Artigo Interno CEG-IST

NON-STATIONARY QUEUE SIMULATION ANALYSIS USING TIME SERIES

Rita Marques Brandão

Departamento de Matemática
Universidade dos Açores
Apartado 1422, 9501-801 Ponta Delgada, Portugal

Acácio M. O. Porta Nova

Departamento de Engenharia e Gestão
Instituto Superior Técnico
Av. Rovisco Pais, 1049-001 Lisboa, Portugal

ABSTRACT

In this work, we extend the use of time series models to the output analysis of non-stationary discrete event simulations. In particular, we investigate and experimentally evaluate the applicability of $ARIMA(p, d, q)$ models as potential meta-models for simulating queueing systems under critical traf£c conditions. We exploit stationarity-inducing transformations, in order to ef£ciently estimate performance measures of selected responses in the system under study.

KEYWORDS

Discrete event simulation, output analysis, simulation meta-models, time series, ARIMA models.

1 INTRODUCTION

Analyzing the results produced by simulation models is certainly an area of utmost practical importance. On the other hand, the highly autocorrelated nature of simulation responses, namely in queuing systems, has challenged simulation analysts to propose ever more innovative approaches.

The use of classical time series models (Box, Jenkins and Reinsel 1994) in the output analysis of stationary discrete event simulations was initially proposed by Fishman (1971). He suggested £tting an autoregressive model of order p, AR(p), to a simulation response, as an intermediate step for estimating reliable variability measures of the response (variance, con£dence intervals). Later, Schriber and Andrews (1984) generalized that approach and used an automated procedure for £tting mixed autoregressive-moving average models, ARMA(p,q). In both cases, the authors reported poor performance results, namely, in the coverage rates of con£dence intervals for simple queueing systems.

Contradicting the apparent incompatibility between queueing systems and time series models, Brandão and Porta Nova (1999) showed that most of those results could be related to either an insuf£cient simulation duration or an excessive initial bias. Keeping under control these two factors, very positive results were observed for an $M/M/3$ stationary queue under moderate and congested traf£c situations.

This paper is organized as follows. In Section 2, we discuss the use of classical time series models to analyze the output of non-stationary discrete event simulations. In Section 3, we investigate the applicability of $ARIMA$ models as potential meta-models for queuing system simulation under critical traf£c conditions. Finally, in Section 4, we draw some conclusions and suggest further work in this area.

2 NON-STATIONARY SIMULATION

The output analysis of non-stationary discrete event simulations is conspicuously absent from the literature of stochastic simulation. The well-known asymptotic result, for simple queueing systems, that most response measures go to in£nity when the utilization factor approaches one, seems to have convinced the simulation community that it was worthless to explore this topic. Uncontrolled evolution, explosive growth, are but two ways of characterizing a situation that has undoubted practical interest. During rush-hour periods, system breakdowns, etc., arrival rates do actually exceed processing rates. And much needed and useful information might be extracted from such non-stationary simulations, many ”what if?” questions could be answered... What is the expected queue length at the end of a rush-hour period? What is the expected sojourn time for an entity arriving halfway through that period? It would be even more interesting if we could predict the evolution in time of these and other performance measures, without having to repeat time consuming simulations and subsequent output analysis. This is the main purpose of pursuing simulation meta-models: £nding analytical models that are simpler but realistic representations of the computer programs implementing simulation models.

Since the output produced by stochastic simulation models is essentially composed of strongly autocorrelated time series, it seemed just natural to investigate the applicability of the classical time series models, $ARIMA(p, d, q)$, as potential meta-models for non-stationary simulations. In most scienti£c areas where the Box-Jenkins methodology is widely applied, data is scarcely available; that is the case, for instance, when econometric models are £t to a single realization of an economical time series. In contrast to that, we can chose any number of independent replications to analyze the output produced in discrete event simulation experiments. Thus, if we analyze the averaged responses across runs, instead of a single realization: (i) we reduce the time series variability; (ii) we are able to identify more clearly the *underlying* evolution of the response with respect to time; and (iii) we make the meta-model £tting process much easier. In addition, this is a valid £tting approach to autocorrelated data, contrarily to regression-based procedures. If the simulation responses are non-stationary in variance, it may be necessary to previously apply a variance-stabilizing transformation.

