18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

Effects of decay uncertainty in the prediction of life-cycle costing for large scale military capability projects

Axel Bender, **Adrian H. Pincombe, Gregory D. Sherman**

Land Operations Division, Defence Science and Technology Organisation, Edinburgh, Australia
Email: Axel.Bender@dsto.defence.gov.au; Gregory.sherman@dsto.defence.gov.au

Abstract: The prediction of life-cycle costing for large-scale future military equipment capabilities is a difficult problem with many complexities. Nevertheless, as part of the capability development decision-making process, strategic planners are often interested in evaluating high-level risks and measures of performance, such as operational availability or the future equipment's life-of-type. This makes necessary the application of modelling and simulation to provide information that links through-life support cost (or some surrogates) to a capability's high-level performance. Such models typically have common characteristics, including decay or degradation, queuing delays, availability of server resources and maintenance processes.

In this paper we explore the effect of uncertainties on high-level performance over a major military equipment's life-cycle. We are particularly interested in answering the question of how confident a strategic planner can be in the estimates of a future capability's life-of-type given uncertainties in the equipment's failure rates and fluctuations in the maintenance throughput rates. For this exploration we analyse the dynamics of a generic maintenance model that captures the basic features of the life-cycle of a degrading major military equipment system. This generic model is based on queue-server discrete event simulations (Mathworks™ SimEvents/Simulink/MATLAB) which emulate macroscopic maintenance processes with time based parameters and statistical distributions as inputs.

When running the simulations for randomised decay distributions with fixed means and variances in the random number generation, we observe variations of various orders of magnitude in the life-of-type estimates. This suggests that uncertainties in microscopic variables (such as inter-arrival times) cause instabilities in high-level strategic performance indicators, and make the prediction of such indicators and consequently through-life support costs hard, if not impossible. Surprisingly, this magnification of low-level uncertainty does not seem to be common knowledge as the prevalence of mean-based estimation methods for inventory provisioning suggests. We further investigate the potential causes of the found instabilities and discuss the decay process' properties, including possible branching behaviour.

Keywords: *Decay processes, uncertainty, capability, life-cycle cost, through-life support, availability, generic queuing models, random-variates, bifurcation, discrete-event simulation.*

1. INTRODUCTION

When developing military major-equipment capability, strategic planners and decision-makers face many complexities and challenges. A particularly difficult task is the estimation of through-life support (TLS) requirements such as the sizes of repair pools and attrition stocks, and the provisioning of spares. TLS prediction models and simulations support capability decision-makers and can potentially help save tens of millions of dollars. Accuracy in TLS prediction reduces, for instance, the risk of over estimating attrition stocks which results in unused and wasted inventory, or under estimating spare provisioning and making necessary the reordering of parts late in the life-cycle when prices are unfavourably high.

During design and development of a military capability, TLS estimation aims at describing operational availability as a function of capability degradation and throughput capacity of associated support systems such as maintenance, inventory and issuing systems. This design and development stage is usually void of comprehensive and accurate data on system, subsystem and component decays; availability of equipments; and, sources of delays in primary and subsidiary support processes. The ambiguity of available data, so it seems, supports the custom of using heuristics and mean-based provisioning and throughput models. Often TLS is estimated from average utility, and enhancements to its prediction primarily focus on improving the accuracy of mean-value data inputs (average inter-arrival times, failure rates, mean times between failures, mean times between repair, average server throughputs, etc) and adding more detail to underlying process models. Such improvement strategies ignore the stochastic nature of decay and degradation processes. Maintenance, warehousing and other supply chain functions suffer inherent uncertainties which neither exhaustive data collection nor refinements to engineering models can reduce.

In this document, we explore the phenomena that can be expected when TLS estimation takes into account real-world uncertainty in equipment decay and queuing processes. In particular, we look for evidence of bifurcation and instability, and quantify the magnification of uncertainty as it propagates through a typical capability support system. While there is queuing theory research which suggests that complexities in logistics might give rise to chaos (Ranjan et al., 2002; Feichtinger et al.,1994), to our knowledge so far no thorough attempt has been made to study complex systems phenomena in models that inform TLS estimates.

