bar{\mathbf{R}}_{\mathbf{0}}$ and $\bar{\mathbf{S}}_{\mathbf{0}}$ remained unchanged during the EM iterations.

Similarly, we evaluated the EM based approximation of the MMFAP defined in (16) based on 1000 samples starting from $\mathbf{Q}_{\mathbf{0}} = \begin{bmatrix} -0.8 & 0.5 & 0.3 \\ 0.6 & -0.7 & 0.1 \\ 0.2 & 0.3 & -0.5 \end{bmatrix}$, $\mathbf{R}_{\mathbf{0}} = \mathbf{R}$, $\mathbf{S}_{\mathbf{0}} = \mathbf{S}$.

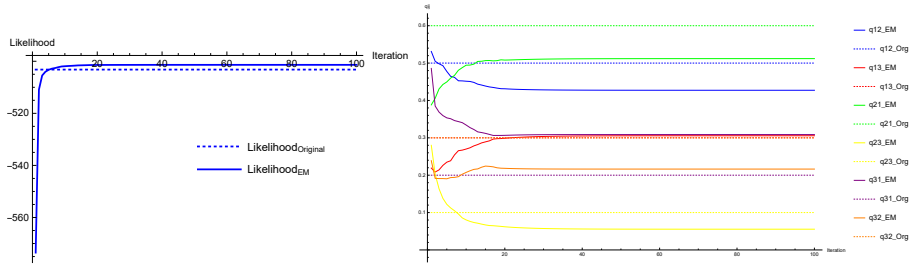

Figure 5: Behaviour of the EM method based on 1000 samples generated from $\mathbf{Q}$, $\mathbf{R}$ and $\mathbf{S}$ in (16) with initial guess $\mathbf{Q}_{\mathbf{0}}$, $\mathbf{R}_{\mathbf{0}}$ and $\mathbf{S}_{\mathbf{0}}$.

The evolution of the log-likelihood value and the transition rates of the Markov chain along the EM iterations are depicted in Figure 5a and 5b, respectively. Figure 5a indicates, that similar to the 2×2 example in Figure 4a, the

likelihood value increased above the one obtained from the original MMFAP. At the same time, the transition rate values in Figure 5b differ more significantly from ones of the original MMFAP than in Figure 4b, which might be a consequence of a looser relation between the transition rates and the likelihood value in higher dimensions.

6.3. Functional approximation of $\mathbf{N}(t,x)$

To accelerate NILT based computation of $\mathbf{N}(t,x)$, a functional approximation procedure is introduced in Section 5.4. Figure 6 and 7 report the accuracy of the proposed approach via a numerical example to approximate $[\mathbf{N}(t,x)]_{1,1}$ with parameters $\mathbf{Q}_{slow}$, $\mathbf{R}$, $\mathbf{S}$ defined in (16). The NILT of $[\mathbf{N}^*(t,v)]_{1,1}$ is computed in $M = 20$ points, and based on those points an order 10 polynomial approximates $[\mathbf{N}(t,x)]_{1,1}$.

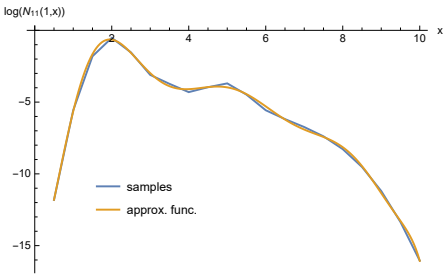

Figure 6: Quality of fitting $\log[\mathbf{N}(t,x)]_{1,1}$.

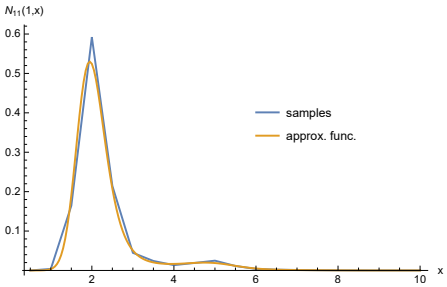

Figure 7: Quality of fitting $[\mathbf{N}(t,x)]_{1,1}$.