We illustrate our meta-modelling approach to non-stationary simulation with two case studies, two simple queues, with utilization factors greater than or equal to one, and we analyze the evolution of two responses: average queue length and average time spent in the system.

3 EXPERIMENTATION

In this section, we present and discuss the experimental evaluation of the applicability of $ARIMA(p, d, q)$ meta-models for the non-stationary simulation of two $M/M/s$ queueing systems—that is, with exponential inter-arrival and service times, and s identical parallel servers. We consider two distinct values for the utilization factor, $\rho = \lambda/(s\mu)$, where λ and μ represent the arrival and service rates, respectively: $\rho = 1$ (a critical traf£c situation) and $\rho = 2$ (super-critical traf£c).

We performed a Monte Carlo experiment consisting of 3000 independent replications of each simulation model with a *reference* duration of 60 time units for the followings queueing systems:

(i) An $M/M/1$ queue with super-critical traf£c ($\rho = 2$); and

(ii) An $M/M/2$ queue with critical traf£c ($\rho = 1$).

The actual duration varied, because each run was only terminated when the last entity that had arrived before 60 time units left the system. The initial conditions for each model were an empty and idle system.

Queue lengths were collected at regular time intervals $\Delta_t = 0.5, 1.0$, with $t \in [0, 60)$; sojourn times were sorted according to the time interval in which the arrival had occurred. Then, for each time interval, the corresponding observations across 30 runs were averaged. Finally, the Box-Jenkins methodology was applied to the averaged time series of each response, through the identi£cation, estimation and diagnostic checking of the $ARIMA(p, d, q)$ models.

Although we performed the experimentation for both interval widths, we only present the graphs corresponding to $\Delta_t = 0.5$, because the results are very similar.

3.1 A Super-critical $M/M/1$ Queue

The super-critical traf£c situation was created with an inter-arrival rate of $\lambda = 2$ and service rate of $\mu = 1$, producing the utilization factor $\rho = 2$.

Analyzing the average queue length for a typical set of 30 runs (see Figure 1, the hardly discernible dash-line), we can conclude that the process is non-stationary on the mean: there is a marked linear trend and the sample autocorrelation function (ACF) decreases very slowly to zero. In the £gure, the sample partial autocorrelation function is also represented. The results produced by the £tted model are also represented (solid line) on the subgraph containing the original series (dash-line). Differentiating the series (see again Figure 1), we observed that it became stationary, without any statistically signi£cative value on the ACF and PACF. This was also con£rmed applying the Box-Ljung test. Thus, in this case, we can £t the $ARIMA(0, 1, 0)$ model to the average queue length. This is an example of the simplest possible $ARIMA$ model, where the £rst difference produces white noise. This process is called a random walk with drift, if it has a nonzero expected value, or simply a random walk, otherwise.

We repeated the analysis for the remaining 99 time series (each corresponding to an average of 30 runs) obtaining the results reproduced in Table 1. In more than 90% of the cases, the same $ARIMA(0, 1, 0)$ model was consistently validated.

Table 1: Valid Fits for Average Queue Length ($M/M/1$ Queue, with $\rho = 2$)

	TIME INTERVAL	
FITTED MODEL	$\Delta_t = 0.5$	$\Delta_t = 1.0$
$ARIMA(0, 1, 0)$	91	93

The 100 original series are represented in Figure 2, as well as the corresponding £tted series. Comparing the two subgraphs, we see that both series have basically the same behavior.

The mean values of the differentiated data (for all the 100 series) are, approximately, 0.5 (for the case $\Delta_t = 0.5$) and 1 (for the case $\Delta_t = 1.0$). This suggests that the average queue length is directly proportional to the elapsed time. This important conclusion is in agreement with a little-known asymptotic result obtained by Bailey (1964) for the $M/M/1$ queue: the mean of the queue length at instant t can be approximated by $(\lambda - \mu)t$ for $\lambda > \mu$, thus validat-

