Proceedings of the 1999 Winter Simulation Conference
P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds.

SENSITIVITY OF OUTPUT PERFORMANCE MEASURES TO INPUT DISTRIBUTION SHAPE IN MODELING QUEUES - 3: REAL DATA SCENARIO

Donald Gross

Department of Systems Engineering and Operations Research
George Mason University
Fairfax, VA 22030, U.S.A

ABSTRACT

This paper follows-on papers presented at the two previous WSC conferences on sensitivity of output measures to input distribution selection in queueing modeling. Here, a real situation is studied, where data on input distributions are utilized and distributions selected by two fitting packages, Arena Input Analyzer and ExpertFit. Empirical distributions made from histograms of the raw data itself , as well as the first two choices from Arena and ExpertFit are compared for this small bank queueing network model, showing that an output measure such as mean wait in queue is quite sensitive to input distribution choice.

1 INTRODUCTION

This is the third paper on sensitivity of output performance measures (e.g., mean wait in queue) to the particular shapes of input distributions (interarrival times and service times) in queueing modeling. Queueing theory shows, under certain conditions (M/G/1 and heavy-traffic approximations), that only the first two moments of interarrival and service times effect the mean queue wait. The question under study in the first two papers was, "how much sensitivity to higher moments is there when these particular conditions are not met. In the first paper, Gross and Juttijudata, 1997, a G/G/1 queue was simulated for traffic intensities (ρ) of .5 to .95, and different families of distributions with the same first two moments (i.e., matching coefficients of variation [CV]) were studied. Mean queue waits, Wq, and the 95th percentiles of waiting time in queue, Wq(.95), were compared. Even for cases of relatively high ρ, sizable differences were noted both in Wq and Wq(.95). In the second paper, Gross and Masi, 1998, a small queuing network (a call center) was studied to see if the sensitivity diminished in a network. The study showed it did not; the sensitivities in both Wq and Wq(.95) were about the same as in the single node, G/G/1 case. Also, both studies showed that the tail measure, Wq(.95), was no more sensitive than the mean measure, Wq, and that in some cases, percentage differences were quite high (e.g., for a ρ of .8 and a CV of 2 for interarrival times and a CV of 0.5 for service times, differences between Gamma and lognormal and Pearson type 5 distributions were almost 30% and 70% respectively).

This study centers on a real situation and compares Wq for various empirical and fitted distributions using real data. Here, in addition to differences in higher moments, mean and variances are not identical either, and we compare empirical fits to those from two fitting packages: Arena Input Analyzer and Expertfit.

2 THE MODEL

A student team for a final project in a Discrete Event Simulation course taught for George Washington University at the Aberdeen Proving Grounds investigated an on-post bank and collected interarrival data and service-time data over the noon rush hour. One hundred seventy eight observations of interarrival times, one hundred twenty four regular teller service times and fifty five express teller service times were collected.

This model, built in Arena, consisted of four regular tellers and one express teller for deposits only. If an express customer came in and found the express teller busy, and if one of the regular tellers were available, the express customer would go to the regular teller for service; otherwise the express customer would join the express queue. Mean waits in the regular and express queues were observed for 25 replications of 50,000 customers (simulating a steady-state during the busy lunch hour period). Traffic intensity was approximately .85 for the regular queue and .75 for the express queue, so that even though the ρ's were relatively high, heavy-traffic conditions were not really met.

3 THE INPUT DISTRIBUTIONS USED

Both empirical distributions made from the actual data and theoretical distributions fitted to the data by the Arena Input

Analyzer and ExpertFit were used in the simulation studies. Mean queue waits (WqReg and WqExpr) were outputted for a variety of distribution combinations comparing empirical, Arena fits and ExpertFit fits to observe how large the percent differences in these output performance measures might be. Confidence intervals were obtained using Arena's Output Analyzer, and the confidence bounds were almost always within 2 to 3 % of the mean values, and never more than 5%.

