Proceedings of the Operational Research Society Simulation Workshop 2012 (SW12)
B.Tjahjono, C. Heavey, S. Onggo, and D-J. van der Zee, eds.

EXPLORATIVE RESEARCH INTO CURRENT PRACTICE OF EXPERIMENTATION IN DISCRETE EVENT SIMULATION

Kathryn Hood

Warwick Business School
University of Warwick
Kathryn.Hoad@wbs.ac.uk

Thomas Monks

Peninsula School of Medicine and Dentistry
University of Exeter
thomas.monks@pcmd.ac.uk

Frances O'Brien

Warwick Business School
University of Warwick
Frances.O-Brien@wbs.ac.uk

ABSTRACT

Experimentation is arguably one the largest and most active research areas within discrete-event simulation. However, studies of discrete-event simulation practice report little transfer of this theory into real world application. This paper explores this gap and presents early results from an explorative investigation into current experimentation practice. Results are similar to previous findings: search experimentation was not in regular use, statistical analysis of results was limited and experimentation is still constrained by time pressures and client demands. One surprise was that optimisation was not found to have made an impact with our participants. This disagrees with previous studies that predict improvements in simulation software will improve uptake of theory by practitioners.

Keywords: Scenario comparison; Search Experimentation; Optimisation; Practice

1 INTRODUCTION

Discrete-event simulation (DES) studies are often split into four phases: problem understanding, conceptual modelling, model building and experimentation (Robinson, 2004). Within these four groupings experimentation has arguably one of the most mature literature bases with active research making theoretical contributions in areas such as meta-modelling (e.g. Friedman and Pressman, 1988) and optimisation (e.g. Fu, 2002). The value of this research, of course, is only realised once it has been transferred into practice. To some degree this transfer of theory to practice will rely on commercial simulation and statistical software implementing techniques in a user friendly manner. However, the use of these techniques in practice will also be affected by simulation modellers' knowledge of them (Hoad and Monks, 2011) and client needs (Fildes and Raynard, 1998, Hollocks, 2001)

The purpose of this paper is to present a study into DES experimentation practice and explore the extent of the gap from theory. To avoid confusion with wider areas of 'output analysis' we note that the scope of the term experimentation used here is restricted to the method of scenario comparison and method of searching the solution space.

The paper starts with a brief review of 'formal' experimental design techniques used in DES as well as the little that is written on informal methods such as Visual Interactive Experimentation (VIE). The previous research into the practice of experimentation is then introduced. In particular, the work of Hollocks (2001) illustrates that a substantial gap between the theory and practice of experimentation was present some ten years ago. A recent explorative case study, consisting of the interviewing of eight DES practitioners, investigating the current practice of simulation is then described. The paper concludes by questioning if the gap between theory and practice has reduced in the last ten years. Further efforts to investigate this area are discussed.

2 EXPERIMENTATION: THEORY AND PRACTICE

The literature base for experimentation is wide and deep and cannot be covered in detail here. However, before proceeding to a review of previous studies of experimentation practice, it is useful to briefly discuss Table 1 that provides two broad groupings of formal experimentation theory, comparing scenarios and search experimentation, along with the informal methods that can be applied within them.

2.1 Scenario Comparison

A scenario refers to an alternative configuration of a system; for example, the throughput of a call centre may be investigated with a different number of call operators or customer handling procedure. Formal methods for comparing performance between scenarios all include some statistical foundation; for example, comparison of scenarios with a base case or in a pairwise manner may be conducted by paired-t confidence intervals for mean differences. More advanced approaches include Ranking and Selection (R&S) such as indifference zone techniques to narrow down scenarios to a 'best' or 'group of best' scenarios (Law, 2007). Away from formal methods of comparison, simulation software provides support for Visual Interactive Experimentation (VIE). That is, the ability for clients to watch and interact with the simulation model in order to see the impact of changes in variables on model performance (Belton and Elder, 1994). Other informal techniques include comparative graphical plots, such as histograms and boxplots, as well as point estimates such as mean outputs and other descriptive statistics.