To simplify the discussion we focus on the analysis of maintenance processes. Maintenance costs are a key contributor to TLS costs. For instance, at time of acquisition the provisioning for repair and attrition stocks in Australia's wheeled military vehicle capabilities makes up over 15% of the fleets. The generic model we study has all the characteristics of typical engineering maintenance models used to inform TLS estimation. This simplification is not a shortcoming of our study. On the contrary, the choice of a generic model for our study will strongly support the finding that inherent uncertainties lead to complex phenomena that cannot be derived from mean-based calculations. We are confident that these phenomena are omnipresent and will also be found in more elaborate maintenance models that have degradation or decay inputs.

The paper is organised as follows. In Section 2 we provide a generic description of the maintenance processes we study, and discuss aspects of decay failure uncertainty for repair provisioning and life-of-type (LOT) estimation. In Section 3 discrete-event simulations are used to correlate high-level planning measures such as equipment availability and LOT to maintenance system inputs that display inherent uncertainties in form of small random fluctuations around mean values.. In Section 4 we discuss our results and conclude.

2. MAINTENANCE PROCESSES AND MODELLING ENVIRONMENT

In our study we use a generic maintenance process as the show case for the effect that uncertainties can have on LOT prediction and consequently TLS estimates. The uncertainties we focus on are inherent to maintenance processes. Unscheduled maintenance, for instance, is often the result of a decay process in which, by and large, component or system failure occurs randomly. Throughput rates in repair servers fluctuate because no maintenance job precisely mirrors another and because human maintainers do not take exactly the same time to perform work of the same type (e.g. oil changes, replacements of spark plugs etc). Even scheduled maintenance does not follow precise schedules; the 30,000 km service of a car might occur at 29,236 km or 32,888 km or two weeks before or after it is actually due.

The maintenance process we model represents the time based flow of the repair demand (or "failure") of a military vehicles (or "entities") capability, its storage and associated maintenance. For simplicity, we merely monitor the repair pool of vehicles that are over and above a constant operational availability threshold. This repair pool reflects the capability's redundancy. Once it is exhausted the vehicle fleet will no longer meet its operational availability target and will be compromised. We thus define LOT as the point in time when the maintenance system can no longer keep up with the demand and the repair pool is depleted by more than 50% of its initial value. After reaching its LOT, the maintenance process is said to be in the "failed state".

There are two generically different approaches to the modelling of maintenance processes: 1) mean-based inventory provisioning whereby capacity and mean volume flow rates are estimated; and, 2) Discrete-Event Simulation (DES) of end-to-end logistical and fault analysis using queue and server design based upon the flows of entities/vehicles. Here we focus on the second case but provide a link to the first one through the study of LOT. We use the MATLAB™, Simulink™, SimEvent™ software packages because of the ease at which statistical and randomisation practices can be applied within the model environment. The layout of the simulation is given in Figures 2 and 3. The types of maintenance process modelling in which we are interested are those consisting of input, server-queue, and output as shown in Figure 1. The inputs represent repair servicing and three common types of repair demands. To take into account uncertainties, these inputs are described by four *random-variate* (randomised statistical) distributions, one for service times and three for time intervals after which entities are "generated" at the model's demand input nodes (i.e. moved from the distribution or output node to these input nodes). Each generated entity is passed to the queue. If a server is available to process the new demand, the entity is removed from the queue, serviced and after completion of service ejected from the maintenance system. The output measures the net number of entities in the whole maintenance system as a function of time.

