OUTPUT ANALYSIS OF CYCLICAL SIMULATIONS

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

KEYWORDS

Discrete event simulation, output analysis, simulation meta-models, time series, SARIMA models

ABSTRACT

In this paper, we present an approach for analyzing the output of non-terminating cyclical simulations. Deterministic components are estimated with nonlinear regression and Fourier series, while SARIMA models are used for the residuals. The approach is illustrated with a single lane traffic analysis. Some conclusions and suggestions for further work are stated.

1 INTRODUCTION

In virtually all simulation studies, the output analysis of discrete event simulation models is critically important. Indeed, the most perfectly crafted model can be rendered useless by a careless statistical analysis of the model-produced results. On the other hand, output analysis has been, as well, one of the most active research areas in simulation, for the last several decades. In fact, many innovative approaches have been proposed to extract useful information from the raw data generated by time-consuming computer programs. One such approach is the *classical time series* (Box, Jenkins and Reinsel 1994) method. Initially proposed by Fishman (1971), it did not receive much attention, except for Schriber and Andrews (1984). In both cases, the context was the steady-state analysis of stationary simulations. Recently, interest in this method revived; see Brandão and Porta Nova (2003a,b). In contrast to previous work, our research has been focused on the use of classical time series models for analyzing *non-stationary* simulations. But, simulations with periodic or cyclical behavior still remain a mostly unexplored topic in the specialized literature. One exception is a brief reference by Law and Kelton (2000) on how the method of independent replications could be used to obtain point and confidence interval estimates for cyclical means. However, many real-life systems of undeniable practical interest actually present some kind of cyclical behavior, namely in areas like traffic engineering, banking, inventory management and manufacturing.

This paper is organized as follows. In Section 2, we discuss how non-terminating cyclical simulations can be analyzed using seasonal time series models. In Section 3, we illustrate our approach, analyzing a traffic problem. Finally, in Section 4, we present some conclusions and suggestions for further work.

2 TIME SERIES FOR CYCLICAL SIMULATIONS

It is well known that the output produced by discrete event simulation models is highly autocorrelated. This has driven analysts to exploit this dependence structure through the use of classical time series models. These models were either used as a means of building confidence intervals for selected performance measures—Fishman (1971) and Schriber and Andrews (1984)—or as meta-models for representing the responses; see Brandão and Porta Nova (2003a,b). In our previous work, we used $ARIMA(p,d,q)$ models for analyzing non-stationary (terminating) simulations. In some situations, it was necessary to previously apply a variance-stabilizing transformation.

A special type of a non-stationary simulation occurs when the target system presents some periodic or cyclical behavior. For example, in traffic management, red and green lights alternate for each direction, and so, queue lengths and sojourn times will oscillate between some limits. Basing our analysis on simple averages will certainly be misleading... We propose building special-purpose meta-models for representing the expected evolution in time of the selected responses.

Our approach consists of the steps that follow.

(i) Averaging the response time series *across* runs.

In fact, we are not restricted to a single realization of the response time series, in contrast to what happens in econometrics, for instance. Thus, the underlying expected behavior of the response becomes more apparent, with a much smaller variability. We have found out that 30 replications is a good choice, from both the statistical and the experimental viewpoints; see Brandão and Porta Nova (2003b).

(ii) Meta-model identification.

We analyze the estimated autocorrelation (ACF) and partial autocorrelation (PACF) functions, in order to identify suitable transformations for achieving a stationary time series. Sometimes, the appropriate transformation involves removing a trend or cyclical *deterministic* component, before applying the Box-Jenkins methodology. If the transformed response still presents a periodic or cyclical behavior, we have to identify the period (s) and determine the two orders of differencing (seasonal, D, and non-seasonal, d) for the candidate $SARIMA$ models. Then, inspecting the estimated ACF and PACF of the transformed series, we select a candidate $ARMA(p,q) \times (P,Q)_s$ model. The orders P and Q are chosen analyzing the estimated ACF and PACF at lags which are multiples of s. The orders p and q are chosen analyzing the estimated ACF and PACF at lags 1,2,...$s-1$.

(iii) Meta-model estimation and diagnostic checking.

It is advisable to estimate the meta-model using an automated procedure. Ours is taken from the R Language (Ihaka and Gentleman 1996). The diagnostic checking of the model is done applying the portmanteau lack-of-fit test (with the modified Ljung-Box-Pierce statistic) and checking for correlation between the estimated parameters. For choosing between several candidate models, we use Akaike's information criterion (AIC).

(iv) Using the fitted meta-model.