Table 1: *Formal and informal experimentation techniques from DES literature*

Grouping	Formal Techniques	Informal Techniques
Scenario Comparison	Comparisons with a base case Comparisons with the best All pairwise comparisons Ranking and selection methods	Visual Interactive Experimentation Graphical plots Descriptive Statistics
Search Experimentation (incl. Sensitivity Analysis)	Experimental Design Meta-modelling Optimisation (heuristic search)	Personal Judgement Client specified

2.2 Search Experimentation

Experimentation can also take the form of a search of the solution space. That is, a number of experimental factors are varied according to a chosen method and performance is monitored to quickly locate particularly good (or bad) combinations or/and improve understanding of how factors influence performance. By this definition, search experimentation also incorporates sensitivity analysis; used either as a method to locate robust solutions or as part of validation (Robinson, 2004). Formal methods can either be statistically based, such as efficient choice of scenarios via experimental design and optimisation via meta-modelling (Friedman and Pressman, 1988); or computational, such as a heuristic search optimisation method (Fu, 2002). The alternative informal methods to searching largely involve a level of personal judgement either by the modeller, client or some combination of the two; for example, a client may have two competing designs of a manufacturing plant to compare - hence the search is limited to those scenarios.

2.3 Experimentation in practice

Insight into the practice of DES experimentation has been built up from studies that have surveyed practitioners specifically about experimentation methodology (Hoad and Monks, 2011, Hollocks, 2001), specifically exploring multiple scenario comparison in practice (Hollocks, 2001) and general studies of DES practice that provide some insight into experimentation (Christy and Watson, 1983). Although the list of studies is small, all of these studies indicate a leaning towards informal methods of experimentation in practice.

In the early 1980s Christy and Watson (1983) found that experimentation conducted was substantially influenced by time pressures. In part, this may have been due to the speed of computers at this time and the relative infancy of commercial simulation software packages. However, almost 20 years later Hollocks (2001) reports a similar result: that experimentation methodology relies heavily on user judgement and is constrained by time pressures – a result echoed in more general Operational Research modelling around the same time (Fildes and Raynard, 1998). Search experimentation was commonly limited to client specified options with little evidence of experimental design, meta-modelling or optimisation playing a role. Sensitivity analysis was often employed, but again this was influenced by personal judgement as opposed to an experimental design (Hollocks, 2001).

In a more recent study, Hoad and Monks (2011) survey practitioners' approaches to comparing multiple scenarios. Although sample size was relatively small, the results indicate that some of the gap between theory and practice may be explained by a simple lack of knowledge of the theory itself. In particular, findings show that only 5 out of 26 (20%) respondents had heard of the Bonferroni Correction for multiple comparisons - even though details of it are available in standard DES textbooks (Banks et al., 2011, Law, 2007, Robinson, 2004). There was also little evidence that alternative scenario comparison methods, such as R&S, were in regular use.

Although each of these studies has limitations, either due to sample size or advances in simulation software since publication, together they provide some evidence of a gap between the theory and practice of experimentation. One explanation for this gap may be that some methods explored by Hollocks (2001) were simply too new at the time to have widespread adoption. For example, the role optimisers play in current practice might be more substantial. The findings of Hoad and Monks (2011), however, suggest that simulation education could also be improved to cover more basic experimentation theory and improve the chances of transfer in practice. It is clear that in order to move forward more research is needed into the current state of experimentation practice.

The remainder of this paper presents an explorative case study conducted with eight DES consultants which provided a basis for substantive discussion on industry practice of DES experimentation. The insights gleaned from these interviews are not generalisable to the population of DES practitioners due to the small, non-random sample. However, we believe the insight that this exploratory research gives is valid and informative and succeeds in provoking interesting discussion into this subject.

3 INSIGHTS INTO EXPERIMENTATION PRACTICE THROUGH INTERVIEWS WITH PRACTITIONERS

The results reported in this section are organised into nine sections, summarising the themes explored in Table **2**, reflecting the use of experimentation theory, comparing scenarios and search experimentation, in practice as well as project issues that influence experimentation.