3.1 Empirical Distributions

Since Arena has a limit of 127 characters on any input string, empirical distributions were based on histograms of the actual data and those that best matched the first four data moments (mean, variance, skewness and kurtosis) were utilized. Several histogram approximations to the data were developed and the first four moments resulting from the histograms were compared to those of the actual data. Figure 1 shows the absolute percent differences of the histogram moments from the data.

It appears fairly clear that for Regular Service, E2 is the closest match to the data and for Express Service, E3 appears best. But for interarrival times, it is not clear. E4 matches the mean exactly, while E1 matches the variance exactly. One might be tempted to select E2 as the best overall compromise, but one might think that the greatest sensitivity

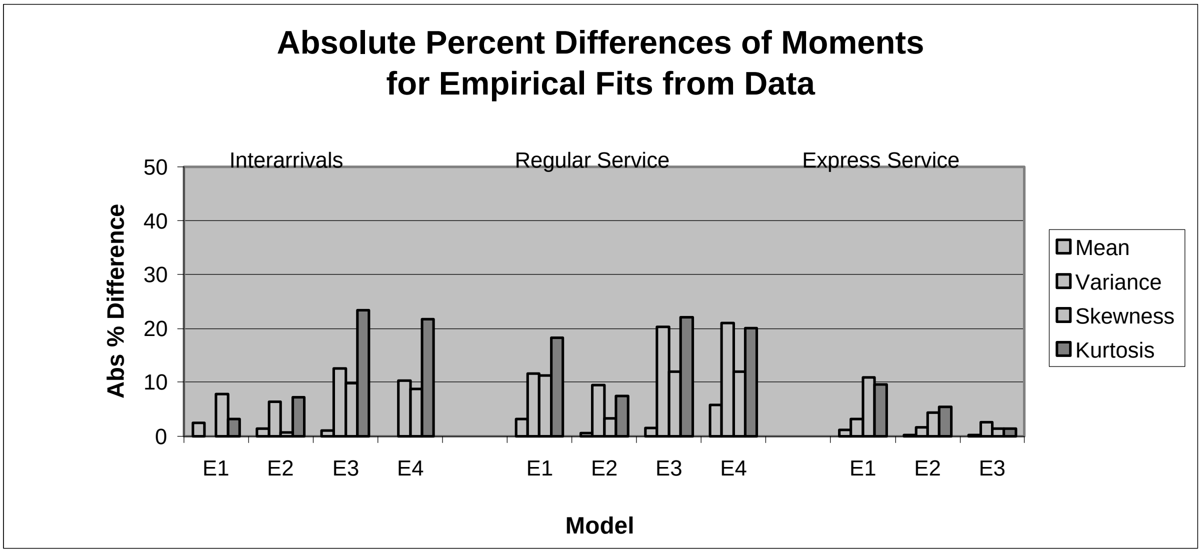


Figure 1: Moment Comparisons of Empirical Fits to Data

of output performance measures such as Wq is to the mean (first moment) and sensitivity lessons as the order of the moments increase. To illustrate that this might be so, we can look at some calculations for $E_k/M/1$ using the QTS software associated with Gross and Harris, 1998. Focusing on the express server only, the data give a mean service time of 1.471, which we use for the exponential service mean. The empirical E4 interarrival time histogram hit the data mean of .486 exactly, while E2 had a mean of .493, about a 1.5 % increase. Since 25% of the arrivals go to the express server, the mean arrival rates become .25/.486 = .514 and .25/.493 = .507 respectively, yielding a traffic intensity of about (0.51)(1.471) $\doteq$ .75. Using the $E_2/M/1$ model, the QTS software gives a Wq of 3.182 for the first case and 3 for the second, an almost 6% difference. Increasing the mean interarrival time by 5% (from .486 to .5103) yields a Wq of 2.61, an almost 18% difference. Interestingly, keeping the mean interarrival time at .486 and using an $E_3/M/1$ model, which reduces the variance from 1.89 to 1.26, (a 33% change) yields a Wq of 2.744, only a 14% difference. This shows how much more sensitive Wq is to the first moment over the second.