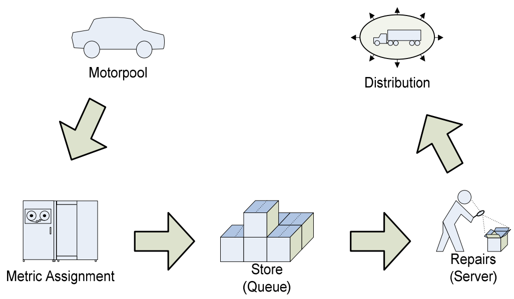

Figure 1 Server-queue flow for maintenance processes

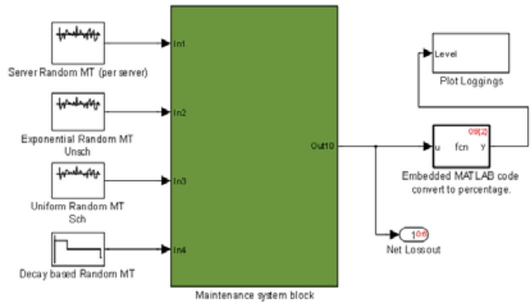

Figure 2 Simulation layout in Simulink™

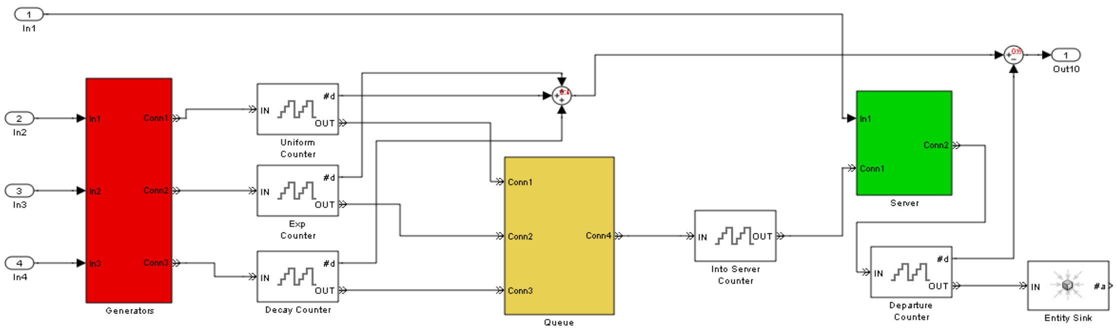

Figure 3 Internal simulation modules of the maintenance block in Simulink™

The key features of our maintenance DES model are:

1. Initially server capacity is tested between 100 and 110% of demand; this range is then extended to between 90% and 135% of demand. This reflects the effectiveness-efficiency trade-off that military decision-makers need to consider when designing the TLS system of a new capability. Over supply of maintenance servers (i.e. high server capacity) corresponds to an inefficient system in terms of idle workshop equipment and personnel. Under supply results in an ineffective system.

2. Two different mechanisms for unscheduled failures are considered. The first one describes failures caused by material fatigue and ageing, and results in unscheduled maintenance requirements that increase slowly over time. Its functional form reflects experience with military (vehicle) capability systems, in which serviceability degrades gradually over the LOT. The second one generates a "background" of accidental, memoryless failures as caused, for instance, by vehicle crashes. Scheduled maintenance requirements are overlaid with these unscheduled maintenance requirements as in (Upadhya and Srinivasan, 2003).

3. *Random-variates* are used to describe the inherent uncertainties in failure rates and servicing times. They cause random, small time-scale fluctuations in maintenance demand and server throughput. These fluctuations are very small in comparison to the total number of entities in the model; i.e. they only perturb slightly a corresponding mean-based inventory provisioning system.

4. There exist no hidden infinite serving times. Also small levels of quantisation are allowable as most maintenance queue systems behave in a quantised manner.

The three statistical demand distributions relate to creeping decay failure, unscheduled memoryless and scheduled maintenance requirements. They describe the inter-arrival times between entities in need of maintenance. The first input is a creeping *decay* with an instantaneous decay time, $A + Be^{\lambda t}$, where t denotes time and $\lambda < 0$ is the decay constant. For small t, inter-arrival rates are close to $1/(A+B)$; for large t the rate approximates $1/A$. This decay time is derived from a failure rate that satisfies the Velhurst equation giving a form of the Logistic curve (Kreysig, 1999) and is often used in models for rates of systems failure in the process of aging (Gavrilov and Gavrilova, 2001).The arrival rate for creeping decay is multiplied by a random-variate Gaussian distribution with unity mean and a variable standard deviation which is determined in testing. This Gaussian distribution gives rise to small-scale fluctuations that are present in real-world repair demands caused by equipment degradation. The other two inputs describe *regular* and *irregular* maintenance arrival times. We use a randomised uniform distribution for inter-arrival times that relate to regular repair demands (Johnson, 2006) and a randomised exponential distribution (the distribution of times between events in a Poisson process) for those relating to irregular memoryless failure (Upadhya and Srinivasan, 2003).

The repair server system is an N-server. Repair processing times for each individual entity are sampled from a randomised Gaussian distribution (the fourth input). When repairing the selected entity a whole server within the N-server is busy for the whole randomly chosen time period.