In total, eight consultants were interviewed[1]. Five of the interviewees are external consultants. The remaining three interviewees are internal consultants: one works for a scientific research facility, while the other two work for the same manufacturer. The amount of experience each interviewee has in DES varied quite widely from as many as thirty years to just two.

[1] For anonymity and ease of writing interviewees will be referred to as "he" regardless of actual gender

Table 2: *Themes explored in the interviews*

Section	Themes Explored
Time spent on experimentation phase	Percentage relative to other aspects of a DES modelling study
Method of experimentation	Frequency of scenario comparison versus search experimentation
Comparing scenario results	Use of theory in practice of scenario comparison
Optimisation	Frequency of use, barriers to use.
Design of Experiments	Frequency of use, types of design, barriers to use.
Meta-modelling	Frequency of use, barriers to use.
Sensitivity Analysis	Frequency of use, types of design, barriers to use.
Constraints on the experimentation phase	Software limitations, client pressures
Presentation/reporting of experimentation results	Use of statistical measures of uncertainty (e.g. confidence intervals), Graphical methods etc.

3.1 Time spent on experimentation phase

From the interviews it was clear that there were two main types of simulation project:

1. The 'simple' model building project, where the built and validated model is handed over to the end-user (client) for them to experiment with or use as they wish.
2. A consulting project where the client seeks recommendations from the simulation practitioner. These types of project therefore include an experimentation phase in order to produce the required recommendations.

Both types of project could be found in the consultant environment but understandably the second type was most prevalent in the in-house practitioners. It was therefore usual for the majority of the interviewees to conduct the experimentation phase themselves. The exceptions described the vast majority (90% in one case) of their work as building models and setting them up for experimentation by the client.

The percentage of time spent on experimentation during these projects varied quite widely between the interviewees. One interviewee stated that his primary work is now maintaining and updating existing models therefore almost all of his time was spent on experimentation. The other interviewees gave answers ranging from 10% to 50% of the project cycle. In particular, some interviewees estimated that they spend only around 10% of project time on experimentation, when experimentation is required. It was the in-house manufacturing practitioners who reported around 50% of their time is spent on experimentation.

3.2 Method of experimentation

Unsurprisingly, the rather open-ended question of how experimentation was conducted obtained very diverse responses, however there was a common thread running through all eight responses which was the overwhelming practice of comparing alternatives (scenarios). Scenarios or alternative system configurations were defined either at the beginning of the project or during its course, and could be provided by the client, developed by the practitioner, or generated through consultation between both parties.

For example, one interviewee was quite clear that in his experience there were basically two types of model: one where there are identified a number of experimental factors that can be varied to obtain the client's objective (search experimentation) and one where there are a set number of specific designs (scenarios) that are to be compared to choose which is 'best'. He acknowledged that when a 'best' scenario is chosen from a set list of alternatives there is then potential to try and optimise around that chosen design, but clients were usually unwilling to go down that route.

3.3 Comparing scenario results

The discussions of how scenario results are generally compared showed that informal comparison using means, medians, variances and/or percentiles was the most common approach. For example, one interviewee stated that in order to decide what design was "acceptable" he usually looked at the mean result as well as the 95^{th} percentile, range etc..., calculated from the replications run for each scenario, and then consult with the client for their views. Another interviewee stated that he typically compared all scenarios to the base case so that "everything else referred to is...an improvement or a worsening compared to that." It was pointed out that, assuming the model didn't take long to run, simply running more and more replications until any difference between results became "obvious", termed as "overkilling the experiment" was a valid alternative to using any "official statistics" tests. There was however, some reported use of basic t tests or consideration of the standard error. One interview even mentioned using a non-parametric median test.

3.4 Optimisation

An unexpected finding was that the optimisation add-ons now regularly included in DES software were rarely utilised by our interviewees, although if search experimentation was used optimisation was the method selected by three of the interviewees. A further interviewee also acknowledged that there are situations where the client requires an "exact number" or "absolute optimum" value, for example when modelling a circulating conveyor where "too many" items on a conveyor is as bad as too few.