6.4. Approximating $\mathbf{R}$ *and* $\mathbf{S}$ *with direct optimization*

6.4.1. Direct optimization example 1

Using the MMFAP simulator we generated 500 samples from the MMFAP with $\mathbf{Q}_{slow}$,

$$\mathbf{R} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 8 \end{bmatrix}, \ \mathbf{S} = \begin{bmatrix} 0.1 & 0 & 0 \\ 0 & 0.5 & 0 \\ 0 & 0 & 0.8 \end{bmatrix} \tag{18}$$

and we performed the gradient based rate and variance optimization starting from $\mathbf{Q}_{slow}$,

$$\mathbf{R}_{init} = \begin{bmatrix} 2.5 & 0 & 0 \\ 0 & 7 & 0 \\ 0 & 0 & 11.5 \end{bmatrix}, \ \mathbf{S}_{init} = \begin{bmatrix} 0.125 & 0 & 0 \\ 0 & 0.35 & 0 \\ 0 & 0 & 0.575 \end{bmatrix} \tag{19}$$

The evolution of the likelihood value during the optimization is depicted in Figure 8 and the final likelihood is obtained by

$$\mathbf{R}_{final} = \begin{bmatrix} 2.49159 & 0 & 0 \\ 0 & 6.96413 & 0 \\ 0 & 0 & 11.456 \end{bmatrix}, \ \mathbf{S}_{final} = \begin{bmatrix} 0.1777 & 0 & 0 \\ 0 & 0.391667 & 0 \\ 0 & 0 & 0.619367 \end{bmatrix} \tag{20}$$

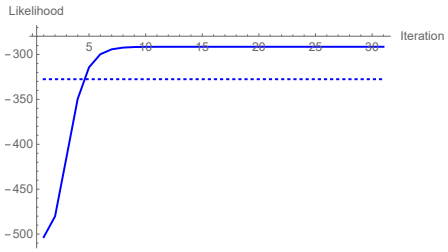

Figure 8: Likelihood value during the optimization of Example 1

In this example the obtained optimum is rather close to the initial guess, which might suggest that there are many local optima in the surface. The next example reports a longer journey in the parameter space.

6.4.2. Direct optimization example 2

We generated 300 samples from the MMFAP provided in (8) and we performed the gradient-based rate optimization starting from $\mathbf{Q}$, $\mathbf{S}$, and $\{r_1, r_2, r_3\} = \{1, 3, 6\}$. This initial guess differs from the one proposed in (9), still it allows us to check the properties of the optimization process starting from a more remote initial point.

The likelihood and the fluid rate values during the optimization are depicted in Figure 9. In this particular example, the optimization procedure finds almost the same fluid rates as we used for generating the samples. The final fluid rates differ only a bit, but this slight difference from the original parameters provides a higher likelihood value than the one obtained from the original parameters in (8).

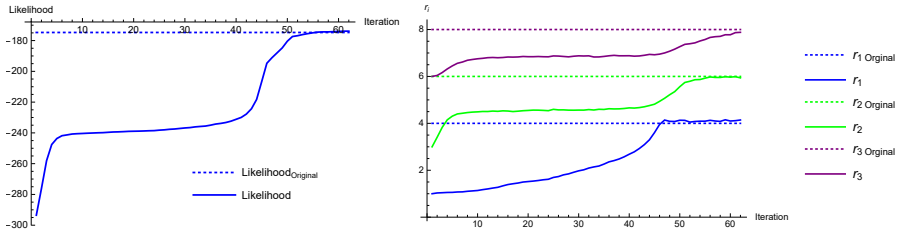

Figure 9: Likelihood and fluid rate values during the optimization of Example 2

6.5. Summary of numerical examples

Based on a much broader numerical experiments, than the ones reported above, we conclude that the proposed combined MMFAP fitting method resembles many properties of the EM method for discrete arrival process fitting:

- The procedure is somewhat sensitive to the initial guess. In many unreported cases, the obtained maximum likelihood remained below the one computed from the original parameters.
- Its stability (ability to obtain a reasonable high optimum) decreases with the number of model parameters. For example, when we optimized 6 parameters in Example 1, the procedure stopped at a local maximum close to the initial point. Instead, when we optimized only 3 model parameters in Example 2, the procedure significantly modified their values during the optimization process.