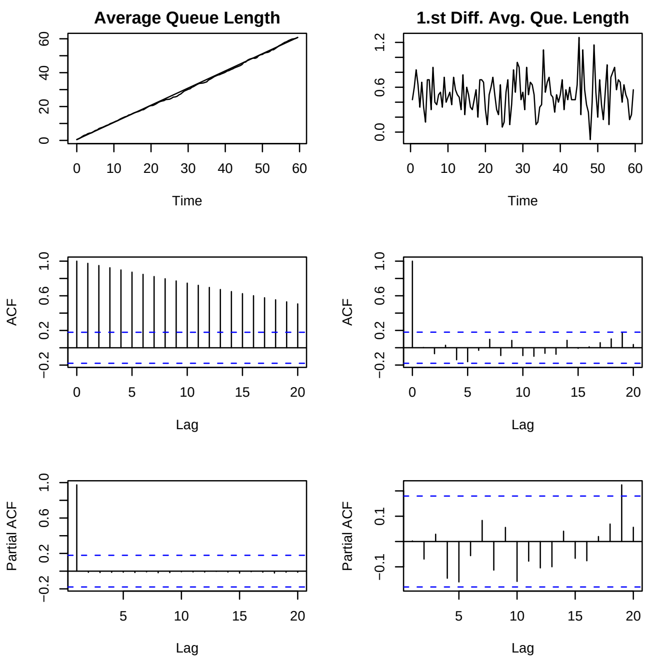

Figure 1: Average queue length ($M/M/1$, $\rho = 2$ and $\Delta_t = 0.5$)

ing the approach and the results presented here. In the same Figure 2, we also represent Bailey's result (the thick white line in the subgraph to the right).

A similar analysis was performed for the other response, the average sojourn time. In this case, the series were non-stationary on both the mean and the variance and it was necessary to perform a Box-Cox transformation to induce variance homogeneity. A logarithmic transformation was applied and then the £rst difference was computed. The original series, the £rst difference of the logarithmically transformed data and its corresponding ACF and PACF are represented in Figure 3 .

In this case, we tried to £t several $ARIMA(p, 1, q)$ models to the logarithms of the values of the 100 series. The corresponding results are reproduced in Table 2.

Table 2: Valid Fits for Average Sojourn Time ($M/M/1$ Queue, with $\rho = 2$)

FITTED MODEL	TIME INTERVAL $\Delta_t = 0.5$	$\Delta_t = 1.0$
$ARIMA(0, 1, 0)$	23	53
$ARIMA(0, 1, 1)$	49	66
$ARIMA(0, 1, 2)$	70	88
$ARIMA(0, 1, 3)$	89	93
$ARIMA(1, 1, 0)$	53	34
$ARIMA(2, 1, 0)$	62	38
$ARIMA(3, 1, 0)$	64	80

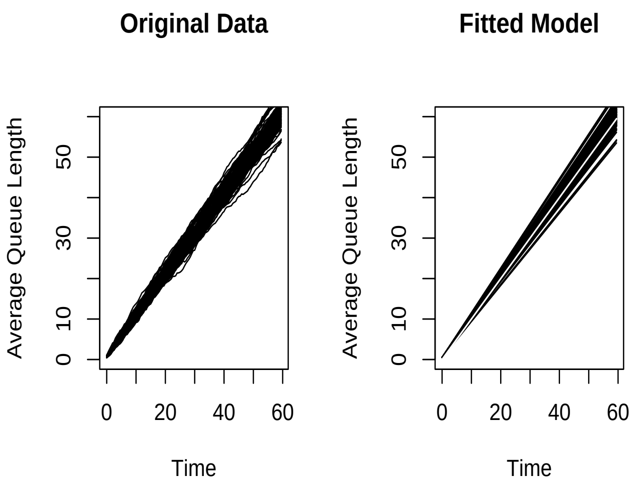

Figure 2: Original and Fitted Series: Average Queue Length ($M/M/1$ Queue, with $\rho = 2$ and $\Delta_t = 0.5$)

We concluded that the best $ARIMA(p, 1, q)$ model for representing the average sojourn time was the $ARIMA(0, 1, 3)$; that is, a moving average of order $q = 3$ is valid for approximately 90% of the differentiated series. Again, in Figure 4 we represent the original series and the corresponding £tted series. As before, the graphs are very

