Proceedings of the 2010 Winter Simulation Conference
B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yücesan, eds.

DOES THE ERLANG C MODEL FIT IN REAL CALL CENTERS?

Thomas R. Robbins

East Carolina University
Department of Marketing and Supply Chain
3212 Bate Building
Greenville, NC 27858, USA

D. J. Medeiros

The Pennsylvania State University
Industrial and Manufacturing Engineering
310 Leonhard Building
University Park, PA 16802, USA

Terry P. Harrison

The Pennsylvania State University
Smeal College of Business
459 Business Building
University Park, PA 16802, USA

ABSTRACT

We consider the Erlang C model, a queuing model commonly used to analyze call center performance. Erlang C is a simple model that ignores caller abandonment and is the model most commonly used by practitioners and researchers. We compare the theoretical performance predictions of the Erlang C model to a call center simulation model where many of the Erlang C assumptions are relaxed. Our findings indicate that the Erlang C model is subject to significant error in predicting system performance, but that these errors are heavily biased and most likely to be pessimistic, *i.e.* the system tends to perform better than predicted. It may be the case that the model's tendency to provide pessimistic (*i.e.* conservative) estimates helps explain its continued popularity. Prediction error is strongly correlated with the abandonment rate so the model works best in call centers with large numbers of agents and relatively low utilization rates.

1 INTRODUCTION

A call center is a facility designed to support the delivery of some interactive service via telephone communications; typically an office space with multiple workstations manned by agents who place and receive calls (Gans, Koole et al. 2003). Call centers are a large and growing component of the U.S. and world economy and are estimated to employ approximately 2.1 million call center agents (Aksin, Armony et al. 2007). Large scale call centers are technically and managerially sophisticated operations and have been the subject of substantial academic research. The literature focused on call centers is quite large, with thorough and comprehensive reviews provided in (Gans, Koole et al. 2003) and (Aksin, Armony et al. 2007). Empirical analysis of call center data is given in (Brown, Gans et al. 2005).

Call centers are examples of queuing systems; calls arrive, wait in a virtual line, and are then serviced by an agent. Call centers are often modeled as M/M/N queuing systems, or in industry standard terminology - the Erlang C model. The Erlang C model makes many assumptions which are questionable in the context of a call center environment. Specifically the Erlang C model assumes that calls arrive at a known average rate, and that they are serviced by a defined number of statistically identical agents with service times that follows an exponential distribution. Most significantly, Erlang C assumes all callers wait as long as necessary for service without hanging up. The model is used widely by both practitioners and academics.

Recognizing the deficiencies of the Erlang C model, many recent papers have advocated using alternative queuing models and staffing heuristics which account for conditions ignored in the Erlang C model. The most popular alternative is the Erlang A model, a simple extension of the Erlang C model that allows for caller abandonment. For example, in a widely cited review of the call center literature (Gans, Koole et al. 2003) , the authors state "For this reason, we recommend the use of Erlang A as the standard to replace the prevalent Erlang C model." Another widely cited paper examines empirical data collected from a call center (Brown, Gans et al. 2005) and these authors make a similar statement; "using Erlang-A for capacity-planning purposes could and should improve operational performance. Indeed, the model is already beyond typical current practice (which is Erlang-C dominated), and one aim of this article is to help change this state of affairs."

The purpose of this study is to systematically analyze the fit of the Erlang C model in realistic call center situations. We seek to understand the nature and magnitude of the error associated with the model, and develop a better understanding of what factors influence prediction error.

The remainder of this paper is organized as follows. In Section 2 we review the Erlang C model and highlight the relevant literature. In section 3 we present a general model of a steady state call center environment and review the simulation model we developed to evaluate it. In section 4 we evaluate the performance of the Erlang C model. We conclude in Section 5 with summary observations and identify future research questions.

2 QUEUING MODELS AND THE ASSOCIATED LITERATURE

Queuing models are used to estimate system performance of call centers so that the appropriate staffing level can be determined to achieve a desired performance metric such as the Average Speed to Answer, or the Abandonment percentage. The most common queuing model used for inbound call centers is the Erlang C model (Gans, Koole et al. 2003; Brown, Gans et al. 2005). A Google search on "Erlang C Calculator" generates about 700,000 items including a large number of downloadable applications to calculate staffing requirements based on the Erlang C model.

The Erlang C model (M/M/N queue) is a very simple multi-server queuing system. Calls arrive according to a Poisson process at an average rate of λ. By nature of the Poisson process interarrival times are independent and identically distributed exponential random variables with mean λ^{-1}. Calls enter an infinite length queue and are serviced on a First Come – First Served (FCFS) basis. All calls that enter the queue are serviced by a pool of n homogeneous (statistically identical) agents at an average rate of $n\mu$. Service times follow an exponential distribution with a mean service time of μ^{-1}.

The steady state behavior of the Erlang C queuing model is easily characterized, see for example (Gans, Koole et al. 2003). The *offered load*, a unit-less quantity often referred to as the number of Erlangs, is defined as $R \equiv \lambda/\mu$. The *traffic intensity* (aka *utilization* or *occupancy*) is defined as $\rho \equiv \lambda/(N\mu) = R/N$.

