Proceedings of the 2006 Winter Simulation Conference
L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, eds.

LESSONS IDENTIFIED FROM DATA COLLECTION FOR MODEL VALIDATION

Alan Cowdale

Operational Analysis Element
Air Warfare Centre
The Thomson Building
Royal Air Force Waddington
Lincoln LN5 9WA, ENGLAND

ABSTRACT

Computer simulation models will ideally be developed within an environment where all the necessary input data is readily available and all the relevant stakeholders are supportive and co-operative. In practice, many models are developed for the purpose of evaluating organisational or process change. As a result, system performance data may often be limited or potentially biased by key parties associated with the results of the simulation model. Ensuring valid input data is therefore a key aspect of model validation and an important component in a successful simulation study. This paper addresses some examples of data collection problems that have been encountered by the author and presents a number of lessons identified.

1 INTRODUCTION

Validation, in a general sense, is the process of assessing analytical capability against standards defining a fully credible system. For a simulation modeller, validation is establishing that the tools, data and models available for a study in a particular domain are capable of providing a credible simulation of the respective systems. This definition acknowledges that a set of validation criteria for the model alone is of limited value since there is almost always a clear linkage between the model and its data, with the latter very often needing to be tailored to reflect its intended use within the model. Therefore, both the raw data and the pre-processing needed to tailor the data for use should be included within the validation process. As a result, validation of a simulation model needs to embrace both the process represented within the model, and the associated assumptions and data items.

Data validation is often not considered to be part of the model validation process because it is usually difficult, time consuming, and costly to obtain sufficient, accurate, and appropriate data (Sargent 2003). As a result, the lack of data validation is often the reason that attempts to validate a model as a whole fail.

The Operational Analysis Element (OAE) of the Air Warfare Centre provides the scientific support to front line units of the Royal Air Force, and consequently is required to develop, maintain and use simulation models and tools for addressing a range of military problems. Over a period of time the OAE have accumulated a number of 'lessons learnt' associated with data validation, and have developed a number of 'rules' to reduce the resultant risks.

2 TYPICAL SOFTWARE LIFE CYCLE

Many simulation models and tools operate within a Simulation Life Cycle (SLC) illustrated in Figure 1. The typical SLC scenario will occur in the following manner.

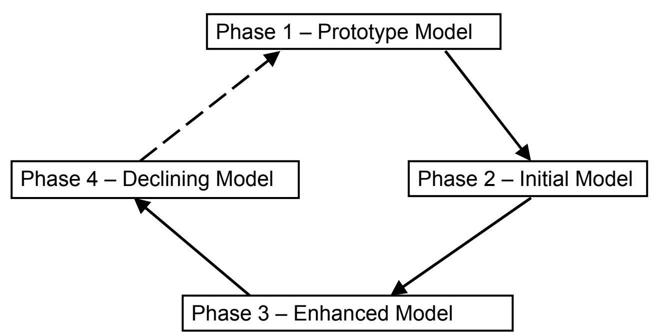

Figure 1: Simulation Life Cycle

<u>Phase 1 consists of the prototype model.</u>

1. The customer has a problem for which a simple simulation modelling has the potential to provide insights.
2. A prototype model is produced, probably a deterministic model in the form of a spreadsheet using highly aggregated input data.

3. The model is easy to use, intuitive and produces simple outputs. The model can be used successfully by either the analyst or the customer.
4. The model is used to produce illustrative results and the customer is persuaded to sponsor the development of a more representative model to address the actual problem.

Phase 2 consists of the initial model.

1. The problem is researched in more detail to understand the processes involved and the input and output measures.
2. A representative model is produced, probably a stochastic model utilising a simulation modelling package.
3. The model is reasonably easy to use, generally intuitive and produces a range of statistical outputs. The model can be used successfully by most analysts or the customer following basic tuition.
4. The model is used to produce representative results for the system.

Phase 3 consists of the model extension phase.

1. The model is modified to either address a slightly different problem or different aspects of the original problem.
2. The extended model is produced, possibly introducing additional simulation modelling features.
3. The model is quite complex to use, requiring a reasonably high level of understanding of the simulation model being employed, and produces a range of statistical outputs. The model can only be used successfully by a small number of dedicated analysts.
4. The model is used to produce representative results for the system, but the outputs invariable require post-processing before presentation to the customer.

Phase 4 comprises the model decline phase.