The fitted meta-model allows us to evaluate the target system and answer many "what-if" questions, without further simulation. For instance, when a road is partially closed for repairs, where do we place the traffic lights, in order to accommodate waiting cars? In a worst case scenario, how long must an incoming driver wait to pass the affected section? In a rush hour period, at what rate will the waiting line increase?

3 A TRAFFIC ILLUSTRATION

Our approach to cyclical non-terminating simulation analysis is illustrated using a slightly modified version of the *Single Lane Traffic Analysis* classical example; see Nozari, Arnold and Pegden (1984) or Pritsker and O'Reilly (1999). In a two-lane road, with the traffic flowing from both directions, one of the lanes has been closed for 500 m, for repairs. At each end of the closed section, traffic lights were installed for controlling the traffic flow, allowing traffic (from only one direction) to flow for a specified time interval. When a light turns green, the waiting cars (from that direction) start moving and pass the light every two seconds. A car arriving when there are no waiting cars and the light is green, will pass through the light without delay. We assume that the time to pass through the repair zone is 60 sec. A light cycle consists of green in direction 1, both red, green in direction 2, both red, and then the cycle is repeated. Both lights remain red for 55 sec to allow the cars in transit to leave the section under repair, before traffic from the other direction is initiated. During most of the day, car inter-arrival times are exponentially distributed with an average of 12 sec from direction 1 and 9 sec from direction 2. However, by the end of the day, there is a one-hour rush period and the expected inter-arrival time from direction 2 is halved to 4.5 sec. In (Brandão and Porta Nova 2003b), we simulated green-light times at 5 sec intervals around 60 sec and concluded that the cycle that performed better was 55-50-55-60.

In this work, our objective is to investigate if it is feasible to build adequate meta-models for characterizing the cyclical behavior of the system, both under normal and congested conditions. Two responses of different nature will be analyzed: the average queue length and the average time spent in the system.

We performed 30 independent replications of the corresponding simulation models, with a *reference* duration of 3600 sec. The duration varied, because each run was only terminated when the last entity that had arrived before 3600 sec left the system. Initially, each model was started in a empty and idle state. Queue lengths (from one direction) were collected at regular 10 sec intervals, with $t = 10, 20, \ldots 3600$, starting at the beginning of the red light for the opposite direction. Then, the corresponding observations across runs were averaged. Sojourn times were sorted according to the time interval in which the arrival had occurred. Then, for each time interval, we collected only the first observation and the corresponding observations across runs were averaged. Additional observations in the same time interval were rejected to maintain independence. This also contributes to a more stable behavior of the response. Then, we tried several transformations for estimating and removing the trend and/or the cyclical *deterministic* components, using Fourier series and the nonlinear least squares method. For removing the trend component, we fitted a straight line. The cyclical *deterministic* component was approximated using an asymmetric triangular wave, expressed as a Fourier series. Finally, we applied the Box-Jenkins methodology to the residuals of the *deterministic* fit of each response (it will be referred as the *stochastic* component), through identification, estimation and diagnostic checking of the $SARIMA(p, d, q) \times (P, D, Q)_s$ models.

Applying our approach to the data corresponding to the normal conditions, we were able to fit an asymmetric triangular wave to the *deterministic* component and a $SARIMA(1, 0, 0) \times (0, 1, 1)_{22}$ model to the *stochastic* component of the average queue length for directions 1 and 2. Although we performed the experimentation for both directions, we only present the graphs corresponding to direction 2, because the results are very similar.

In Figure 1, the original series (dash-line) and the corresponding fitted *deterministic* series (solid blue line) of direction 2 are represented.

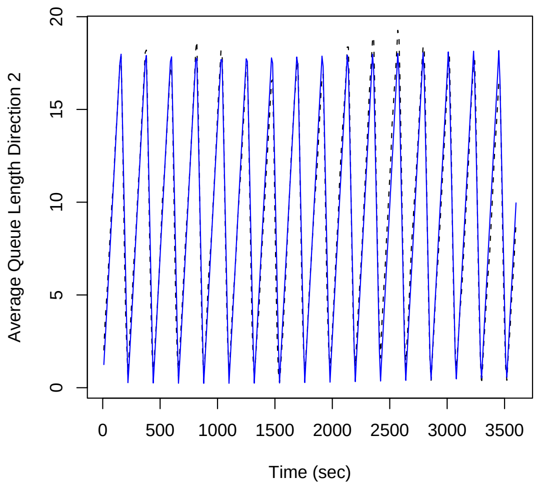

Figure 1: Original and Fitted Series: Average Queue Length, Direction 2

Comparing the two curves, we see that both series have basically the same behavior, indicating that the fitted *deterministic* model properly represents the original data series. A more detailed analysis (see Figure 2) reveals insignificant discrepancies between the two curves.