Given the assumption that all calls are serviced, the traffic intensity must be strictly less than one or the system becomes unstable, *i.e.* the queue grows without bound. This system can be analyzed by solving a set of balance equations and the resulting steady state probability that all N agents are busy is

$$P\{Wait > 0\} = 1 - \left(\sum_{m=0}^{N-1} \frac{R^m}{m!}\right) \Bigg/ \left(\sum_{m=0}^{N-1} \frac{R^m}{m!} + \left(\frac{R^m}{N!}\right)\left(\frac{1}{1-R/N}\right)\right) \quad (1)$$

Equation (1) calculates the proportion of callers that must wait prior to service, an important measure of system performance. Another relevant performance measure for call centers managers is the *Average Speed to Answer* (ASA).

$$\begin{aligned} ASA \quad &= E[Wait] = P\{Wait > 0\} \cdot E[Wait \mid Wait > 0] \\ &= P\{Wait > 0\} \cdot \left(\frac{1}{N}\right) \cdot \left(\frac{1}{\mu_i}\right) \cdot \left(\frac{1}{1-\rho_i}\right) \end{aligned} \tag{2}$$

A third important performance metric for call center managers is the *Telephone Service Factor* (TSF), also called the "service level." The TSF is the fraction of calls presented which are eventually serviced and for which the delay is below a specified level. For example, a call center may report the TSF as the percent of callers on hold less than 30 seconds. The TSF metric can then be expressed as

$$\begin{aligned} TSF \quad &= P\{Wait \le T\} = 1 - P\{Wait > 0\} \cdot P\{Wait > T \mid Wait > 0\} \\ &= 1 - C(N, R_i) \cdot e^{-N\mu_i(1-\rho_i)T} \end{aligned} \tag{3}$$

A fourth performance metric monitored by call center managers is the *Abandonment Rate*; the proportion of all calls that leave the queue (hang up) prior to service. Abandonment rates cannot be estimated directly using the Erlang C model because the model assumes no abandonment occurs.

A substantial amount of research analyzes the behavior of Erlang C model, much of it seeks to establish simple staffing heuristics based on asymptotic frameworks applied to large call centers. (Halfin and Whitt 1981) develop a formal version of the square root staffing principle for M/M/N queues in what has become known as the Quality and Efficiency Driven (QED) regime. (Borst, Mandelbaum et al. 2004) develop a framework for asymptotic optimization of a large call center with no abandonment.

As is the case with any analytical model, the Erlang C model makes many assumptions, several of which are not wholly accurate. In the case of the Erlang C model several assumptions are questionable, but clearly the most problematic is the no abandonment assumption, as even low levels of abandonment can dramatically impact system performance (Gans, Koole et al. 2003). Many call center research papers however analyze call center characteristics under the assumption of no abandonment (Jennings and Mandelbaum 1996; Green, Kolesar et al. 2001; Green, Kolesar et al. 2003; Borst, Mandelbaum et al. 2004; Wallace and Whitt 2005; Gans and Zhou 2007).

The Erlang C model assumes also that calls arrive according to a Poisson process. The interarrival time is a random variable drawn from an exponential distribution with a known arrival rate. Several authors assert that the assumption of a known arrival rate is problematic. Both major call center reviews (Gans, Koole et al. 2003; Aksin, Armony et al. 2007) have sections devoted to arrival rate uncertainty. (Brown, Gans et al. 2005) perform a detailed empirical analysis of call center data. While they find that a time-inhomogeneous Poisson process fits their data, they also find that arrival rate is difficult to predict and suggest that the arrival rate should be modeled as a stochastic process. Many authors argue that call center arrivals follow a doubly stochastic process, a Poisson process where the arrival rate is itself a random variable (Chen and Henderson 2001; Whitt 2006; Aksin, Armony et al. 2007). Arrival rate uncertainty may exist for multiple reasons. Arrivals may exhibit randomness greater than that predicted by the Poisson process due to unobserved variables such as the weather or advertising. Call center managers attempt to account for these factors when they develop forecasts, yet forecasts may be subject to significant error. (Robbins 2007) compares four months of week-day forecasts to actual call volume for 11 call center projects. He finds that the average forecast error exceeds 10% for 8 of 11 projects, and 25% for 4 of 11 projects. The standard deviation of the daily forecast to actual ratio exceeds 10% for all 11 projects. (Steckley, Henderson et al. 2009) compare forecasted and actual volumes for nine weeks of data taken from four call centers. They show that the forecasting errors are large and modeling arrivals as a Poisson process with the forecasted call volume as the arrival rate can introduce significant error. (Robbins, Medeiros et al. 2006) use simulation analysis to evaluate the impact of forecast error on performance measures demonstrating the significant impact forecast error can have on system performance.