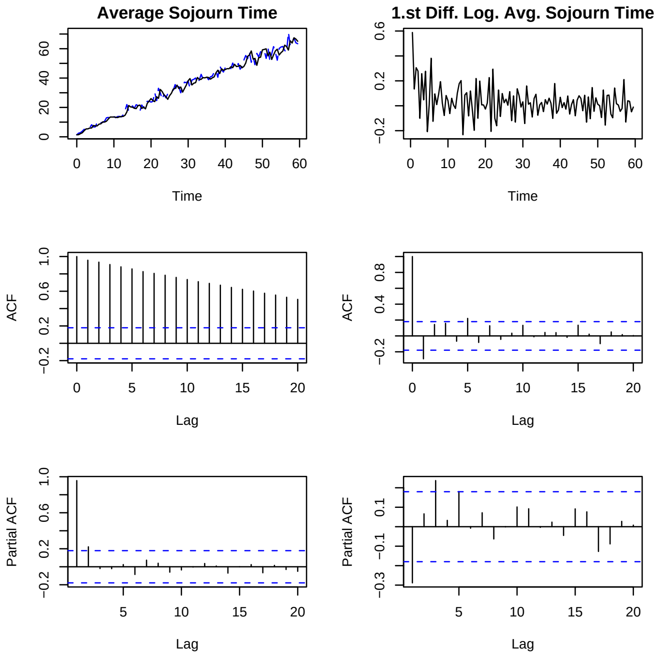

Figure 3: Average Sojourn Time ($M/M/1$ Queue, with $\rho = 2$ and $\Delta_t = 0.5$)

similar indicating a good £t.

3.2 A Critical $M/M/2$ Queue

To create, now, a situation with a critical value for the utilization factor, $\rho = \lambda/(s\mu)$, we chose an arrival rate of $\lambda = 2$ and a service rate of $\mu = 1$, resulting in $\rho = 1$ for the utilization factor.

Analyzing, for this case, the average queue length series (dash-line in Figure 5), we can observe, as before, a clear linear trend on the data; again, the ACF decreases very slowly to zero indicating a non-stationarity on the mean. In solid line we represent the results produced by the £tted model. Repeating the application of the Box-Jenkins methodology to the average queue length response, it was again possible to £t an $ARIMA(0,1,0)$ model.

Repeated analysis of the remaining series yielded the results reproduced in Table 3.

Table 3: Valid Fits for Average Queue Length ($M/M/2$ Queue, with $\rho = 1$)

FITTED MODEL	**TIME INTERVAL** $\Delta_t = 0.5$	$\Delta_t = 1.0$
$ARIMA(0,1,0)$	97	92

Again, we can conclude that the $ARIMA(0,1,0)$ is a good model for the average queue length. In Figure 6 we represent the original series and the corresponding £tted series.

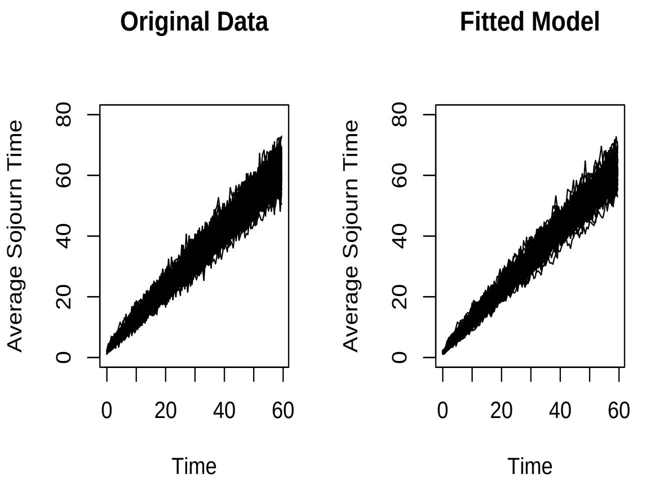

Figure 4: Original and Fitted Series: Average Sojourn Time ($M/M/1$ Queue, with $\rho = 2$ and $\Delta_t = 0.5$)

The results obtained for the average sojourn time in the $M/M/2$ queue, with $\rho = 1$, were very similar to those of the $M/M/1$ queue. Again, it was necessary to apply the logarithmic transformation and to differentiate the series in order to make it stationary (see Figure 7).