3.2 Arena and ExpertFit Distributions

All 178 interarrival time observations, as well as the 124 regular service time observations and the 55 express service time observations were run through both Arena's Input Analyzer and ExpertFit. Table 1 shows the first three choices and their relative scores. Arena's fits are based essentially on mean squared error between the data histogram and the candidate theoretical distribution (low number best). ExpertFit uses a secret formula and comes out with a score from 0 to 100 (high number best). Arena has essentially

Table 1: Arena and ExpertFit First Three Choices

Fit Package	Interarrival Times		Regular Service Times		Express Service Times	
	Distribtn	Score	Distribtn	Score	Distribtn	Score
Arena-1st	Weibull	0.00146	Lognormal	0.0100	Lognormal	0.0010
Arena-2nd	Beta	0.00183	Gamma	0.0195	Gamma	0.0254
Arena-3rd	Lognormal	0.00523	Weibull	0.0212	Erlang	0.0282
XFit-1st	Gamma(E)	97.06	Log-logistic	100.00	Pearson-5	97.83
XFit-2nd	Gamma	89.71	Pearson-6	95.00	Log-logistic	96.74
XFit-3rd	Weibull(E)	86.76	Lognormal	88.75	Pearson-6	89.13
XFit-Beta*	Beta	97.22				

* When given an upper bound value the same as determined in the Arena 2nd Choice, Beta scored 1st

eleven theoretical distributions it considers for fitting the data. These include unbounded distributions such as gamma and Weibull, as well as bounded distributions such as beta and uniform. Arena chooses the best from these, according to its least mean square error. Parameters are generally estimated from the data by maximum likelihood. ExpertFit has many more candidate distributions to choose among, however, unless the user specifies an upper bound value to the variate, ExpertFit will not consider bounded distributions in its automatic (guided) fitting mode. ExpertFit also uses maximum likelihood in estimation of distribution parameters.

From Table 1, we see that Arena and ExpertFit generally picked different distribution families for their first three choices. None of the first choices match. The (E) after some of the Expertfit distributions indicate that a location parameter was added. Expertfit considers a two-parameter family with a location parameter added as a separate choice from the same family without adding the extra location parameter. Arena will add a location parameter if it gives a better fit, but does not consider the same family without the location parameter as a competitor. Figure 2, shows, for service times, the top choices for Arena and ExpertFit compared to the data histograms.