1. The model is now very complex and any modifications to either address a slightly different problem or different aspects of the original problem take significant time and effort.
2. The model is complex to use, requiring a very high level of understanding of the simulation model being employed, and produces a range of statistical outputs which require considerable post-processing .
3. The model is used to produce results for the system, but the outputs are not conveyed directly to the customer due to complexity and non-intuitive nature of the results.
4. The customer has misgivings about the model and requests that the potential for a much simpler model be investigated, probably a deterministic model in the form of a spreadsheet using highly aggregated input data.

3 SIMULATION MODEL VALIDATION

Many of the issues illustrated in this SLC scenario can be related to the validation of the simulation model. The Phase 4 decline of the model is often precipitated by the customer perception that the simulation is no longer providing credible results. It is therefore clearly in the simulation modeller’s best interests to have the process and the evidence available to pursued the customer that the model remains ‘fit for purpose’ (Sadowski 2005).

There are many books and papers dealing with methodologies for simulation validation (for example Law and Kelton 2000). The majority of the methodologies will involve a combination of:

1. Conversations with subject matter experts.
2. Comparison with existing theory.
3. Comparison with observations of the current system (Lada et al. 2005).
4. Intuition.

The simulation model will however only ever be a representation of the full system, and as such the key decision for the modeller and the customer is invariable ‘is the model valid enough for the simulation study being undertaken?’. Whilst the fidelity of elements within the simulation model can be enhanced as the model matures, high quality input data will need to be applied throughout the SLC.

4 INPUT MODEL DANGERS

A key aspect of model validation is to ensure the accuracy of the input data being used within the simulation. The use of poor quality input data significantly increases the risk of the simulation study providing incorrect information to the customer and decision maker. This can be illustrated through a simple queuing simulation example. The following example has been constructed using SIMUL8.

Consider a single server queue for which the customer arrival rate is reported to be 10 per hour and the service time is reported to be 6 minutes (Figure 2). If customers were to arrive at a constant rate (i.e. one every 6 minutes) and the server takes exactly 6 minutes to deal with the customers request, then the customers would never have to wait and the server would have no ‘idle’ time.

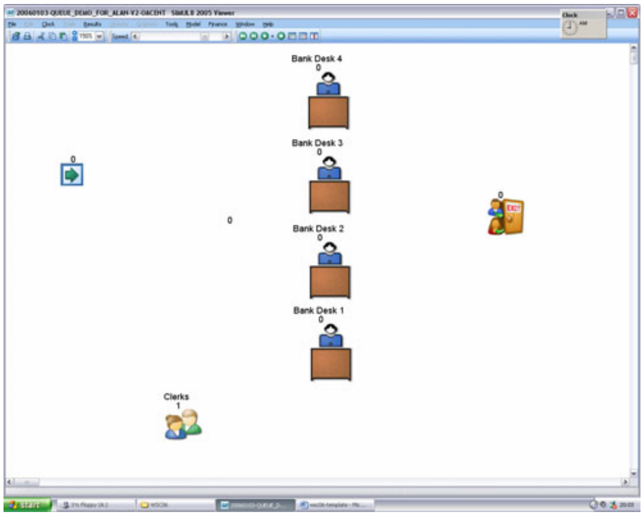

Figure 2: Simple Queuing Simulation

Simulation modellers will immediately identify that the model does not take into account any random variations in either the customer arrival times or the time taken by the server to meet the customer's requests.

The importance of considering randomness in the model can be illustrated by considering the effect of changing the customer arrival time from exactly 6 minutes to a value sampled from a Normal Distribution with a Mean of 6 minutes and a Standard Deviation of 1 minute. Under these circumstances, the average waiting time is just over 4 minutes with a maximum waiting time of just over 10 minutes. Similarly, if the server time is modeled as a Normal Distribution with a mean of 6 minutes and a Standard Deviation of 1 minute, then the average waiting time increases to over 5 minutes, with a maximum waiting time of over 13 minutes.

The response from the server's management to seeing this type of result, could be to consider introducing new processes or retraining the server to reduce the time taken to meet the customer's request. The model could be used to identify that reducing the time taken to meet the customer's requests to a Mean of 5 minutes and a Standard Deviation of 1 minute would result in only minimal waiting time for the average customer, and a maximum waiting time of around 3.5 minutes.