However, they did not, on the whole, use the optimiser supplied with their respective simulation package. Instead the in-house consultants said they used their own "optimisation tool which allows us to do all kinds of scenarios." Others alluded to using an informal optimisation methodology or "experimentation framework" to home in on the "optimal area", whilst a further interviewee preferred using ad-hoc methods: "run...the extremes,...then using educated guesswork,...narrow your [search area]... and then experiment around those areas". He explained that the main reason he had never chosen to use the provided optimiser was "speed"; "A lot of times, if you're going to use [the optimiser], you have to build your model with that in mind [as] it has to run very quick; otherwise ...you're waiting a long time for your results."

One interviewee went so far as to suggest that simulation and optimisation should be regarded as distinct issues; he implied that simulation is a tool used to perform optimisation rather than the other way around.

3.5 Design of Experiments

Although experimental designs (e.g. full ($2k$) factorial designs) were only ever named by a single interviewee, other interviewees mentioned attempting to identify key inputs; suggesting an informal approach approximating experimental design (Robinson, 2004).

It was conjectured by one interviewee that it was rare that you would need any proper experimental design because it was often possible to simply run a full factorial range of experimentation that covered all that was of interest. Another interviewee stated that creating formal experimental designs was impractical since the questions posed requiring answers often changed from day to day.

3.6 Meta-modelling

Although most of the interviewees had heard of the concept of meta-modelling (after it was explained to some of them), none of them practised it. One interviewee referred to the approach as a "holy grail" due to its ability to save time instead of running the simulation many times. However, he also conjectured that meta-models might not be able to capture the complexity of bigger models. This is perhaps at odds with the perception of another interviewee, who thought meta-models were "massively complicated".

3.7 Sensitivity Analysis

Every one of the interviewees carried out sensitivity analysis, and they did so regularly. Typically, this was done using personal judgement, restricting the analysis to key inputs and within a feasible range of levels. One interviewee explained that using sensitivity analysis helped in "understanding what the important areas are to investigate or to [obtain] more accurate data" from the clients.

There was no mention of formal techniques like experimental design to speed up the process, but one interviewee recalled setting the key inputs to extreme levels (these had been determined during the scoping phase of his project) and monitoring the variation in the results. Another interviewee also used some kind of informal experimental design in that the model would be run with parameter values increased or decreased by a couple of set amounts and if a "significant impact" was observed then more detailed evaluation would take place.

3.8 Constraints on the experimentation phase

Another common thread emerged in the interviewees' perceptions of what constrained their experimentation: limited time. The running times of the simulations were deemed to be a major impediment, in line with the findings of previous surveys on simulation practice (Christy and Watson, 1983, Hollocks, 2001). Two of the interviewees mentioned that larger, more complex models required a long run-time with which computing technology is yet to catch up. Time (which was also referred to in terms of cost) was also deemed to be a problem for clients, constraining the amount and type of experimentation that could be practiced.

One interviewee made reference to the gap between theory and practice, arguing that academic methods were too complicated to practise in the real world: "...people in university ...sometimes don't understand...the constraint in the reality; the constraint in reality is the time."

Two interviewees raised the issue of client understanding; this reflects the main problem encountered by respondents in Christy and Watson (1983). It was explained that some clients would struggle with the use of experimentation on top of understanding simulation. This also led to time pressures where a lack of understanding by the client about the time required to run simulations and experimentation could lead to infeasible deadlines.

Software inadequacy was raised by two interviewees. One found problems with inflexible interfacing between simulation software and external data bases e.g. Microsoft Excel, while the other reflected the opinion of Fu(2002) that optimisation algorithms had not yet successfully translated into software programmes. One of the interviewees also pointed out that he, as a relatively knowledgeable person about optimisation, found the existing software difficult to use, and also therefore found it almost impossible to get a client to successfully use it should they show an interest in doing so.