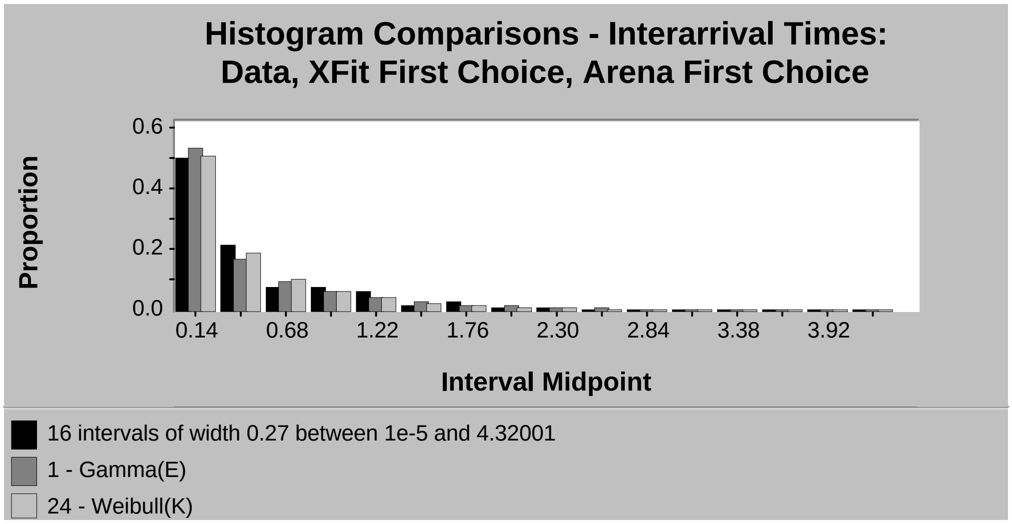
4 RUN RESULTS

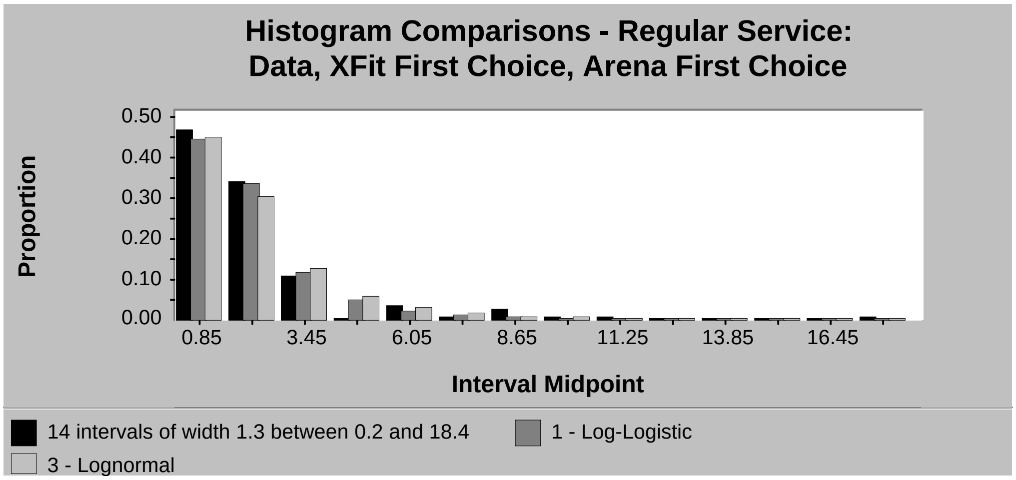
The first set of runs involved only the empirical distributions to see how sensitive Wq was to the particular empirical distribution chosen for interarrivals. Since E2 appeared the best match for regular service and E3 for express service, the models compared were EiE2E3, i=1,2,3,4, where Ei is the interarrival distribution, E2 is the regular service distribution and E3 is the express service distribution. Figure 3, shows the Arena results for the four cases.

Figure 3 shows the Wq for regular service and express service for each of the four empirical models, as well as 95% confidence bounds and the minimum and maximum Wq values over the 25 replications. Note the tightness of the confidence bounds; all are less than ± 5% from the mean. All models, except E4E2E3 were within each others confidence bounds. E4E2E3 shows significant differences (e.g., a percent difference for express service Wq between E2/E2/E3 and E4/E2/E3 of about 8% and a percent difference for regular service Wq of 12%). This is consistent with our earlier analysis using the E_k/M/1 theoretical model from the QTS software, showing a percent difference in express Wq of about 6%, for this slight change in mean interarrival time (.486 vs. .493).

The second set of runs compares Arena and ExpertFit two top choices (models designated as A1A1A1, A2A2A2, X1X1X1 and X2X2X2 respectively) to E4E2E3 which was chosen because of its better match to the data mean. Figure 4 shows percentage differences in Wq and percentage differences in the moments of the Arena and Expert Fit first and second choice models from the empirical model. We set the empirical model as the base from which to compute the percent differences. We do not mean to imply that this is the *correct* model - we will beg the "age-old" question of whether it is better to use an empirical distribution or a fitted distribution. However, one had to be chosen from which to compute percentage differences and the empirical model was chosen rather than to pick either Arena or ExpertFit.