Some recent papers address staffing requirements when arrival rates are uncertain. (Bassamboo, Harrison et al. 2005) develop a model that attempts to minimize the cost of staffing plus an imputed cost for

customer abandonment for a call center with multiple customer and server types when arrival rates are variable and uncertain. (Harrison and Zeevi 2005) use a fluid approximation to solve the sizing problem for call centers with multiple call types, multiple agent types, and uncertain arrivals. (Whitt 2006) allows for arrival rate uncertainty as well as uncertain staffing, i.e. absenteeism, when calculating staffing requirements. (Steckley, Henderson et al. 2004) examine the type of performance measures to use when staffing under arrival rate uncertainty. (Robbins and Harrison 2010)develop a scheduling algorithm using a stochastic programming model that is based on uncertain arrival rate forecasts.

The Erlang C model also assumes that the service time follows an exponential distribution. The memoryless property of the exponential distribution greatly simplifies the calculations required to characterize the system's performance, and makes possible the relatively simple equations (1)-(3). If the assumption of exponentially distributed talk time is relaxed, the resulting queuing model is the $M/G/N$ queue, which is analytically intractable (Gans, Koole et al. 2003) and approximations are required. However, empirical analysis suggests that the exponential distribution is a relatively poor fit for service times. Most detailed analysis of service time distributions find that the lognormal distribution is a better fit (Mandelbaum, Sakov et al. 2001; Gans, Koole et al. 2003; Brown, Gans et al. 2005).

Finally, the Erlang C model assumes that agents are homogeneous. More precisely, it is assumed that the service times follow the same statistical distribution independent of the specific agent handling the call. Empirical evidence supports the notion that some agents are more efficient than others and the distribution of call time is dependent on the agent to whom the call is routed. In particular more experienced agents typically handle calls faster than newly trained agents (Armony and Ward 2008). (Robbins 2007) demonstrated a statistically significant learning curve effect in an IT help desk environment.

3 CALL CENTER SIMULATION

3.1 The Modified Model

In this section we present a revised model of a call center, relaxing key assumptions discussed previously. In our model calls arrive at the call center according to a Poisson process. Calls are forecasted to arrive at an average rate of $\hat{\lambda}$. The realized arrival rate is λ, where λ is a normally distributed random variable with mean $\hat{\lambda}$, standard deviation σ_λ and coefficient of variation $c_\lambda = \sigma_\lambda / \hat{\lambda}$. The choice of the normal distribution gives us a symmetric distribution centered on the forecasted value. A disadvantage of the normal distribution is the possibility of generating negative values. However, in our experiments the mean value is sufficiently positive, a minimum of 5 standard deviations, that this is not a concern. The time required to process a call by an average agent is a lognormally distributed random variable with mean μ^{-1} and standard deviation σ_μ. Arriving calls are routed to the agent who has been idle for the longest time if one is available. If all agents are busy the call is place in a FCFS queue. When placed in queue a proportion of callers will balk; *i.e.* immediately hang up. Callers who join the queue have a patience time that follows a Weibull distribution. If wait time exceeds their patience time the caller will abandon. Calls are serviced by agents who have variable relative productivity r_i. Agent productivity is assumed to be a normally distributed random variable with a mean of 1 and a standard deviation of σ_r. An agent with a relative productivity level of 1, for example, serves calls at the average rate. An agent with a relative productivity level of 1.5 serves calls at 1.5 times the average rate, an agent with a productivity level of .75 serves calls at .75 times the average rate. Given the mean productivity level of 1, on average calls are served at the rate λ.

3.2 Experimental Design

In order to evaluate the performance of the Erlang C against the simulation model we conduct a series of designed experiments. Based on the assumptions for our call center discussed previously, we define the following set of nine experimental factors.

Table 1: Experimental Factors

	Factor	Low	High
1	Number of Agents	10	100
2	Offered Utilization ($\hat{\rho}$)	65%	95%
3	Talk Time (mins)	2	20
4	Patience β	60	600
5	Forecast Error CV (c_λ)	0	.2
6	Patience α	.75	1.25
7	Talk time CV	.75	1.25
8	Probability of Balking	0	.25
9	Agent Productivity Standard Deviation	0	.15

The forecasted arrival rate in the simulation is a quantity derived from other experimental factors by

$$\hat{\lambda} = \hat{\rho} N \mu \tag{4}$$

Given the relatively large number of experimental factors, a well designed experimental approach is required to efficiently evaluate the experimental region. A standard approach to designing computer simulation experiments is to employ either a full or fractional factorial design (Law 2007). However, the factorial model only evaluates corner points of the experimental region and implicitly assumes that responses are linear in the design space. Given the anticipated non-linear relationship of errors we chose to implement a Space Filling Design based on Latin Hypercube Sampling (LHS) as discussed in (Santner, Williams et al. 2003). Given a desired sample of *n* points, the experimental region is divided into n^d cells. A sample of *n* cells is selected in such a way that the centers of these cells are uniformly spread when projected onto each axis of the design space. While the LHS design is not perfectly orthogonal like a factorial design, the design does provide for a low correlation between input factors greatly reducing the risk of multicollinearity. We chose our design point as the center of each selected cell.