3.9 Presentation/reporting of experimentation results

Client demands and perceived understanding were found to regularly inform the format for reporting experimentation results. Client understanding was an important driver for all the interviewees. For example, one interviewee mentioned revising his typical format of presenting percentile figures to conform to a certain industry standard. He explained that he tends to use the statistics that the clients use themselves "because that's [what] they'll understand."

All the interviewees reported using a variety of graphical techniques to display and explain results. A variety of charts were mentioned, most were generally well known, (scatter plots, column and bar charts, Gantt charts, radar charts, time series plots, line graphs, pie charts, 100% stacked column charts and S-charts), but some of them seemed unique to the particular organisation (e.g. bow-tie and H-charts). Radar charts were mentioned as being useful for displaying variance. S-charts which are a type of process variation chart were also used. The 100% stacked column chart was reported by one interviewee as being a "fairly powerful" way to show "changes over time" and at diagnosing important occurrences in the model over time. Time series graphs were also used to show "how one or more values changes over time". Line graphs were mentioned as being useful for displaying sensitivity analysis results by one interviewee. General column and bar charts were deemed particularly useful for displaying scenario results side by side ("cross-scenario reporting") for direct comparison with each other or a base case. Simple tables were also used to display resulting KPI values and to display scenario results side by side for easy comparison.

Confidence intervals were provided depending on the client; not all of them required it and it was suggested that their significance was often downplayed. Indeed, one interviewee stated that he normally provided confidence intervals but did not call them confidence intervals, instead opting to refer to them as lower and upper bounds. Another interviewee found confidence intervals to be less important than percentiles as a measure of variation.

4 SUMMARY AND DISCUSSION

These interviews with DES practitioners were exploratory in nature, rather than being intended to test any preconceived theories about the real world practice of DES experimentation.

It was usual for all but two of the interviewees to conduct the experimentation phase. The common thread running through all eight interviews was the overwhelming practice of comparing alternatives. Scenarios or alternative system configurations were defined either at the beginning of the project or during its course, and could be provided by the client, developed by the practitioner, or generated through consultation between both parties.

An unexpected finding was that the optimisation software now regularly included in DES software was rarely utilised, although optimisation (informal or otherwise) did emerge as the only formal search experimentation technique that saw any use. The general ease of use, ease of adaptability and time taken for the software to obtain a solution were the main reasons given for the reluctance to utilise this kind of software.

Although most of the interviewees had heard of the concept of meta-modelling (after it was explained to some of them), none of them practised it. In general, there was a lack of knowledge on the time-saving advantages of meta-models, their applicability to both search experimentation and sensitivity analysis, as well as how they can be validated to be as good as the simulation model upon which they are based (Friedman and Pressman, 1988). Improving understanding of this technique in combination with software support for meta-modelling, could greatly improve the search of the solution space.

There was evidence to suggest that consultants do conduct experimental design, albeit not always formally. However, all the interviewees carried out sensitivity analysis, and did so regularly. It was evident that the need for sensitivity analyses is well-understood by all interviewees (if not some of their clients), but they seemed unaware of the potential to speed up the process through experimental design or meta-modelling.

The running times of the simulations were deemed to be a major impediment to experimentation. This common response underscores the need for better design of experimentation and faster, easy-to-use software features to speed up experimentation.

In general, results analysis tended to be very basic and sometimes failing to account for statistical issues in estimation of and the uncertainty that surrounds the difference in scenario performance. This is a matter for concern as it could lead to a client regarding scenario differences as 'black and white' as well as missing valuable statistical information that helps inform understanding of the practical differences between scenarios; it also pinpoints another area for potential improvement in education and software.

5 CONCLUSION

In conclusion, the exploratory interviews were a rich source of information and suggest that the practice of experimentation does indeed continue to deviate from theory despite the advances in DES software. The interviewees displayed a reasonably good understanding of various aspects of DES theory, even if they did not practise it. However, there was a general lack of appreciation for formal search experimentation procedures, and practitioners tended to shy away from them in favour of comparing alternatives. Time was the major impediment. Software that was not up to speed with the latest theoretical developments meant that practitioners missed on opportunities to perform better, faster experimentation even if they were aware of them. A key revelation was the apparent failure of optimisation software to increase the use of optimisation.