We note from Figure 4 that the percentage differences in the means (we also show percentage differences in the approximate ρ's for regular and express service since differences in mean interarrival times and mean service times could tend to cancel if interarrival means were underestimated and service time means were overestimated) are quite small (< 3 %), but we have seen before that small





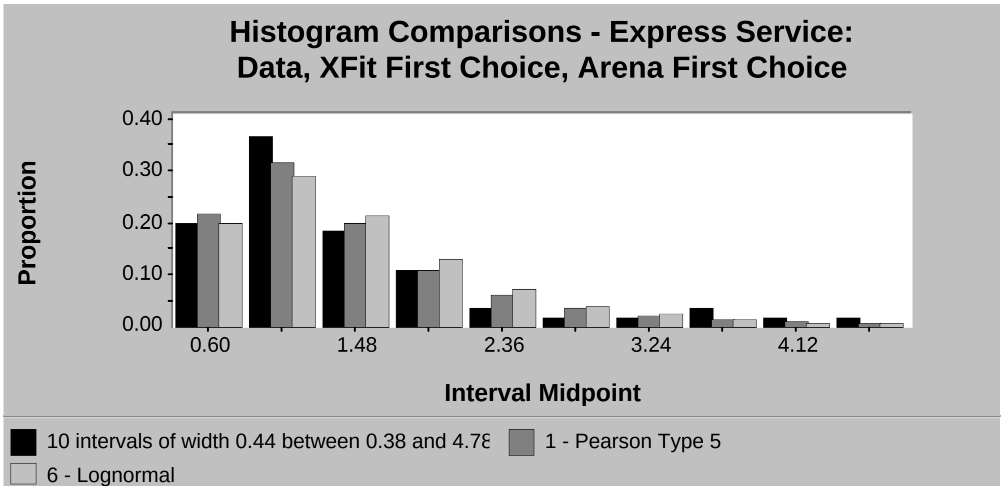


Figure 2: Histogram Comparisons Among Data, ExpertFit First Choice and Arena First Choice