This level of analysis may however result in misleading conclusions. If the variability in customer arrival times were also to be considered, for example by obtaining customer arrival times by sampled from a Poisson Distribution with a mean of 6 minutes, a single server would struggle to meet the customer requirements. A server meeting the customer's requests to a Mean of 5 minutes and a Standard Deviation of 1 minute could still result in significant waiting time for the customers. The server processing time would need to be reduced further to reduce the maximum customer waiting time to a minimal level. Under these circumstances, introducing an additional server may be a more practical proposition.

Although the customer arrival rate was reported at 10 per hour, if this arrival rate is not randomly distributed throughout the day, but has particular peak times, then again the simulation results may lead to misleading conclusions. For example if the hourly arrival rate of Table 1 were to occur, then 2 servers would result in an average time in queue of just over 8 minutes with a maximum of around 14 minutes, although both servers would be 'idle' for over 1/3 of the day. In these circumstances, having 2 servers for the first and last hour of the day, and three servers for the midday peak hour would result in minimal average wait time and a maximum wait of less than 4 minutes.

Table 1: Customer Arrival Rate

Time period	Customers per hour
9-10	16
10-11	6
11-12	6
12-1	24
1-2	6
2-3	6
3-4	6
4-5	10

5 INPUT DATA ISSUES

While the previous queuing example is artificial, it effectively illustrates the importance of obtaining and implementing the correct input data. In practice, obtaining valid input data is often difficult and is rarely given the same priority as other aspects of the simulation model development.

Input data issues can take several forms, covering the full spectrum from too little to too much available data.

1. Data not available.
2. Data not easily available.
3. Data incomplete.
4. Data inaccurate.
5. Data misleading.
6. Data deluge.

5.1 Data Not Available

Having no data can result from a number of reasons, but usually it is because the simulation model is being designed to address a new system or new process and hence there is no data available. When there is no data available, the modeller can estimate the input parameters by using a combination of;

1. Conversations with subject matter experts.

2. Comparison with existing theory.
3. Comparison with observations of the current system.
4. Comparison with similar models and studies.
5. Intuition.

5.2 Data Not Easily Available

While data may potentially be available, it may be too expensive either in time or cost to collect. In a military environment this is a common situation, when, for example, data from training exercises could be used for model inputs, but recording and collating data from all the participants can be hugely expensive a time consuming. If it is impractical to collect the data, then the modeller will need to estimate the input parameters as if no data was available.

5.3 Data Incomplete

In circumstances when data is available, it will often be incomplete. Data may for example have only been collected over a short period of time, or from a limited sub-set of the overall system. Incomplete data can be used to estimate input parameters using appropriate statistical techniques, or by using directly within 'bootstrapping' methods. While statistical techniques can be used to estimate the confidence intervals for the input parameter estimates, for these to be valid, the data set must be a representative sample of the overall system

5.4 Data Inaccurate

When data is available, there is a risk that it is inaccurate. This may be because the fidelity of the data is not a high as that required by the modeller, or even that the data has been corrupted or is in error. Gross inaccuracies are often easy to identify, but data recording issues such as using 'yards' rather than 'metres' are much harder to identify.

5.5 Data Misleading

When accurate data is available, it may be misleading to the modeller. Misleading data may result from a number of causes; either the data that has been collected is not what the modeller was expecting, or the true system is more complex that that envisioned by the modeller. Misleading data invariably means that additional effort and data analysis will be required by the modeller and will impact the study timelines.

5.6 Data Deluge

In the electronic age, many automated or semi-automated systems keep detailed data logs. While the data may be available and easily accessible, extracting the relevant data and processing the data into a usable format may involve considerable time and effort.

6 DATA COLLECTION

When considering the collection of input data, the analyst has got three options.

1. Use pre-collected data.
2. Get somebody else to collect the data.
3. Collect the data himself/herself.

6.1 Use Pre-Collected Data

Using data that has already been collected is always the easiest method of obtaining input data for the simulation model. If the data is being reproduced from a previous study or from the academic literature it will bring with it a level of peer review and accreditation, and takes little additional effort to convert into a form suitable for the model to use. There are however a number of risks associated with using pre-collected data:

1. The data may not be representative of the scenario to be modeled.
2. There may be implicit assumptions associated with the data.
3. The accuracy and fidelity of the data may be different to that needed in the model.

6.2 Get Somebody Else to Collect the Data

Having somebody else collect the data you require as part of a wider data management process has a number of attractions. The principal attraction is the small cost in terms of time and effort for the modeller, but also the added credibility of being part of a wider organisational process. The risk with this method is that your data may not be a high priority element, and hence may not be collected with the quantity, quality and completeness that the modeller requires.