In this paper we present three experiments which consist of multiple simulation runs. In each run there are 300,000 simulation time steps. The following aspects of the random-variate input distributions of inter-arrival times are fixed throughout this investigation:

1. Decay distribution: $\lambda = -0.000005$ $A = 16$, $B = 4$ and the normalising randomised Gaussian distribution has mean = 1. Figure 4 shows an example cumulative distribution function.
2. Regular maintenance distribution: mean = 70, uniform in the [60,80] simulation time step interval.
3. Irregular maintenance distribution: mean = 70 simulation time steps.
4. Service time distribution per server: mean = 360 simulation time steps.

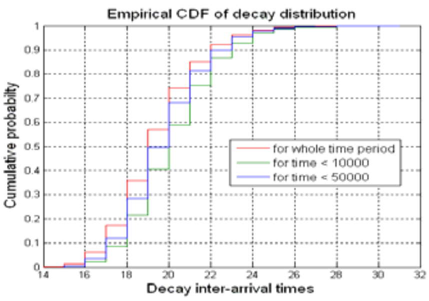

Figure 4 Cumulative function for a sample decay distribution for a variety of inter-arrival times generated under the first experiment conditions with standard deviation in decay random-variate at 10%

The normalising Gaussian standard deviation, the number of servers and the standard deviation in the server random-variate are fixed in each simulation run. The first experiment has 32 servers and a standard deviation in service time random variate of 11.1%. Over multiple simulations the Gaussian standard deviation in the creeping decay-variate is changed, in order to study the effect caused by growing uncertainties of repair arrival times arising from equipment degradation. In the second experiment we vary deviations in the service time random variate. The final experiment then investigates varying server numbers. These input parameters were chosen to meet the three criteria discussed above. From simple steady state calculations, 1/(16+4)+1/70+1/70 = 11/140 = (99/112)*(32/360), i.e. the mean arrival rate at time = 0 is 99/112 = 88.4% of the average capacity of the N-server. As time increases to infinity the mean arrival rate approaches 459/448 = 102.5% of the N-server's average capacity i.e. 1/16+1/70+1/70 = (459/448)*(32/360). This asymptotic limit, however, is not reached in the experiments. After 300,000 simulation steps, the mean arrival rate is around 0.878, i.e. 98.7% of the average server capacity. Under these conditions one would expect from a mean-based calculation of flows that the performance outcomes in the experiments would only show small availability losses over time and that the system would rarely if ever enter the failed state.

In all simulations, we start with 600 serviceable vehicles in the repair pool (i.e. there is a total number of 600 entities in the system). Throughout any simulation, we monitor the proportion of available stock which we call the *availability level* given by the equation:

$$100 \times (\text{total number of entities - number entities in maintenance}) \Big/ \text{total number of entities } \%. \qquad (1)$$

The other performance measures relate to time (measured in simulation steps) when instability sets in, time of particular loss thresholds in availability levels, and rates of such losses. From these measures we extract: when and under what conditions significant events occur, the time of any evidence for instability, the nature of any effects such as branching and, if such events occur, the rates of change in availability levels. We also seek to obtain the net qualitative effect of any uncertainty and the propagation of randomisation effects.

3. SIMULATION RESULTS AND ANALYSIS

In the first experiment we studied the impact of increasing uncertainty in inter-arrival times of failures caused by creeping decay. We ran the simulation for 32 different widths of the Gaussian distribution (in steps of 1% standard deviation from 0 to 31%) and, for each variance, ran the simulation 50 times. Different runs reflect different samples from the random-variate input distributions. In each simulation the vehicle availability performance measure showed marginal fluctuations in fleet availability caused by the random (emulated day-to-day) input variability. However, after around 80,000 to 120,000 time steps, large discrepancies in availability levels between the simulation runs of exactly the same model emerge. An example of two typical simulation instantiations is shown in Figure 5.