The fitted meta-model allows us to answer one of the

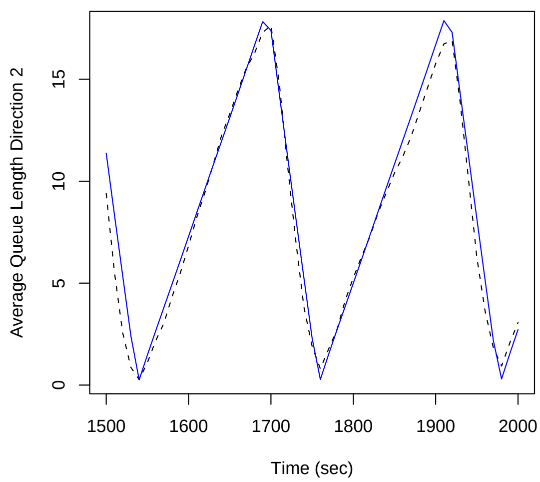

Figure 2: Original and Fitted Series: Average Queue Length, Direction 2 (1500 to 2000 sec)

questions formulated at the end of Section 2. The maximum value for the average queue length for direction 1 is approximately 14 and 18.5 for the direction 2. Considering that an average car length is 4.5 meters and that, for instance, 0.5 meters should be left between adjacent cars, then, at least, 70 and 93 meters must be left before the crossroads for directions 1 and 2, respectively, in order to avoid traffic jams.

We performed a similar analysis for the other response, the average sojourn time. Again, it was possible to fit an asymmetric triangular wave to the *deterministic* components for the two directions. With respect to the *stochastic* components, we fitted a $SARIMA(3, 0, 0) \times (0, 1, 1)_{22}$ model for direction 1 and a $SARIMA(2, 0, 0) \times (0, 1, 2)_{22}$ model for direction 2.

Except for a few extremal values, the global behavior of the original series is mostly captured (see Figure 3).

Even a more detailed scrutiny (see Figure 4) only shows insignificant discrepancies.

Having this meta-model available, we can answer the second question of Section 2. Under normal traffic conditions, the expected maximum sojourn time is approximately 235 and 222 for directions 1 and direction 2, respectively.

The one-hour rush period turns out to be a very interesting situation if the 55-50-55-60 cycle is left unchanged. Analyzing both system responses for direction 1 yielded similar results, although the fitted models for the *stochastic* components were distinct: a $SARIMA(1, 0, 1) \times (0, 1, 1)_{22}$ model for the average

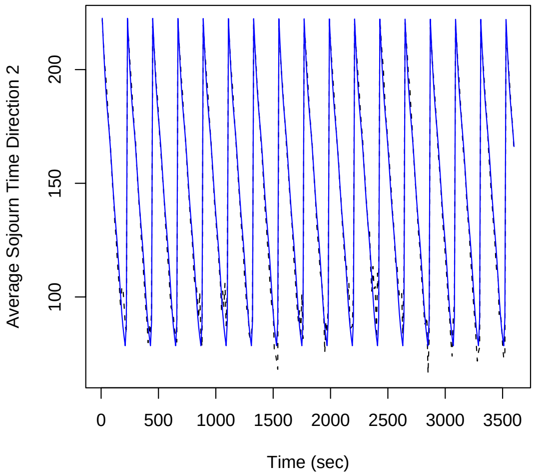

Figure 3: Original and Fitted Series: Average Sojourn Time, Direction 2

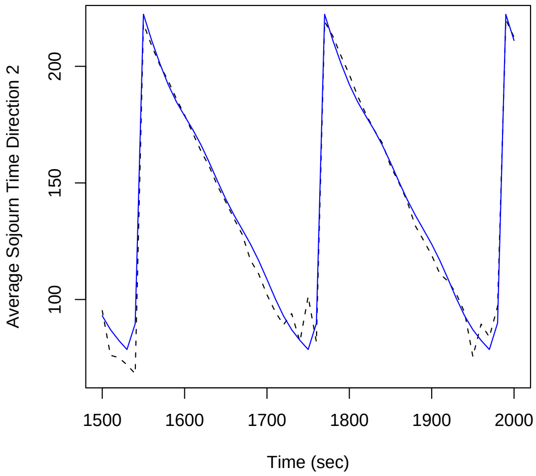

Figure 4: Original and Fitted Series: Average Sojourn Time, Direction 2 (1500 to 2000 sec)

queue length and a $SARIMA(2,0,0) \times (0,1,1)_{22}$ model for the average sojourn time.