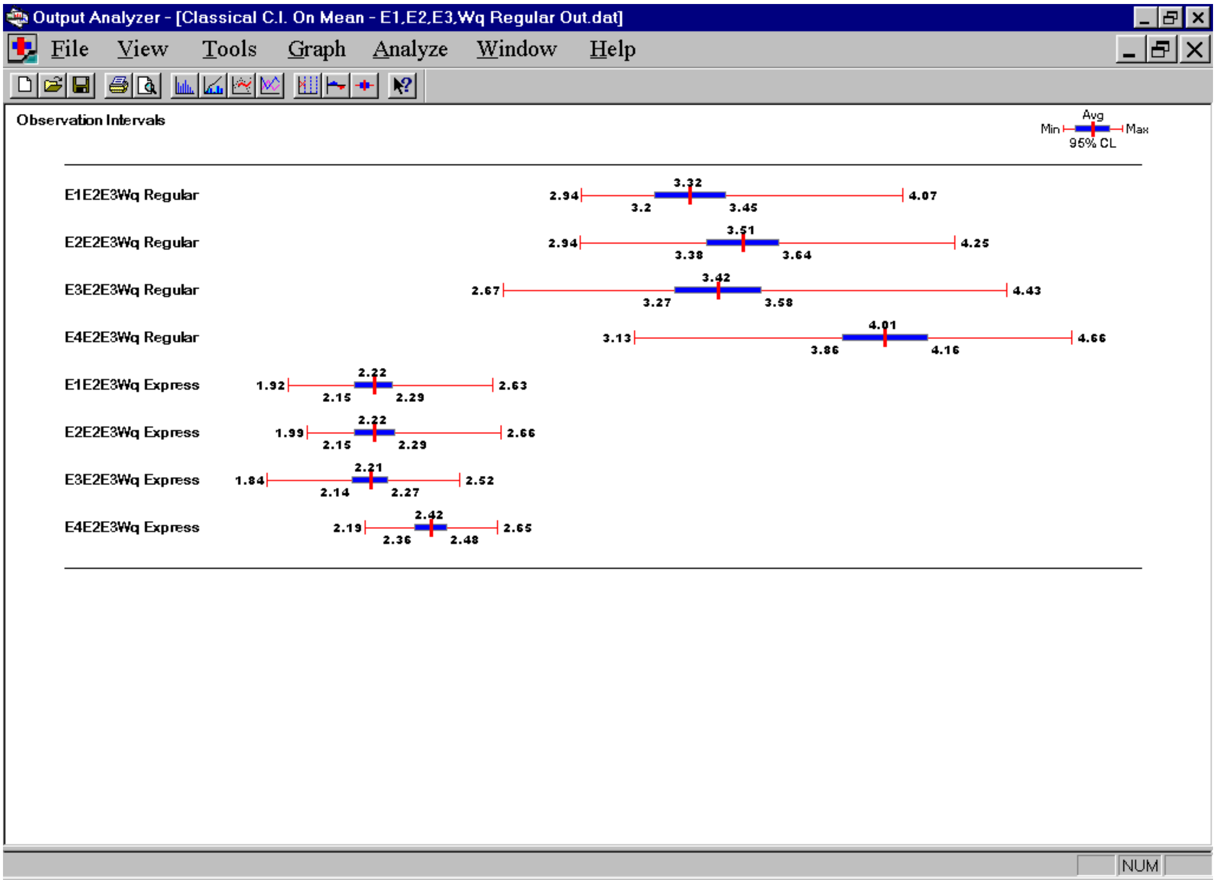


Figure 3: Arena Runs for EiE2E3

differences in means can result in much larger differences in Wq. Because of so many things to compare (two Wq values, four means, four variances, four skewness measures and four kurtosis measures), it is not easy to see what really has the major influence on the Wq differences. The smallest percent difference for WqExpr from E4E2E3 is A1A1A1 (not even shown on the chart since it is so close to zero), but A1A1A1 has the third largest difference in its ρ, indicating that closeness in some of the other moments have an effect. The previous studies referenced in the introduction, where distributions with the same first two moments were compared, already showed that other factors contributed to Wq. We see this again here in that the smallest percent differences in Wq are not necessarily associated with the smallest differences in the first two moments and vice versa. Further, A1A1A1 has the smallest percent difference in <u>both</u> Wq-Reg and Wq-Expr, but had the third largest difference in <u>both</u> ρ values and the second largest difference in interarrival-time variance, although it had the smallest differences in both regular and express service variance and skewness. A2A2A2 had the largest percent difference in Wq-Reg, but was second closest in difference for ρ-Reg.

5 CONCLUSIONS

We can definitely conclude that output performance measures such as mean wait in queue are quite sensitive to the particular input distribution family chosen. Even matching the first two moments was shown not to be sufficient (directly in the previous studies and indirectly here) and that higher moments as reflected by the particular distribution shape, interact in complex ways and significantly influence the output measures.

This suggests that great attention be given to input modeling, for "garbage in gives garbage out" was never truer than in simulation modeling. Sizable effort should be devoted to obtain very large, accurate samples (much larger than those obtained for this study - ideally at least 500 observations) of input distributions, several fitting packages utilized, and significant sensitivity analyses done with the input distributions chosen from the fitting packages, as well as using the empirical distributions themselves.