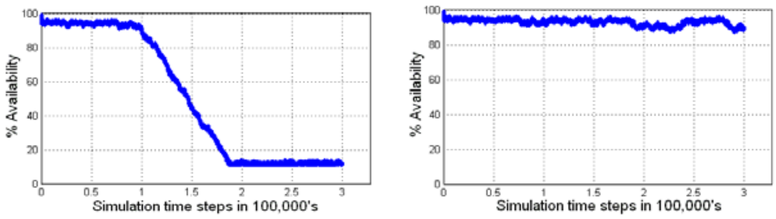

Figure 5 Typical fleet availability output from two experiments of the same model (fixed mean and standard deviation) under the same conditions and only differing in the instantiation of the random variates

In collating results from each simulation run, we sampled output (availability) data only every 50,000 time steps and for those times when it falls below 80%, 50% and 20%. These sample values provide sufficient information to plot an approximate availability prediction curve without the smaller disturbance effects. Figure 6 shows that instability sets in somewhere around the 100,000 time step mark. The possible availability prediction curves are clustered into branching sets, or "*cluster bands*", some of which ("failing runs") correspond to large losses over the simulation time and some of which ("successful runs") do not reach a LOT. These clusters are not characterised by the same decay standard deviation but contain simulations of models with varying standard deviations.

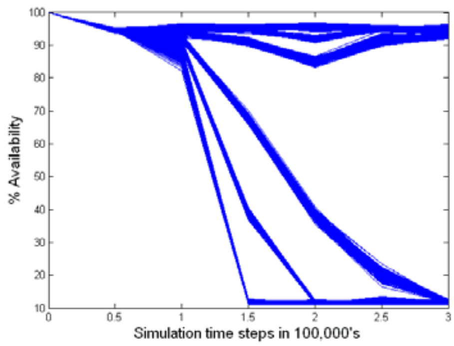

Figure 6 Availability prediction for a model with varying decay standard deviation

The scatter plot in Figure 7 displays the change in availability levels over the second 100,000 simulation time steps, and clearly shows *bifurcation* effects in availability losses.

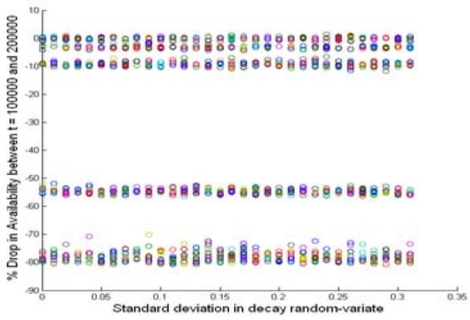

Figure 7 Drop in availability between simulation time steps 100,000 and 200,000. Same data as Figure 6

Of those simulation runs that lead to failure, Figure 8 shows two scatter plots with results derived from measurements of availability levels of 20%, 50% and 80%. The first scatter plot (left) shows at least five cluster bands of LOT as a function of decay standard deviation. The second plot shows slope values for the availability drop from 80% to 20%. There obviously exist close to consistent differences in slope values between successive cluster bands. Figures 7 and 8 suggest that the standard deviation for generating the decay failure distribution has little or no impact on outcome likelihoods. While the observed instabilities are undoubtedly caused by uncertainty (randomisation in the model distribution functions), the magnitude of uncertainty seems to have no noticeable effect.

Our second experiment looks at a select subset (1%, 6%, 11%, 16%, 21%, 26% and 31% for the Gaussian standard deviation in the creeping decay) of the first experiment's simulations under changes to the variance in the service time random-variate.

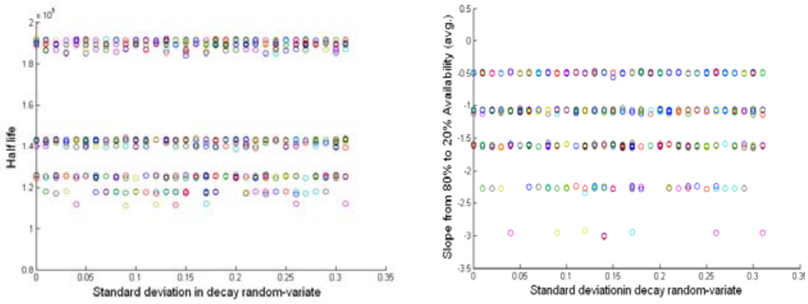

Figure 9 shows a subtle increase in fluctuation of availability loss as the variance for the service time random-variate increases. This increase is minimal compared to the branching effect. Fluctuations in service rates have virtually no effect on availability levels within each cluster band.