3.3 Simulation Model

The model is evaluated using a straightforward discrete event simulation model. The purpose of the model is to predict the long term, steady state behavior of the queuing system. The model generates random numbers using the a combined multiple recursive generator (CMRG) based on the Mrg32k3a generator described in (L'Ecuyer 1999). Common random numbers are used across design points to reduce output variance. To reduce any start up bias we use a warm up period of 5,000 calls, after which all statistics are reset. The model is then run for an evaluation period of 25,000 calls and summary statistics are collected. For each design point we repeat this process for 500 replications and report the average value across replications.

The specific process for each replication is as follows. The input factors are chosen based on the experimental design. The average arrival rate is calculated based on the specified talk time, number of agents, and offered utilization rate according to equation (4). A random number is drawn and the realized arrival rate is set based on the probability distribution of the forecast error. That arrival rate is then used to generate Poisson arrivals for the replication. Agent productivities are generated using a normal distribution with mean one and standard deviation σ_p. Each new call generated includes an exponentially distributed interarrival time, a lognormally distributed average talk time, a Weibull distributed time before abandonment, and a Bernoulli distributed balking indicator. When the call arrives it is assigned to the longest idle agent, or placed in the queue if all agents are busy. If sent to the queue the simulation model checks the balking indicator. If the call has been identified as a balker it is immediately abandoned, if not an abandonment event is scheduled based on the realized time to abandon. Once the call has been assigned to an agent, the realized talk time is calculated by multiplying the average talk time and the agent's productivity. The agent is committed for the realized talk time. When the call completes the

agent processes the next call from the queue, or if no calls are queued becomes idle. If a call is processed prior to its time to abandon, the abandonment event is cancelled. If not, the call is abandoned and removed from the queue when the patience time expires.

After all replications of the design point have been executed the results are compared to the theoretical predictions of the Erlang C model. We calculate the error as the difference between the theoretical value and the simulated value. We make a relative error calculation so that the sign of the error indicates the bias in the calculation. In our experiment we evaluated an LHS sample of 1,000 points.

4 EXPERIMENTAL ANALYSIS

4.1 Summary Observations

Based on our analysis we can make the following summary observations:

- The Erlang C model is, on average, subject to a reasonably large error over this range of parameter values.
- Measurement errors are highly positively correlated across performance measures.
- The Erlang C model is on average pessimistically biased (the real system performs better than predicted) but may become optimistically biased when utilization is high and arrival rates are uncertain.
- Measurement error is high when the real system exhibits higher levels of abandonment. The error is strongly positively correlated with realized abandonment rate and predicted ASA.
- The Erlang C model is most accurate when the number of agents is large and utilization is low.
- Errors decrease as caller patience increases.

We will now review our experimental results in more detail.

4.2 Correlation and Magnitude of Errors

The magnitude of errors generated by using the Erlang C model across our test space is high on average, and very high in some cases. The errors across the key metrics are highly correlated with each other, and highly correlated with the realized abandonment rate. Table 2 shows a correlation matrix of the errors generated from the Erlang C model.

Table 2: Error Correlation Matrix

	Simulated Abandonment Rate	*Prob Wait Error*	*ASA Error*	*TSF Error*	*Utilization Error*
Simulated Abandonment Rate	1.000				
Prob Wait Error	.867	1.000			
ASA Error	.766	.722	1.000		
TSF Error	-.880	-.987	-.759	1.000	
Utilization Error	.970	.861	.745	-.873	1.000

Correlations between measure errors are strong. The measured errors all move, on average, in an optimistic or pessimistic direction together. ProbWait and ASA are positively correlated; it is desirable for both these measures to be low. ProbWait is negatively correlated with TSF; a measure for which a high value is desirable. Measurement error is also highly correlated with abandonment rate. Given the high correlation between measures we will utilize ProbWait as a proxy for the overall error of the Erlang C model.

Average error rates are reasonably high under the Erlang C model, with errors being pessimistically skewed. Figure 1 shows a histogram of the ProbWait error.