We guess, that these similarities of the EM based discrete arrival process fitting and the combined MMFAP fitting methods have common roots in the properties of their respective likelihood functions, and the applied optimization procedure slightly affects the optimization quality.

Anyhow, starting from a proper initial guess, the proposed combined optimization procedure can obtain a higher likelihood than the one computed from the original parameters, which means that the original model parameters are not necessarily the optimal ones for a given data set, and the quality of the fitting method cannot be judged based on the identity of the model parameters, but it has to be judged based on the likelihood values.

6.6. Running time

The set of evaluated experiments, is limited by the computational complexity of our Mathematica implementation. There are many elements of the implementation which affects the computational complexity. Mathematica is not the most efficient programming environment for such numerical computations, but we used it because the applied NILT method, the CME method, is conveniently available in Mathematica. During the computations, we used the CME method with order 30. The parameter tuning the trade off between accuracy and the number of points where the NILT is computed for functional approximation was set to $M = 20$ (c.f. Section 5.4 and 6.3). All in all, the computation time of the combined fitting in Algorithm 1 with 3-state background Markov chain, 300 samples and ≈ 100 iterations was 4-8 hours on our regular desktop computers.

7. Conclusions

The EM method is commonly applied for parameter estimation of Markov modulated models. In this paper, we considered the fitting of MMFAP and showed that the EM method is not applicable for optimizing the fluid rate and variance parameters. As a result, we propose a combined fitting approach where the modulating Markov chain is optimized via the EM method, and the fluid rate and variance parameters are optimized via direct likelihood optimization.

Numerical examples investigate the properties of the proposed combined MMFAP fitting method. The results of the numerical experiments suggest that the properties of the combined method for MMFAP fitting are similar to the ones of the EM based fitting of discrete arrival processes.

[1] M. F. Neuts, A versatile Markovian point process, Journal of Applied Probability 16 (1979) 764–779.

[2] K. Mitchell, A. van de Liefvoort, Approximation models of feed-forward G/G/1/N queueing networks with correlated arrivals, Performance Evaluation 51 (2) (2003) 137–152.

[3] M. Telek, G. Horváth, A minimal representation of Markov arrival processes and a moments matching method, Performance Evaluation 64 (9-12) (2007) 1153–1168.

[4] H. Okamura, T. Dohi, Faster maximum likelihood estimation algorithms for Markovian arrival processes, in: 2009 Sixth international conference on the quantitative evaluation of systems, IEEE, 2009, pp. 73–82.

[5] P. Buchholz, An EM-algorithm for MAP fitting from real traffic data, in: International Conference on Modelling Techniques and Tools for Computer Performance Evaluation, Springer, 2003, pp. 218–236.

[6] T. Rydén, An EM algorithm for estimation in Markov-modulated Poisson processes, Computational Statistics & Data Analysis 21 (4) (1996) 431–447.

[7] G. Horváth, H. Okamura, A fast EM algorithm for fitting marked Markovian Arrival Processes with a new special structure, in: Computer Performance Engineering, Springer Berlin Heidelberg, 2013, pp. 119–133.

[8] L. Breuer, An EM algorithm for batch Markovian arrival processes and its comparison to a simpler estimation procedure, Annals of Operations Research 112 (1) (2002) 123–138.

[9] H. Okamura, T. Dohi, K. Trivedi, Markovian arrival process parameter estimation with group data, IEEE/ACM Trans. Netw. 17 (2009) 1326–1339.

[10] D. Anick, D. Mitra, M. M. Sondhi, Stochastic theory of a data-handling system with multiple sources, Bell System technical journal 61 (8) (1982) 1871–1894.

[11] D. Mitra, Stochastic theory of a fluid model of producers and consumers coupled by a buffer, Advances in Applied Probability 20 (3) (1988) 646–676.