6.3 Collect the Data Yourself

If the modeller collects their own data, this will be time consuming and detract from the model development activities. Collecting your own data however ensures that the modeller is assured of the quality, completeness and accuracy of the data being used within the model, and probably more importantly, the process of collecting the data will often provide additional insights into how the current system actual performs.

7 REAL WORLD INPUT EXAMPLES

Data is defined as a series of observations or measurements (Collins 1988). When the data are placed into context they become Information. The addition of judgement to the interpretation of the Information constitutes Knowledge. Whilst inputs to the simulation model will consist of data, knowledge will invariably be required to ensure the data is relevant for the analysis task.

7.1 Weapon Accuracy Example

The accuracy of free-fall bombs has traditionally been modelled in simulations as a bi-variate Normal Distribution. By examining down-range and across-range miss distances for different aircraft and weapon combinations obtained from trials, exercises and real military operations one would expect that it should be relatively simple to select the most appropriate mean and standard deviation parameters for the bi-variate Normal Distribution.

In practice, examination of raw miss-distance data often generates a number of problematic issues. Principal amongst these is that the raw data rarely looks 'Normal'. A typical plot of down-range miss distance will tend to be similar to Figure 3, comprising a large 'tail' to the distribution and with a significant number of 'zeros'. The large tail is invariably caused by data points where the aircrew are aiming the weapon at a target location which is different to the aim-point from which the miss distance is being measured, and the large number of 'zero' misses is invariably the result of using a non-point target. If the target is relatively small (say a tank or an artillery piece) then target size is probably not a major factor, but some targets, for example large buildings, can be of the same order of magnitude as the expected miss distances. The RAF occasionally uses Garvie Island off Cape Wrath in North Eastern Scotland as a bombing range during exercises. When the target is a large granite island, unsurprisingly almost all the weapons on these exercises 'hit' the target.

7.2 Aircraft Reliability Example.

A squadron of aircraft, comprising 12 airframes, deploy to an operational theatre. During the tour of duty, the squadron flew 8 missions per day for 10 days. Examination of the aircrew Mission Reports showed that 4 missions involved ground aborts (i.e. an aircraft was unable to launch due to system faults). A simple examination of the data could suggest that with 4 ground aborts in 80 missions (5%) a simulation model of the system could assume an input value of 95% for aircraft reliability.

An analyst with knowledge of the system would however know that 95% is probably a significant overestimate of aircraft reliability. This is because the ground abort data has been derived from aircrew Mission Reports. Of the 12 aircraft deployed, the flight engineers will only offer up to the aircrew airframes which are believed to be fully serviceable. Consequently the 95% value, is the reliability of aircraft believed to be serviceable when handed over to the aircrew. To obtain a true value for aircraft reliability the analyst would need to have obtained the engineering status reports for each of the individual airframes deployed on the operation.

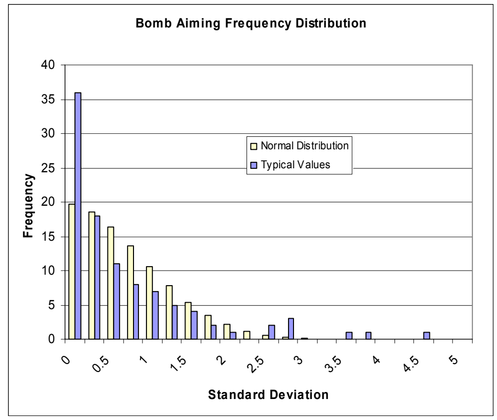

Figure 3: Typical Miss-Distance Distribution

7.3 Mission data accuracy example

Since MISREPs are filled in by the aircrew, the quality of the data is totally depended on the accuracy of the data provided by the individual filling in the data form. During Operation GRANBY (Iraq 1991) the overall quality of MISREPs tended to be poor, with a significant number of forms either incomplete or inaccurate. Following Operation GRANBY the OAE undertook a major education campaign amongst the RAF aircrew to improve the quality of MISREP reporting. This involved liaison with the squadrons, providing data collection tools and offering feedback and evidence of the value of the data that was being collected.