Figure 8 Failure cases: LOT, or half-life (left) and slopes of descent from 80% to 20% availability (right) vs. standard deviation in arrival rate of random-variate in decay

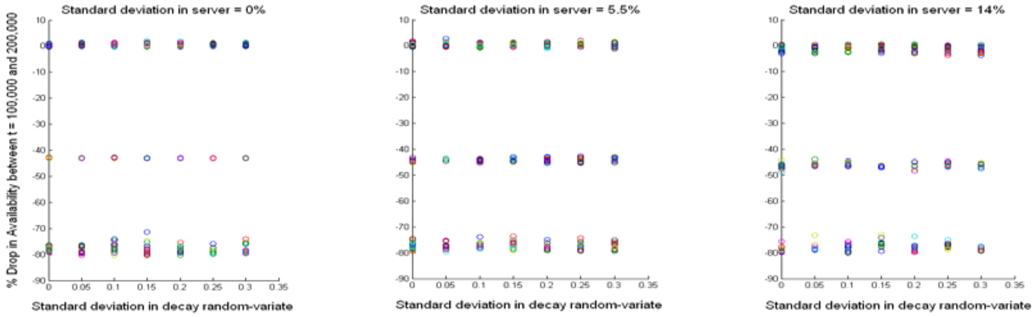

Figure 9 Same as Figure 7 but for varying servicing time variance. Fluctuations in servicing time increase from the left to the right panel. Note that in Figure 7 the standard deviation in the server is 11.1%

Our final experiment investigates the existence of the branching phenomenon when the server throughput capacity (number of servers) changes. This is done by running a select subset (1%, 11%, 21% and 31% for the Gaussian standard deviation in the creeping decay and 0%, 10%, 20% and 30% for the standard deviation in the service time variate) of the second experiment's simulations and then sweeping through different numbers of servers. As can be seen in Figure 10, the branching phenomenon is strongly affected by the queue-server system capacity. Clearly the number of servers has the most dramatic effect on the position of the onset of instability, the number of cluster bands, and the probability of system success/failure (i.e. the number of members in the various cluster bands). Figure 10 indicates that there are parameter regions that generate almost certain outcomes, such as "regular failure" (leftmost panel) and "regular success" (rightmost panel), the transition region between these two stable regions exhibits complex topological structure, both in terms of numbers and shapes of cluster bands.

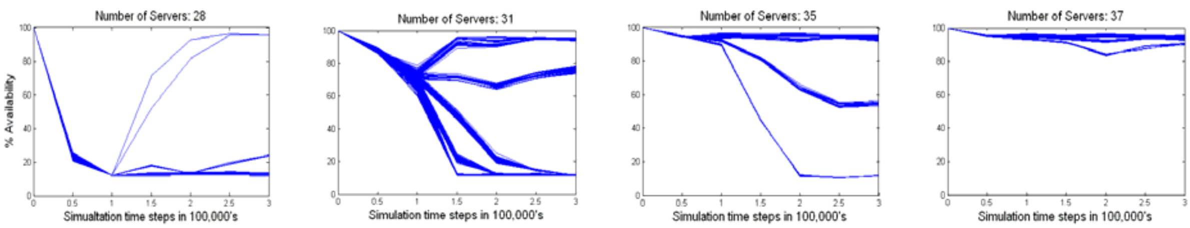

Figure 10 Same as Figure 6. However, simulations are run for a subset of the simulations presented in Figure 9. The number of servers increases from the left to the right panel

Table 1. Comparison of outcomes for provisioning techniques with respect to the fixed inputs.

Server numbers	**<29**	**<32**	**<35**	**<38**	**>38**
Percentage chance of failure (determined by simulation)	>96	>50	>2	<2	0
Likely outcome / Risk factor of failure	Fail	High	Med	Low	Success
Mean arrival rate at $t = 0$ (% of mean service capacity)	>97.5	>88.4	>80.8	>74.4	<74.4
Mean arrival rate at $t = 300,000$ (% of mean service capacity)	>109	>98.7	>90.3	>83.1	<83.1
Mean arrival rate at $t \to \infty$ (% of mean service capacity)	>113	>102	>93.7	>86.3	<86.3

The relation between system success probability and server throughput capacity can be compared with simple estimates (Table 1). From a mean-based throughput calculation one would expect that maintenance systems with more than 32 servers will be able to deal (all the time) with the estimated repair demand. However, inherent uncertainty introduces large risks. For instance, a maintenance system designed to meet the initial repair demand at 88% of capacity and the demand of the degraded fleet with around 99% of capacity, might in reality have a high risk of failing to meet the demand owing to small variances in arrival times.