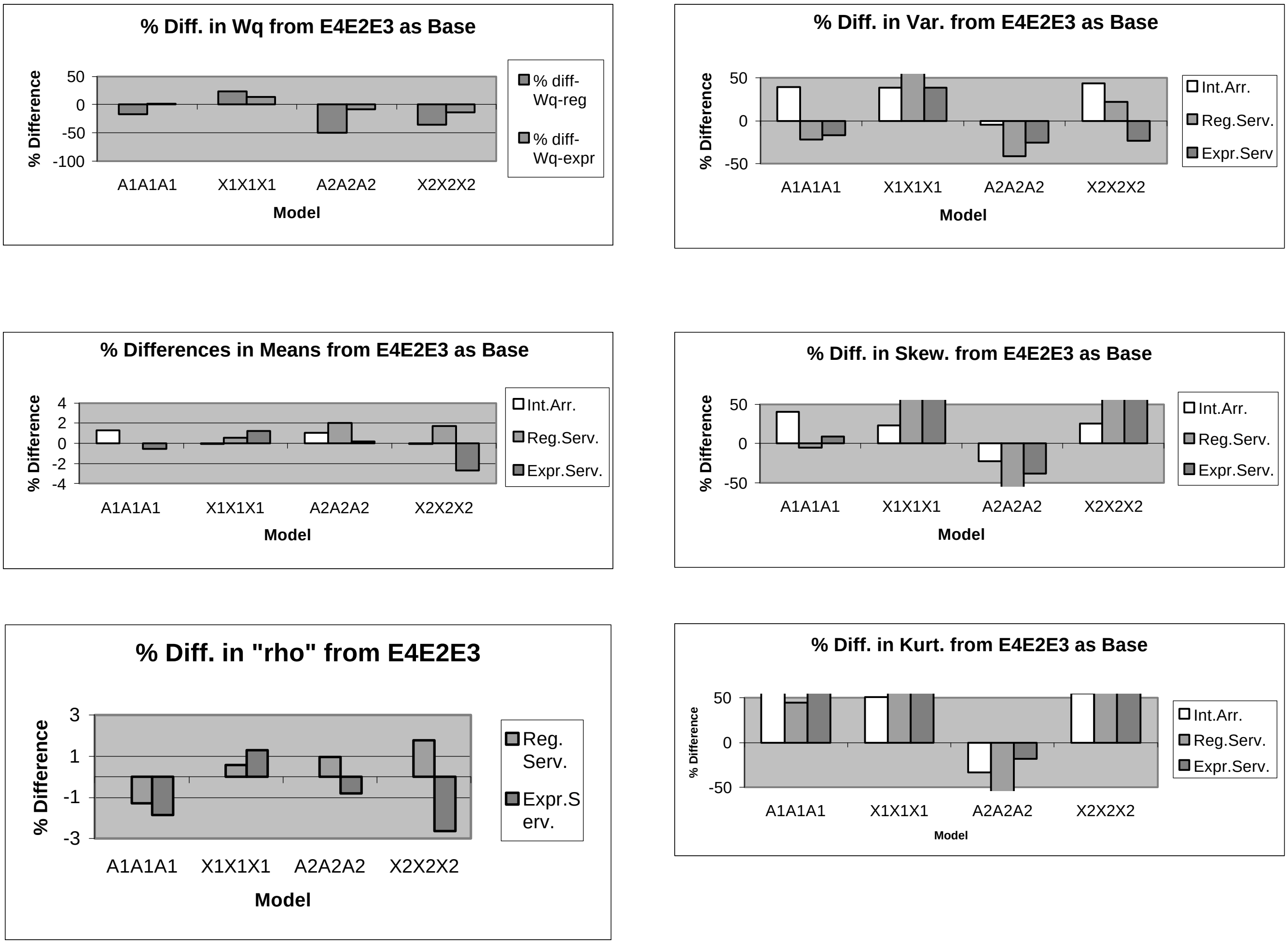


Figure 4: Comparisons Among Arena, ExpertFit and Empirical Models

REFERENCES

Gross, D. and Harris, C.M. 1998. *Fundamentals of Queueing Theory*, 3rd ed., John Wiley and Sons, New York.

Gross, D. and Juttijudata, M. Sensitivity of Output Measures to Input Distributions in Queueing Simulation Modeling. In *Proceedings of the 1997 Winter Simulation Conference*, ed. S. Andradottir, K. J. Healy, D. H. Withers, B. L. Nelson, 296-302.

Gross, D. and Masi, D.M.B., Sensitivity of Output Performance Measures to Input Distributions in Queueing Network Modeling. In *Proceedings of the 1998 Winter Simulation Conference*, ed. D. J. Madeiros, E. F. Watson, J. S. Carson , M.S. Manivannan, 629-635.

AUTHOR BIOGRAPHY

DONALD GROSS is a research professor in the Department of Systems Engineering and Operations Research at George Mason University and Professor Emeritus, Department of Operations Research, at The George Washington University. His research interests are queueing theory and simulation input modeling.