However, when we consider direction 2, both responses grow without bounds; see Figures 5 and 6, respectively. This is no surprise, since, in this case, the corresponding utilization factor is greater than 1. We can see that the average queue length shows a linear trend, with a cyclic component superimposed on it. On the other hand, only a linear trend is visible for the average sojourn time. The fitted models for the *stochastic* component were a $SARIMA(0,1,0) \times (0,1,1)_{22}$ model for the average queue length and an $ARMA(3,1,1)$ model for the average sojourn time. Again, the fitted *deterministic* components keenly capture the essential behavior of the original series.

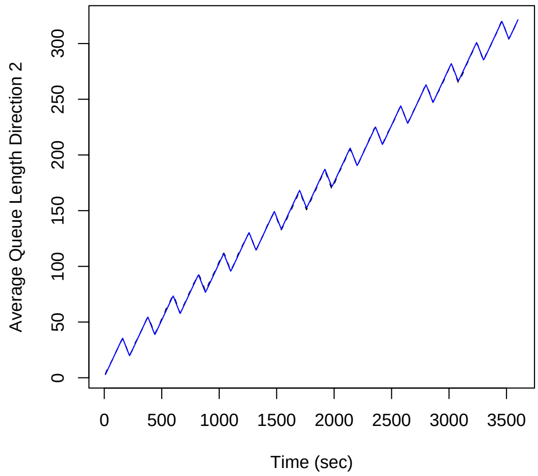

Figure 5: Original and Fitted Series: Average Queue length, Direction 2

These meta-models finally allow us to answer the last question of Section 2. Thus, under congested traffic, the average queue length for direction 2 linearly increases at a rate of about 0.086/sec (5.16/min), with an asymmetric triangular wave of approximate amplitude of 21.5 superimposed on it. The average sojourn time basically follows a straight line, with a slope of 0.63 and an intercept of 145 seconds, approximately.

4 CONCLUSIONS

Our recent work has been focused on extending the use of classical time series models to non-stationary simulations. In this paper, we present a very effective approach for analyzing cyclic behavior in non-terminating discrete event simulations. When applied to a classical example from the simulation literature, our approach was capable of producing reliable results of undeniable practical interest.

However, we feel that it is still possible to improve the approach and extend it to other classes of terminating or non-terminating simulations. Also, more extensive

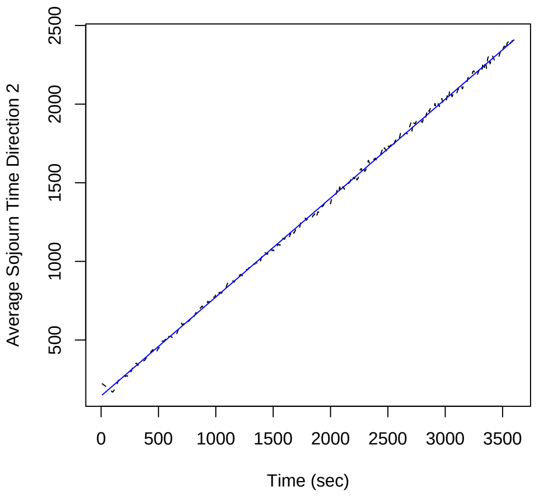

Figure 6: Original and Fitted Series: Average Sojourn Time, Direction 2

experimentations may be performed, and many other analytical results investigated...

REFERENCES

Box, G.E.P., Jenkins, G.M., and Reinsel, G.C. (1994) *Time Series Analysis: Forecasting and Control*, 3rd edn, Prentice Hall, Englewood Cliffs, NJ.

Brandão, R.M., and Porta Nova, A.M.O. (2003a) Non-stationary queue simulation analysis using time series, in *Proceedings of the 2003 Winter Simulation Conference*, Chick, S., Sánchez, P.J, Ferrin, D., and Morrice, D.J., (eds), Institute of Electrical and Electronics Engineers, Piscataway, NJ, 408–413.

Brandão, R.M., and Porta Nova, A.M.O. (2003b) Analysis of non-stationary stochastic simulations using time series, Working Paper 2/03, CEGIST, Lisbon (submitted to *IIE Transactions*).

Fishman, G.S. (1971) Estimating sample size in computing simulation experiments. *Management Science* **18**(1), 21–38.

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

Law, A.M. and Kelton, W.D. (2000) *Simulation Modeling and Analysis*, 3rd edn, McGraw-Hill, New York.

Nozari, A., Arnold, S.F., and Pegden, C.D. (1984) Control variates for multipopulation simulation experiments. *IIE Trans.* **16**,159–169.

Pritsker, A.A.B., and O'Reilly, J.J. (1999) *Simulation with Visual SLAM and Awesim*, 2sd edn, John Wiley & Sons, New York.

Schriber, T.J., and Andrews, R.W. (1984) ARMA-based confidence 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>`.