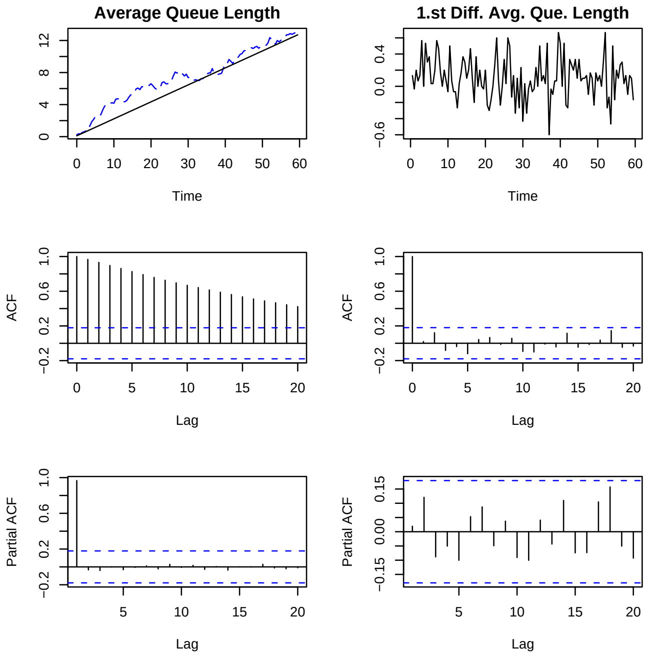

Figure 5: Average Queue Length ($M/M/2$ Queue, with $\rho = 1$ and $\Delta_t = 0.5$)

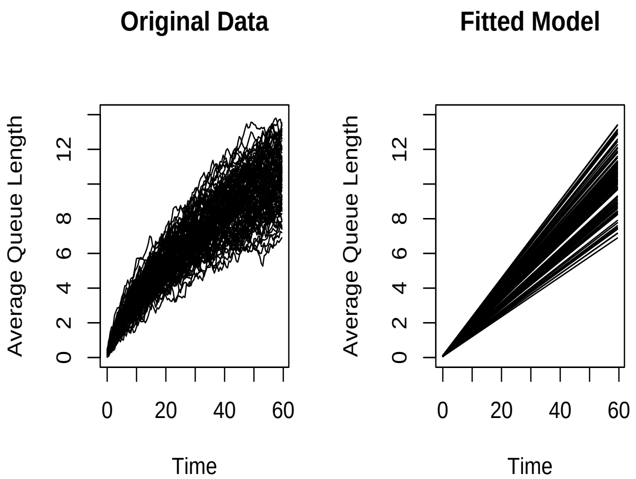

Figure 6: Original and Fitted Series: Average Queue Length ($M/M/2$ Queue, with $\rho = 1$ and $\Delta_t = 0.5$)

We then tried to £t the same type of $ARIMA(p, 1, q)$ models to the 100 series, obtaining the results presented in Table 4.

For $\Delta_t = 0.5$, we have two candidate models: the $ARIMA(0, 1, 2)$ and the $ARIMA(3, 1, 0)$ models. In Figure 8, we represent the original series, the $ARIMA(0, 1, 2)$

Table 4: Valid Fits for Average Sojourn Time ($M/M/2$ Queue, with $\rho = 1$)

	TIME INTERVAL	
FITTED MODEL	$\Delta_t = 0.5$	$\Delta_t = 1.0$
$ARIMA(0, 1, 0)$	0	45
$ARIMA(0, 1, 1)$	91	94
$ARIMA(0, 1, 2)$	96	99
$ARIMA(0, 1, 3)$	96	100
$ARIMA(1, 1, 0)$	36	96
$ARIMA(2, 1, 0)$	83	99
$ARIMA(3, 1, 0)$	98	99

£tted series and the $ARIMA(3, 1, 0)$ £tted series. We can conclude that the two £tted models behave similarly and represent quite well the original series.

4 CONCLUSIONS AND RECOMMENDATIONS

In this work, we propose an approach that is valid for meaningfully analyzing the output produced by non-stationary stochastic simulations. Based on what we might call *the classical time series method* for simulation output analysis, the approach can be used to obtain effective meta-models, namely, for queueing system simulation. A signi£cant experimental evaluation of our approach showed that it performed quite well for two queueing systems under critical traf£c conditions.

It is clear that much has to be done to develop and fundament the approach presented here. Also more comprehensive experimental evaluations, new examples of applications, other analytical models relating selected simulation responses with model parameters... However, we feel that this is a very promising area of undoubtedly practical interest, and we intend to continue exploring it.