[12] S. Asmussen, Stationary distributions for fluid flow models with or without Brownian noise, Communications in Statistics. Stochastic Models 11 (1) (1995) 21–49.

[13] S. A.-D. Almousa, G. Horváth, M. Telek, EM based parameter estimation for Markov modulated fluid arrival processes, in: 26th International Conference on Analytical & Stochastic Modelling Techniques & Applications (ASMTA), Vol. 13104 of LNCS, Tsukuba, Japan, 2021, pp. 226–242. doi:10.1007/978-3-030-91825-5_14.

[14] G. Horváth, S. Rácz, M. Telek, Analysis of second-order Markov reward models, in: The International Conference on Dependable Systems and Networks, DSN/PDS 2004, IEEE CS Press, Florence, Italy, 2004, pp. 845–854.

[15] I. Horváth, G. Horváth, S. A.-D. Almousa, M. Telek, Numerical inverse Laplace transformation using concentrated matrix exponential distributions, Performance Evaluation.

[16] C. Van Loan, Computing integrals involving the matrix exponential, IEEE Transactions on Automatic Control 23 (3) (1978) 395–404.

[17] M. Bražėnas, G. Horváth, M. Telek, Parallel algorithms for fitting Markov Arrival Processes, Performance Evaluation 123-124 (2018) 50 – 67.

Appendix

Appendix A. Maximizing the model parameters

Assuming $q_i = \sum_{j,j\neq i} q_{i,j}$, the derivatives of $\log \mathcal{L}$ are as follows:

$$\begin{aligned}\frac{\partial}{\partial q_{i,j}} \log \mathcal{L} &= \frac{\partial}{\partial q_{i,j}} \sum_{k=1}^{K} \left(-\Theta_i^{(k)} q_{i,j} + M_{i,j}^{(k)} \log q_{i,j}\right) \\ &= \sum_{k=1}^{K} \left(-\Theta_i^{(k)} + M_{i,j}^{(k)} \frac{1}{q_{i,j}}\right),\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial r_i} \log \mathcal{L} &= \frac{\partial}{\partial r_i} \sum_{k=1}^{K} \left(-\Theta_i^{(k)} \frac{r_i^2}{2\sigma_i^2} + F_i^{(k)} \frac{r_i}{\sigma_i^2}\right) \\ &= \sum_{k=1}^{K} \left(-\Theta_i^{(k)} \frac{r_i}{\sigma_i^2} + F_i^{(k)} \frac{1}{\sigma_i^2}\right),\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial \sigma_i^2} \log \mathcal{L} &= \frac{\partial}{\partial \sigma_i^2} \sum_{k=1}^{K} \left(-\Theta_i^{(k)} \frac{r_i^2}{2\sigma_i^2} + F_i^{(k)} \frac{r_i}{\sigma_i^2} - \frac{1}{2\sigma_i^2} F\Theta_i^{(k)} - M_i^{(k)} \frac{\log \sigma_i^2}{2}\right) \\ &= \frac{\partial}{\partial \sigma_i^2} \sum_{k=1}^{K} \left(\left(-\Theta_i^{(k)} r_i^2 + 2F_i^{(k)} r_i - F\Theta_i^{(k)}\right) \frac{1}{2\sigma_i^2} - M_i^{(k)} \frac{\log \sigma_i^2}{2}\right) \\ &= \sum_{k=1}^{K} \left(\left(-\Theta_i^{(k)} r_i^2 + 2F_i^{(k)} r_i - F\Theta_i^{(k)}\right) \frac{-1}{2(\sigma_i^2)^2} - M_i^{(k)} \frac{1}{2\sigma_i^2}\right).\end{aligned}$$