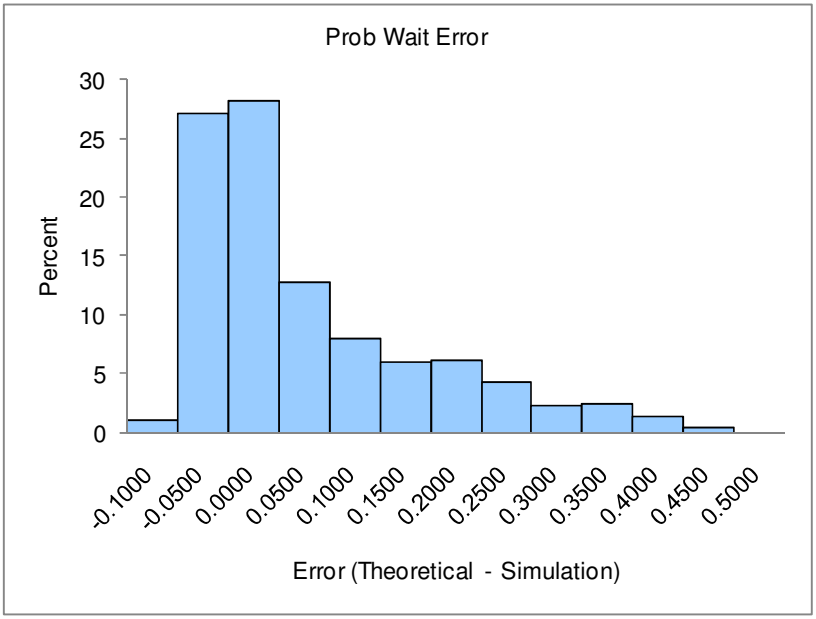

Figure 1: Histogram of Erlang C Prob Wait Errors

The average error is 7.96%, and the data has a strong positive skew; 72% of the errors being positive. The largest error is 49.4%, the smallest is -8.0 %.

4.3 Drivers of Erlang C Error

Having established that error rates are high under the Erlang C model, we now turn out attention to characterizing the drivers of that error. As discussed in the previous section, Erlang C errors are highly correlated with the realized abandonment rate. The notion that abandonment is a major driver of errors in the Erlang C model is further illustrated in Figure 2. This graph shows the error in the ProbWait measure on the vertical axis and the abandonment rate from the simulation analysis on the horizontal axis.

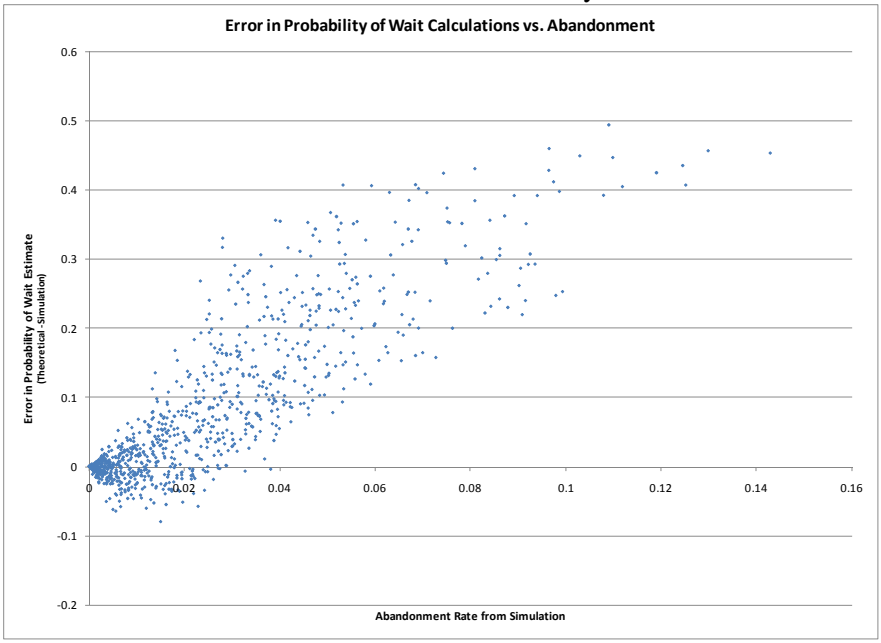

Figure 2: Scatter Plot of Erlang C Errors and Abandonment Rate

The graph clearly shows that as abandonment increases, the error in the ProbWait measure increases as well. The graph also reveals that optimistic errors, i.e. errors in which the system performed worse than predicted, only occur with relatively low abandonment rates. The average abandonment rate for optimistic predictions was .74%. The graph also reveals that significant error can be associated with even low to moderate abandonment rates. For example, for all test points with abandonment rates of less than 5%, the average error for ProbWait is 4.8%. For test points in which abandonment ranged between 2% and 5% the average ProbWait error is 12.2%.

To assess how each of the nine experimental factors impacts the error, we perform a regression analysis. The dependent variable is the ProbWait error. For the independent variable we use the nine experimental factors normalized to a [-1,1] scale. This normalization allows us to better assess the relative impact of each factor. The LHS sampling method provides an experimental design where the correlation between experimental factors is low, greatly reducing risks of multicollinearity. The results of the regression analysis are shown in Table 3.