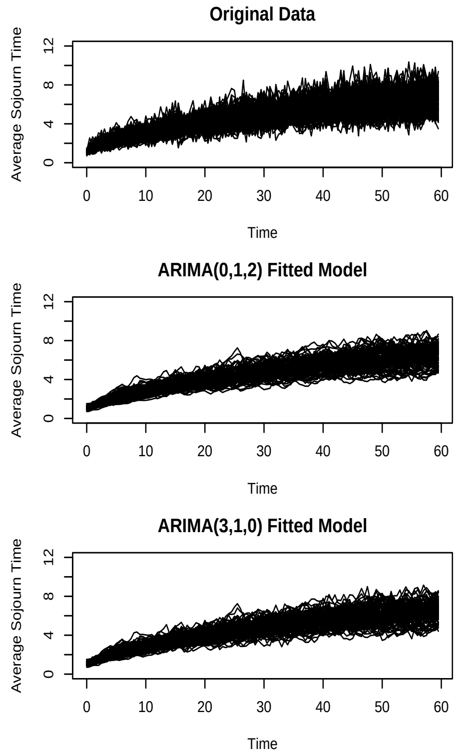

Figure 8: Original and Fitted Series: Average Sojourn Time ($M/M/2$ Queue, with $\rho = 1$ and $\Delta_t = 0.5$)

REFERENCES

Bailey, N.T.J. 1964. *The Elements of Stochastic Processes with Applications to the Natural Sciences*. John Wiley and Sons.

Box, G. E. P., G. M. Jenkins, and G. C. Reinsel. 1994. *Time series analysis: Forecasting and control*. 3d ed. Prentice Hall.

Brandão, R.M. and A.M Porta Nova. 1999. *An Experimental Evaluation of Methods for Simulation Output Analysis*. In *Proceedings of the 1999 European Simulation Symposium*(Horton, G., Möller, D. and Rüde, U., eds.), p. 601-607. The Society for Computer Simulation.

Fishman, G.S. 1971. Estimating Sample Size in Computing Simulation Experiments. *Management Science* 18 (1):21–38.

Ihaka, R. and R. Gentleman. 1996. R: A Language for Data Analysis and Graphics. *J. of Computational and Graphical Statistics* 5 (3):299–314.

Schriber, T. J. and R. W. Andrews. 1984. ARMA-Based Con£dence Intervals for Simulation Output Analysis. *Am. J. Math. Management Science* 4 (3 & 4):345–373.

AUTHOR BIOGRAPHIES

RITA M. BRANDÃO is a Lecturer and Ph. D. student of the Department of Mathematics at the University of Azores. Her e-mail address is `<rita@notes.uac.pt>`.

ACÁCIO M. O. PORTA NOVA is an Associate Professor of the Department of Engineering and Management at the Superior Technical Institute (IST) of the Technical University of Lisbon. He is a member of IIE. His e-mail address is `<apnova@ist.utl.pt>`.

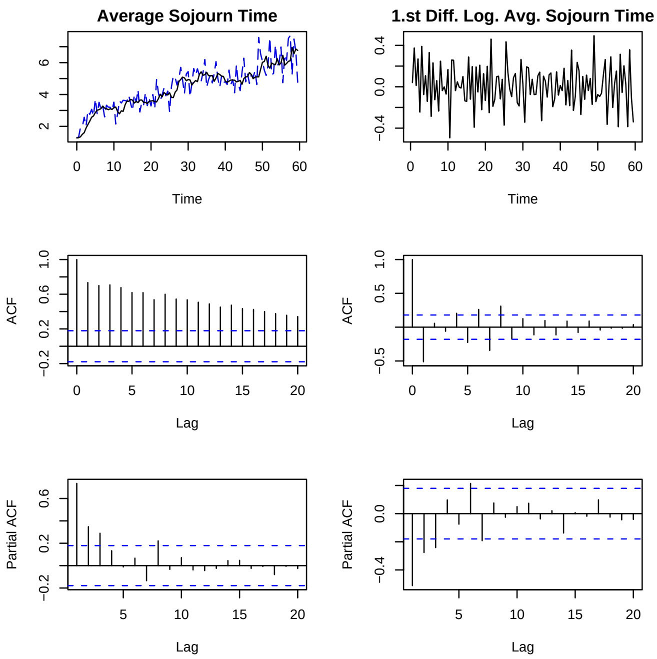

Figure 7: Average Sojourn Time ($M/M/2$ Queue, with $\rho = 1$ and $\Delta_t = 0.5$)