The optimal parameter values are obtained where the derivative is zero:

$$0 = \sum_{k=1}^{K} \Theta_i^{(k)} - M_{i,j}^{(k)}/q_{i,j} \longrightarrow q_{i,j} = \frac{\sum_{k=1}^{K} M_{i,j}^{(k)}}{\sum_{k=1}^{K} \Theta_i^{(k)}}. \tag{A.1}$$

$$0 = \sum_{k=1}^{K} \frac{\Theta_i^{(k)} r_i - F_i^{(k)}}{\sigma_i^2} \longrightarrow r_i = \frac{\sum_{k=1}^{K} F_i^{(k)}}{\sum_{k=1}^{K} \Theta_i^{(k)}}. \tag{A.2}$$

$$0 = \sum_{k=1}^{K} \left(\Theta_i^{(k)} r_i^2 - 2F_i^{(k)} r_i + F\Theta_i^{(k)} \right) \frac{1}{\sigma_i^2} - M_i^{(k)}$$

$$\longrightarrow \sigma_i^2 = \frac{\sum_{k=1}^{K} \Theta_i^{(k)} r_i^2 - 2F_i^{(k)} r_i + F\Theta_i^{(k)}}{\sum_{k=1}^{K} M_i^{(k)}}. \tag{A.3}$$

Appendix B. Expected values of the hidden parameters

For $E(F_i^{(k)})$ we have

$$\mathrm{E}\left(F_i^{(k)}\right) = E_{\Theta_i^{(k)}} E_{F_i^{(k)}|\Theta_i^{(k)}}(F_i^{(k)}) = E_{\Theta_i^{(k)}} r_i \Theta_i^{(k)} = r_i E(\Theta_i^{(k)}). \tag{B.1}$$

For $F\Theta_i^{(k)} = \sum_{n=1}^{n^{(k)}} \frac{{f_n^{(k)}}^2}{\theta_n^{(k)}} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}}$, we have

$$\mathrm{E}\left(F\Theta_i^{(k)}\right) = \mathrm{E}\left(\sum_{n=1}^{n^{(k)}} \frac{{f_n^{(k)}}^2}{\theta_n^{(k)}} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}}\right)$$

$$= \mathrm{E}_{\{n^{(k)},\theta_1^{(k)},\ldots,\theta_{n^{(k)}}^{(k)}\}}\left(\sum_{n=1}^{n^{(k)}} \mathrm{E}_{\{f_n^{(k)}|\theta_n^{(k)}\}}\left(\frac{{f_n^{(k)}}^2}{\theta_n^{(k)}} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}}\right)\right). \tag{B.2}$$

Let $\mathcal{X}(\mu,\sigma^2)$ denote a normal distributed random variable with mean μ and variance σ^2. Its second moment is $\mathrm{E}\left(\mathcal{X}^2(\mu,\sigma^2)\right) = \sigma^2+\mu^2$. When $\theta_n^{(k)} = x$ then $f_n^{(k)}$ is $\mathcal{X}(xr_i, x\sigma_i^2)$ distributed and $\mathrm{E}\left({f_n^{(k)}}^2\right) = \mathrm{E}\left(\mathcal{X}^2(xr_i, x\sigma_i^2)\right) = x\sigma_i^2 + x^2 r_i^2$, that is

$$\mathrm{E}_{\{f_n^{(k)}|\theta_n^{(k)}\}}\left(\frac{{f_n^{(k)}}^2}{\theta_n^{(k)}} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}}\right) = \frac{\mathrm{E}_{\{f_n^{(k)}|\theta_n^{(k)}\}}\left({f_n^{(k)}}^2\right)}{\theta_n^{(k)}} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}}$$

$$= \frac{\theta_n^{(k)}\sigma_i^2 + {\theta_n^{(k)}}^2 r_i^2}{\theta_n^{(k)}} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}} = \left(\sigma_i^2 + \theta_n^{(k)} r_i^2\right) \mathcal{I}_{\left\{J_n^{(k)}=i\right\}}. \tag{B.3}$$