Overall, these findings indicate a significant gap between the practice and academic theory of experimentation, and that both practitioners and clients could be far more aware of the advantages of search experimentation and statistically sound results analysis. However, it also poses the question of whether academics in this research field are sufficiently aware of and sensitive to the needs and pressures encountered by DES practitioners, in order to bridge this apparent gap.

To further explore the issues raised by this research, a larger more statistically robust survey instrument in the form of an online questionnaire is being constructed. The creation of this questionnaire has been guided and informed by the results discussed in this paper.

ACKNOWLEDGMENTS

We especially wish to thank the interviewees for their generosity in giving their time and expertise to help us conduct this research. The majority of this research was first carried out by Ashwin Rajaskanda in part fulfilment of the Degree of MSc in Management Science and Operational Research at the University of Warwick (2011), under the supervision of Dr K. Hoad and F. O'Brien.

REFERENCES

Banks, J., Carson, J. S., Nelson, B. L. & Nicol, D. M. (2011). *Discrete Event System Simulation,* Pearson Education.

Belton, V. & Elder, M. (1994). Decision support systems: Learning from visual interactive modelling. *Decision Support Systems,* 12**,** 355-364.

Christy, D. P. & Watson, H. (1983). The Application of Simulation: A Survey of Industry Practice. *Interfaces,* 13(5)**,** 47-52.

Fildes, R. & Raynard, J. (1998). The foundations, development and current practice of OR. *Journal of the Operational Research Society,* 49**,** 304-443.

Friedman, L. W. & Pressman, I. (1988). The Metamodel in Simulation Analysis: Can It Be Trusted? *Journal of the Operational Research Society,* 39(10)**,** 939-948.

Fu, M. C. (2002). Optimization for Simulation: Theory vs. Practice. *INFORMS Journal on Computing,* 14(3)**,** 192-215.

Hoad, K. & Monks, T. (2011). A note on the use of multiple comparison scenario techniques in education and practice. *In:* S. Jain, R. R. C., J. Himmelspach, K.P. White, and M. Fu (ed.) *Proceedings of the 2011 Winter Simulation Conference.* Institute of Electrical and Electronics Engineers,Inc, Piscataway, New Jersey.

Hollocks, B. W. (2001). Discrete-event simulation: an inquiry into user practice. *Simulation Practice and Theory,* 8**,** 451-471.

Law, A. (2007). *Simulation Modeling and Analysis,* New York, McGraw Hill.

Robinson, S. (2004). *Simulation: The Practice of Model Development and Use,* Chicester, John Wiley.

AUTHOR BIOGRAPHIES

KATHRYN HOAD is an assistant professor in the Operational Research and Management Sciences Group at Warwick Business School. She holds a BSc (Hons) in Mathematics and its Applications from the University of Portsmouth, an MSc in Statistics and a PhD in Operational Research from the University of Southampton. Her research interests include Discrete Event Simulation Practice, Output analysis techniques and experimentation.

THOMAS MONKS is an associate research fellow in the Peninsula Medical School, University of Exeter. He holds a BSc (Hons) in Computer Science and Applicable Mathematics from Staffordshire University, an MSc in Operational Research from Lancaster University and a PhD in Operational Research from Warwick Univesity. He has worked as both a Software Engineer in the private sector and an Operational Research Analyst within the public sector. His research interests include Discrete Event Simulation Practice, Reuse and Management Learning.

FRANCES O'BRIEN is an associate professor in the Operational Research and Management Sciences Group at Warwick Business School. She holds a BSc (Hons) in Mathematics and Classical Studies from Surrey University and an MSc in Operational Research from the University of Southampton. Prior to taking up an academic career, she worked for Ford of Europe as an Operational Research Analyst (simulation & LP modeling). Her current research interests include the practice of Operational Research and the development of frameworks, methods and models to support corporate strategy development.