Table 3: Regression Analysis of ProbWait Error

Regression Analysis

R^2	0.746		
Adjusted R^2	0.744	n	1000
R	0.864	k	9
Std. Error	0.058	Dep. Var.	**Prob Wait Error**

ANOVA table

Source	*SS*	*df*	*MS*	*F*	*p-value*
Regression	9.6689	9	1.0743	323.87	8.38E-288
Residual	3.2839	990	0.0033		
Total	12.9529	999			

Regression output					*confidence interval*	
variables	*coefficients*	*std. error*	*t (df=990)*	*p-value*	*95% lower*	*95% upper*
Intercept	0.0797	0.0018	43.745	1.52E-233	0.0761	0.0832
Num Agents	-0.0721	0.0032	-22.778	1.11E-92	-0.0783	-0.0658
Utilization Target	0.1500	0.0032	47.365	1.09E-256	0.1438	0.1562
Talk Time	0.0184	0.0032	5.829	7.53E-09	0.0122	0.0246
Patience	-0.0134	0.0032	-4.233	2.52E-05	-0.0196	-0.0072
AR CV	-0.0260	0.0032	-8.206	7.05E-16	-0.0322	-0.0198
Talk Time CV	-0.0035	0.0032	-1.096	.2734	-0.0097	0.0027
Patience Shape	-0.0027	0.0032	-0.858	.3912	-0.0089	0.0035
Probability of Balking	0.0228	0.0032	7.172	1.44E-12	0.0165	0.0290
Agent Heterogeneity	0.0050	0.0032	1.585	.1133	-0.0012	0.0112

Given the normalization of the experimental factors, the magnitude of the regression coefficients provides a direct assessment of the impact that a factor has on the measurement error. The factor that most strongly influences the error is the offered utilization, the magnitude of its coefficient being more than twice the value of the next measure and more than five times the magnitude of all other factors. The size of the call center, measured as the number of agents, has a major impact on errors. Factors related to willingness to wait, *i.e.* Patience, Patience Shape, and Probability of Balking, all have low to moderate impacts, but with exception of Patience Shape are statistically significant. Talk time is also a statistically significant factor with a moderate impact. The variability of talk time and agent heterogeneity both have low impacts that are not statistically significant.

The most important drivers of Erlang C errors are the size and utilization of the call center. This is further illustrated in Figure 3. This graph shows the results of an experiment where the number of agents and utilization factors are varied in a controlled fashion. All other experimental factors are held at their mid-point.

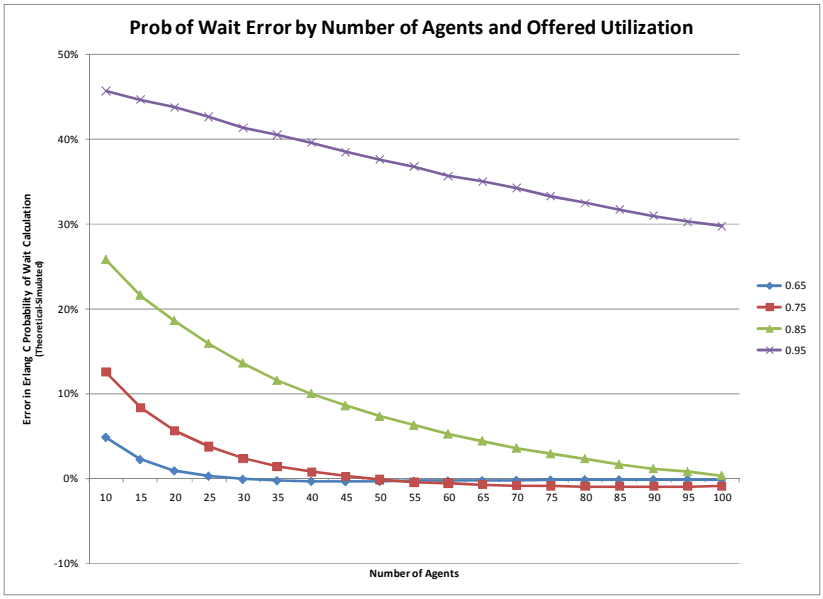

Figure 3: Erlang C ProbWait Errors by Call Center Size and Utilization

This graph demonstrates that the Erlang C model tends to provide relatively poor predictions for small call centers. This error tends to decrease as the size of the call center increases. However, the graph also illustrates that for busy centers the error remains high. For a very busy call center, running at 95% offered utilization, the error rate remains at 30%, even with a pool of 100 agents. The errors tend to track with abandonment; abandonment rates increase with utilization and decrease with the agent pool.

The conclusion that abandonment behavior drives the Erlang C error is further illustrated in Figure 4. In this experiment we systematically vary two Willingness to Wait parameters. Specifically, we vary the balking probability and the β factor of patience distribution.