Substituting (B.3) into (B.2) results

$$\mathrm{E}\left(F\Theta_i^{(k)}\right) = \mathrm{E}\left(\sum_{n=1}^{n^{(k)}} \frac{{f_n^{(k)}}^2}{\theta_n^{(k)}} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}}\right)$$

$$= \mathrm{E}_{\{n^{(k)},\theta_1^{(k)},\ldots,\theta_{n^{(k)}}^{(k)}\}}\left(\sum_{n=1}^{n^{(k)}} \mathrm{E}_{\{f_n^{(k)}|\theta_n^{(k)}\}}\left(\frac{{f_n^{(k)}}^2}{\theta_n^{(k)}} \mathcal{I}_{\left\{J_n^{(k)}=i\right\}}\right)\right)$$

$$= \mathrm{E}_{\{n^{(k)},\theta_1^{(k)},\ldots,\theta_{n^{(k)}}^{(k)}\}} \left(\sum_{n=1}^{n^{(k)}} \left(\sigma_i^2 + \theta_n^{(k)} r_i^2 \right) \mathcal{I}_{\left\{ J_n^{(k)}=i \right\}} \right)$$
$$= \mathrm{E}\left(M_i^{(k)} \right) \sigma_i^2 + \mathrm{E}\left(\Theta_i^{(k)} \right) r_i^2.$$

Appendix C. Numerical computation of the expected value of the hidden parameters

In the E-step we compute the expected value of the hidden parameters for given $\underline{\alpha}$, $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{S}$ and observed data (t_k, x_k) for $k = 1, \ldots, K$. For the expected values of $\Theta_i^{(k)}$ we have

$$\mathrm{E}\left(\Theta_i^{(k)} | t_k, x_k \right) = \mathrm{E}\left(\sum_{n=1}^{n^{(k)}} \theta_n^{(k)} \mathcal{I}_{\left\{ J_n^{(k)}=i \right\}} | t_k, x_k \right) = \mathrm{E}\left(\int_{t=0}^{t_k} \mathcal{I}_{\{\mathcal{J}(t)=i|x_k\}} dt \right) \tag{C.1}$$

$$= \int_{t=0}^{t_k} \mathrm{E}\left(\mathcal{I}_{\{\mathcal{J}(t)=i|x_k\}} \right) dt = \int_{t=0}^{t_k} \Pr\left(\mathcal{J}(t) = i | x_k \right) dt$$
$$= \sum_k \sum_\ell \int_{t=0}^{t_k} \Pr\left(\mathcal{J}(0) = k, \mathcal{J}(t) = i, \mathcal{J}(t_k) = \ell | x_k \right) dt$$
$$= \sum_k \sum_\ell \int_{t=0}^{t_k} \Pr\left(\mathcal{J}(0) = k \right)$$
$$\int_{x=0}^{x_k} \lim_{\Delta \to 0} \frac{1}{\Delta} \Pr\left(x \le \mathcal{X}(t) < x + \Delta, \mathcal{J}(t) = i | \mathcal{J}(0) = k, \mathcal{X}(0) = 0 \right)$$
$$\lim_{\Delta \to 0} \frac{1}{\Delta} \Pr\left(x_k \le \mathcal{X}(t_k) < x_k + \Delta, \mathcal{J}(t) = \ell | \mathcal{J}(t) = i, \mathcal{X}(t) = x \right) dx dt$$
$$= \underline{\alpha}_k \int_{t=0}^{t_k} \int_{x=0}^{x_k} \mathbf{N}(t, x) \underline{e_i} \underline{e_i}^T \mathbf{N}(t_k - t, x_k - x) \mathbb{1} dx dt \tag{C.2}$$

where the jth element of vector $\underline{\alpha}_k$ is $\Pr\left(\mathcal{J}(0) = j \right)$ and $\underline{e_i}$ is the ith unit column vector.

According to (C.2), (14) and (15) , the expected value of $\Theta_i^{(k)}$ is

$$E(\Theta_i^{(k)}) = \int_{x=0}^{x_k} \int_{t=0}^{t_k} [\underline{f}_k(t, x)]_i \cdot [\underline{b}_k(t_k - t, x_k - x)]_i \, dt \, dx. \tag{C.3}$$

In a similar manner, the expected value of $M_{i,j}^{(k)}$ is

$$E(M_{i,j}^{(k)}) = \int_{x=0}^{x_k} \int_{t=0}^{t_k} [\underline{f}_k(t, x)]_i \cdot q_{i,j} \cdot [\underline{b}_k(t_k - t, x_k - x)]_j \, dt \, dx. \tag{C.4}$$