4. DISCUSSION AND CONCLUSIONS

In this paper we presented discrete event simulations (DES) of maintenance models derived from generic server-queue models. Our main focus was to study the effect of input uncertainties on the predictability of life-of-type (LOT) and operational availability. LOT prediction is an important component of through-life support (TLS) cost estimation. Improving its accuracy in large-scale military capability planning can potentially save millions of dollars.

Unfortunately, our study demonstrates that there are fundamental limitations to LOT prediction accuracy. These limitations arise from the stochastic nature of decay and degradation processes, i.e. from the fact that times of entity arrival in repair workshops or servicing times are not deterministic but distributions governed by random processes. Even very small fluctuations around mean times between failures, mean times between repair, etc lead to unexpected complex phenomena. Small uncertainties are amplified by orders of magnitude, and there is an observable change in the topological structure of output data from which high-level measures of performance, such as LOT and operational availability, are derived. More precisely, in our investigation we discovered fluctuation-induced branching phenomena in temporal operational availability profiles of generic maintenance server-queue models. The bifurcation properties showed structure in the onset of instability and the slopes of system degradation towards failure.

To our knowledge these complex maintenance dynamics have so far not been discovered in TLS estimation models. However, they might explain why estimators such as OmegaPS Analyzer 4.0 show an extraordinary sensitivity to small variations in input parameters. Of course, instability and branching phenomena add a degree of ambiguity to the definitions of system “success” and “failure”. What comprises failure or success of the maintenance system depends on how certain a planner or decision-maker needs to be with respect to the availability of entities. Risk evaluation needs to change in the presence of bifurcation in prediction curves, and needs to differ from that derived in mean-value based estimates. Stated differently, the *robustness* requirement for the maintenance system should be an important consideration in TLS estimation processes. Effectiveness-efficiency trade-off analysis should make room for effectiveness-efficiency-*robustness* trade-off considerations when estimating TLS costs.

The observed complex phenomena suggest that a fine-tuned maintenance system’s predominant reaction to inherent uncertainties is that of a complex system in an unstable dynamic regime. Uncertainty does not seem to propagate in a continuous manner but undergoes discontinuous transitions that express themselves in the emergence of topological change. The existence of bifurcation is a likely result of delay effects in queue-server dynamics such as described in Iooss and Joseph, 1980. Whilst known that this is sometimes linked to non-linear dynamics (D. Driebe and R.R McDaniel, 2005), such behaviour is by and large ignored in provisioning planning and subsequent TLS estimates. As illustrated here, this ignorance might result in failure to notice significant risks and thus might prove to be very costly.

REFERENCES

Driebe D. J and McDaniel R. R., (2005), Uncertainty and Surprise in Complex Systems: Questions on Working with the Unexpected, *Springer: Complexity*, 3, 24-27

Feichtinger G., Hommes C. H and Herold W., (1994), Chaos in a Simple Deterministic Queueing System, *ZOR – Mathematical Methods of Operations Research*, 40: 109-119

Gavrilov L.A., Gavrilova, N.S, (2001), The reliability theory of aging and longevity. *Journal of Theoretical Biology*, 213(4): 527-545.

Iooss G., Joseph D. D., (1980), Elementary Stability and Bifurcation Theory, *Springer-Verlag*, 4-11

Kreysig E., (1999), Advanced Engineering Mathematics, *John Wiley and Sons Inc*, 8th Edition, 31

Johnson R. W., (1994), Estimating the Size of a Population, *Teaching Statistics*, 16:2, 50-52

Ranjan P., Kumara S, Suruna A., Manikonda V., Greaves M., Peng W. (2002), Decision Making in Logistics: A Chaos Theory Based Analysis, *AAAI Technical Report*, SS-02-03.

Upadhya S. K. and Srinivasan N. K., (2003), Availability of Weapon Systems with Logistic Delays: A Simulation Approach, *International Journal of Quality & Reliability Management* 20:7, 836-846.