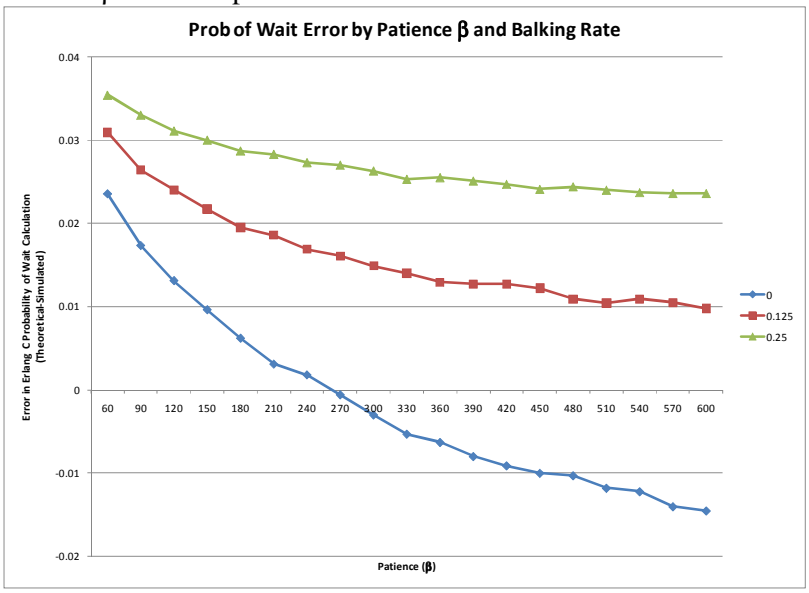

Figure 4: Erlang C ProbWait Errors by Willingness to Wait

This analysis verifies that the more likely callers are to balk, the higher the error rate. The analysis also shows that when callers are more patient, the error rates decrease. The more likely callers are to abandon, either immediately or soon after being queued, the higher the abandonment rate and the less accurate the Erlang C measures become.

An additional factor of interest is the uncertainty associated with the arrival rate. While it's overall effect is not large, about 1.8%, it has effects that are dissimilar to other experimental factors as illustrated in Figure 5. This graph shows the results of an experiment that varies the coefficient of variation of the arrival rate error and the number of agents while holding all other factors at their mid-points.

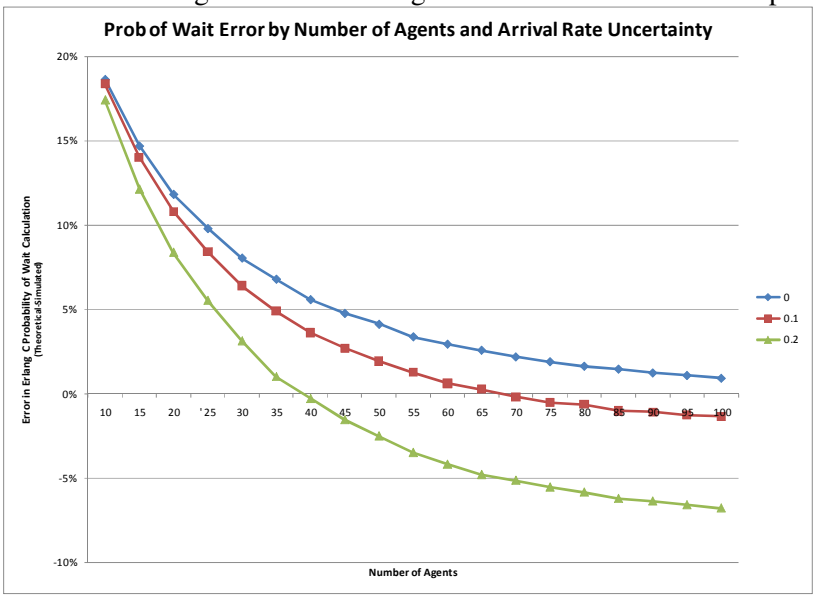

Figure 5: Erlang C ProbWait Errors by Call Center Size and Forecast Error

This experiment shows that for small call centers arrival rate uncertainty has a small effect, but that effect becomes more pronounced for larger call centers. It is also worth noting that arrival rate uncertainty has an optimistic effect, and for high levels of uncertainty the model exhibits an optimistic bias. Arrival rate uncertainty is a major factor leading to an optimistic estimate from the Erlang C model; of the 21.9% of test points with an optimistic bias the average arrival rate uncertainty cv was 14.0%. Since arrival rate uncertainty tends to bias the prediction in the opposite direction of most other factors, it also has the effect of reducing error in many situations. For example, high utilization tends to bias the estimate pessimistically, a bias reduced when arrival rate uncertainty is present.

5 SUMMARY AND CONCLUSIONS

The Erlang C model is commonly applied to predict queuing system behavior in call center applications. Our analysis shows that when we test the Erlang C model over a range of reasonable conditions predicted performance measures are subject to large errors. The Erlang C model works reasonably well for large call centers with low to moderate utilization rates, but factors that tend to generate caller abandonment; such as high utilization, small agent pools, and impatient callers, cause the model error to become quite large. While the model tends to provide a pessimistic estimate, arrival rate uncertainty will either reduce that bias or lead to a optimistic bias. It may be the case that the model's tendency to provide pessimistic (*i.e.* conservative) estimates helps explain its continued popularity. It is clear that great care must be taken before using the Erlang C model to make any calculations that require a high level of precision.

Our future research is focused on analyzing the increasingly popular Erlang A model and comparing it's performance to the Erlang C model to test the growing consensus that Erlang A is a superior model for call center analysis.

REFERENCES

Aksin, Z., M. Armony and V. Mehrotra. 2007. The Modern Call-Center: A Multi-Disciplinary Perspective on Operations Management Research. *Production and Operations Management* 16: 665-668.

Armony, M. and A. R. Ward 2008. Fair Dynamic Routing in Large-Scale Heterogeneous-Server Systems, Stern School of Business, NYU.