As a result, by Operation ALLIED FORCE the quality of the RAF MISREP reporting was considerably increased. Data collectors need to remain aware however of the circumstances under which the data is being collected. For example, during Operation ALLIED FORCE, the quality of the data collected by the Harrier GR7 crews flying from Italy tended to be higher than from the Tornado GR1 crews flying from Germany. This should not have been a surprise however, since the Tornado crews were tending to fly missions lasting over 6 hours longer than the Harrier crews (due to overland routing issues to reach Kosovo). As a result the Tornado crews were filling in the MISREPs and data collection forms at the end of a 14 hour day in a highly stressed environment, and unsurprisingly attention to detail when filling in the forms was not always a high priority.

7.4 Weapon Usage Example

During Operation ALLIED FORCE (Kosovo in 1999) and IRAQI FREEDOM (2003) the OAE were tasked to maintain a database on air-weapon expenditure by RAF aircraft. During these Operations, the RAF was operating from a number of bases, and the OAE had insufficient manpower to maintain a data collection team at all of the bases, all of the time. Consequently, the database was compiled principally from aircrew Mission Reports (MISREPs). Relying solely on MISREPs would however had been problematic however, since cross-referencing with armourers logs and engineer's authorization sheets occasionally identified missions that had been flown and weapons released form which the OAE had not collected a corresponding MISREP.

7.5 Heisenberg's Principle Example

In predominately human based systems, it is quite possible that collecting the data will raise the level of awareness of the process or system amongst the various stakeholders. In these circumstances the data collection may actually alter the process or the system itself.

In the fall of 2002 the Fire Brigade Union within the United Kingdom proposed a ballot on strike action in support of a wage demand. Consequently the Ministry of Defence agreed that in the event of strike action, emergency assistance to the local authorities would be provided by deploying military staff to man fire-fighting equipment and to provide command and control functions. The OAE developed two simple simulation models to investigate asset utilisation and to act as a training aid for the military mission planners and control centre staff in the Norfolk region (Cowdale.2003). The Norfolk Fire Service and the Office of the Deputy Prime Minister had kept detailed records on the type and location of fires which had historically occurred in the Norfolk region, and hence this data was considered an excellent source of input data for the simulation models. During the days of industrial action although the number of malicious false-alarms increased, the overall number of reported incidents in the Norfolk region (and within the United Kingdom as a whole) were significantly less than the historical predictions. Whilst bad weather and heavy rain invariably contributed to the number of fire incidents, the build-up to the industrial action received significant media coverage and resulted in a major increase in public awareness of fire risks and preventative measures.

8 CONCLUSION

If you are in a position to define a data collection plan.

1. Think very hard about what you want.
2. If in doubt collect it
3. Make sure you are collecting what you think you are collecting
4. Ensure you document what you collected and what you didn't
5. If possible confirm via two sources

Data collection and model validation, is nothing new. In a letter to Nathaniel Hawes dated 25 May 1694, Sir Issac Newton wrote "If instead of sending the observations of able seamen to able mathematicians on land, the land would send able mathematicians to sea, it would signify much more to the improvement of navigation and the safety of men's lives and estates on the element." .

REFERENCES

Collins, *Reference English Dictionary*. HarperCollins 1988.

Cowdale, A. C. Simulation modelling in the support of emergency fire-fighting in Norfolk. *Proceedings of the 2003 Winter Simulation Conference.* Editors Chick, Sanchez, Ferrin and Morrice 1707-1710.

Lada, E. K., N. M. Steiger, M. A. Wagner, and J. R. Wilson. Introduction to modeling and generating probabilistic input processes for simulation. *Proceedings of the 2005 Winter Simulation Conference.* Editors Kuhl, Steiger, Armstrong, Joines. 41-55. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Law, A. M., and D. W. Kelton. 2000. *Modelling and analysis.* 3rd Edition. New York: McGraw-Hill.

Sargent, R. G. Verification and validation of simulation models. *Proceedings of the 2003 Winter Simulation Conference.* Editors Chick, Sanchez, Ferrin and Morrice 37-48. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Sadowski, D. A. Tips for successful practice of simulation. *Proceedings of the 2005 Winter Simulation Conference.* Editors Kuhl, Steiger, Armstrong, Joines. 56-61. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

AUTHOR BIOGRAPHY

ALAN COWDALE is the senior scientific advisor at the Air Warfare Centre in Lincolnshire, UK. He holds a BSc in Mathematics from the University of Southampton and an MSc in Operational Research from the University of Birmingham. He has undertaken a wide range of simulation projects for both the Royal Navy and Royal Air Force in connection with assessing weapon system performance and effectiveness.