Bassamboo, A., J. M. Harrison and A. Zeevi. 2005. Design and Control of a Large Call Center: Asymptotic Analysis of an LP-based Method. *Operations Research* 54: 419-435.

Borst, S., A. Mandelbaum and M. I. Reiman. 2004. Dimensioning Large Call Centers. *Operations Research* 52: 17-35.

Brown, L., N. Gans, A. Mandelbaum, et al. 2005. Statistical Analysis of a Telephone Call Center: A Queueing-Science Perspective. *Journal of the American Statistical Association* 100: 36-50.

Chen, B. P. K. and S. G. Henderson. 2001. Two Issues in Setting Call Centre Staffing Levels. *Annals of Operations Research* 108: 175-192.

Gans, N., G. Koole and A. Mandelbaum. 2003. Telephone call centers: Tutorial, review, and research prospects. *Manufacturing & Service Operations Management* 5: 79-141.

Gans, N. and Y.-P. Zhou. 2007. Call-Routing Schemes for Call-Center Outsourcing. *Manufacturing & Service Operations Management* 9: 33-51.

Green, L. V., P. Kolesar and J. Soares. 2003. An Improved Heuristic for Staffing Telephone Call Centers with Limited Operating Hours. *Production and Operations Management* 12: 46-61.

Green, L. V., P. J. Kolesar and J. Soares. 2001. Improving the SIPP Approach for Staffing Service Systems That Have Cyclic Demands. *Operations Research* 49: 549-564.

Halfin, S. and W. Whitt. 1981. Heavy-Traffic Limits for Queues with Many Exponential Servers. *Operations Research* 29: 567-588.

Harrison, J. M. and A. Zeevi. 2005. A Method for Staffing Large Call Centers Based on Stochastic Fluid Models. *Manufacturing & Service Operations Management* 7: 20-36.

Jennings, O. B. and A. Mandelbaum. 1996. Server staffing to meet time-varying demand. *Management Science* 42: 1383.

L'Ecuyer, P. 1999. Good Parameters and Implementations for Combined Multiple Recursive Random Number Generators. *Operations Research* 47: 159-164.

Law, A. M. 2007. *Simulation modeling and analysis*. Boston, McGraw-Hill.

Mandelbaum, A., A. Sakov and S. Zeltyn 2001. Empirical Analysis of a Call Center, Technion - Israel Institute of Technology.

Robbins, T. R. 2007. Managing Service Capacity Under Uncertainty - Unpublished PhD Dissertation Pennsylvania State University. University Park, PA. Avaialble via (http://personal.ecu.edu/robbinst/) [accessed August 31, 2010]

Robbins, T. R. and T. P. Harrison. 2010. Call Center Scheduling with Uncertain Arrivals and Global Service Level Agreements. *European Journal of Operational Research* Forthcoming.

Robbins, T. R., D. J. Medeiros and P. Dum 2006. *Evaluating Arrival Rate Uncertainty in Call Centers*. 2006 Winter Simulation Conference, Monterey, CA.

Santner, T. J., B. J. Williams and W. Notz 2003. *The design and analysis of computer experiments*. New York, Springer.

Steckley, S. G., S. G. Henderson and V. Mehrotra. 2009. Forecast Errors in Service Systems. *Probability in the Engineering and Informational Sciences*: 305-332.

Steckley, S. G., W. B. Henderson and V. Mehrotra 2004. Service System Planning in the Presence of a Random Arrival Rate, Cornell University.

Wallace, R. B. and W. Whitt. 2005. A Staffing Algorithm for Call Centers with Skill-Based Routing. *Manufacturing & Service Operations Management* 7: 276-294.

Whitt, W. 2006. Staffing a Call Center with Uncertain Arrival Rate and Absenteeism. *Production and Operations Management* 15: 88-102.

AUTHOR BIOGRAPHIES

THOMAS R. ROBBINS is an Assistant Professor in the department of Marketing and Supply Chain at East Carolina University. He holds a PhD in Business Administration and Operations Research from Penn State University, an MBA from Case Western Reserve and a BSEE from Penn State. Prior to beginning his academic career he worked in professional services for approximately 18 years. His email address is <robbinst@ecu.edu>.

D. J. MEDEIROS is Associate Professor of Industrial Engineering at Penn State University. She holds a Ph.D. and M.S.I.E from Purdue University and a B.S.I.E. from the University of Massachusetts at Amherst. She has served as track coordinator, Proceedings Editor, and Program Chair for WSC. Her research interests include manufacturing systems control and CAD/CAM. She is a member of IIE and SME. Her email address is <djm3@psu.edu>.

TERRY P. HARRISON is the Earl P. Strong Executive Education Professor of Business and Professor of Supply Chain and Information Systems at Penn State University. He holds a Ph.D. and M.S. degree in Management Science from the University of Tennessee and a B.S in Forest Science from Penn State. He was formerly the Editor-in-Chief of *Interfaces* and is currently Vice President of Publications for INFORMS. His mail address is <tharrison@psu.edu>.