A Critique of the Live Synthetic Trials Balance to Support the Smart Acquisition Cycle – Better Dead than Alive?

Michael Kelly
QinetiQ
Cody Technology Park
Room 2067/04 Building A5
Ively Road
Farnborough
Hampshire
GU14 0LX
United Kingdom

mkelly1@qinetiq.com

ABSTRACT

A recent project was carried out for UK MoD to assess the advantages and disadvantages of changing the balance of trials, exercising and experimentation versus modelling and simulation techniques (including Synthetic Environments (SEs)) used in support of the various stages of the 'smart' equipment procurement regime. The findings from which have generated the debate presented here into the advantages and disadvantages of field trials over simulation techniques. Anecdotal and substantive evidence on the effectiveness of Synthetic Environments (SE) is presented, drawn from recent history of their use in significant modelling and simulation events. These events include the Synthetic Environment Coordination Office national capability demonstrator programmes, synthetic environment experiments, and Niteworks programmes among others. The modelling errors that can occur in SE's and the Verification, Validation and Accreditation issues are considered. A critique of field trials and experimentation is also presented highlighting experimental design limitations and human factor issues arising from them. The gathering of data (objective and subjective), the experimental designs and the cost issue of field trials are addressed. The two broad analytical domains (trials and SEs) are then compared and suggestions are made for selection of one technique over the other for the various phases of the CADMID cycle (Concept, Assessment Demonstration, Manufacture In-Service and Disposal). Recommendations are also made for modification to the UK defence equipment procurement business process to maximise the effective use of supporting trials of modelling techniques. The evidence presented suggests that the assumed superiority and frequent choice of field trial's over SE modelling techniques needs more careful and substantive consideration than it commonly receives.

1.0 INTRODUCTION

A recent research programme sought evidence for the superiority of manned field trials over synthetic environments in data gathering. This paper seeks to highlight the lessor known failings in field trials along side those of SEs, demonstrate that it is an arbitrary dichotomy and highlight where synthetic environments can be of benefit in the MoD procurement process.

Kelly, M. (2005) A Critique of the Live Synthetic Trials Balance to Support the Smart Acquisition Cycle – Better Dead than Alive? In *The Effectiveness of Modelling and Simulation – From Anecdotal to Substantive Evidence* (pp. 15-1 – 15-10). Meeting Proceedings RTO-MP-MSG-035, Paper 15. Neuilly-sur-Seine, France: RTO. Available from: http://www.rto.nato.int/abstracts.asp.

2.0 RECENT RESEARCH

A research study was carried to examine the way that geo-spatial and temporal information is generated, managed and exchanged (across all three domains) - with a view to refining the way that MoD currently procures equipment, thus ensuring that the future real time Networked Enabled Capability (NEC) need for accurate and timely information can be met (and maintained). The study was largely predicated on the premise that there is a need to move more towards the use of real world measurement techniques using independent ground truth as opposed to modelling and simulation techniques, where it is both appropriate and practical. This real world measurement can then be used to assess performance against agreed standards or levels of required capability. It was believed that such a move would lead to a number of important benefits: improved rigour during the acquisition process; assurance that the necessary level of system performance or capability is being achieved and that in-service equipment is operating effectively ensuring that maximum military capability is continuously made available.

A subtask within this study was to examine the UK Defence Procurement Agency smart procurement process to see if there was a case for changing the business process to rely more on man in the loop field trials and exercises (so called ground truth based assessment) and less on modelling and simulation techniques. The simulation technique 'Synthetic Environments' (SE's) had come under particular scrutiny.

This paper expands the debate generated by the part of the study which was addressing the pros and cons of both Modelling & Simulation (M & S) techniques and man in the loop techniques.

3.0 THE BUSINESS PROCESS

The new Defence Procurement Agency (DPA) business process focuses on managing procurements through six key stages; Concept, Assessment Demonstration, Manufacture In-Service and Disposal (CADMID). While Synthetic Environment Based Acquisition (SEBA) had been suggested to contribute to a faster, cheaper and better procurement [1] the objective evidence was limited. However the recent trend in procurement was away from demonstrably expensive field trials ("live" trials) to usually cheaper modelling and simulation techniques ("dead" trials) including SEs.

While the inherent inaccuracies and assumptions in modelling and simulation were by and large well known and are reviewed in this paper; similar inaccuracies and assumptions in field trials have in many cases been ignored.

4.0 THE USE OF FIELD TRIALS

The range of trials that could be covered by use of the "real world measurement techniques using independent ground truth" is considerable and covers a large spectrum of human in the loop trials (Table.1.). This spectrum ranges from laboratory man-in-the-loop trials with real or simulated systems (including, prototypes sub-systems or equipment components), through minor field trials of manned equipments to full field trials or instrumented exercises of multiple manned platforms interacting with each other in a real synthetic military operation. All require control of variables, accurate measurement and statistical analysis to determine the significance of any results.

It should be noted that as the complexity of a field trial goes up so does the cost and that the increase in complexity proportionately reduces the experimental control. There is an order of magnitude increase in cost the moment one steps out of doors on to the ranges and if significant platforms are involved like ships and planes then the costs become very large.
Despite these costs (or because of them) every attempt should be made to introduce as much rigour into the trials design and data collection as possible, to ensure any results have some scientific validity over casual observation.

Table 1: Spectrum of man-in-the-loop trials

Technique	Example	Benefit	Disadvantage	Notes
Laboratory based equipment trials	Selecting optimum man–machine interface alternatives in Command & Information Systems displays. Optimum altimeter display design	Manipulate independent variables, measure effect on dependant variable, control spurious variables.	Impact of,& interaction with wider system not tested	Single or multiple subjects. Simple, low cost & repeatable
Laboratory based systems trials	Full cockpit/ or crew station design trials	Manipulate independent variables, measure effect on dependant variable, control spurious variables.	Impact of,& interaction with wider system not tested	Single or team subjects. Low cost and repeatable
Single platform Field trials	Prototype vehicle trials	Manipulate some independent variables, measure effect on dependant variable.	Effect of uncontrolled & spurious variables not known (weather fatigue, terrain, crew fatigue etc)	Cost going up. Not repeatable. Hard to get statistical significant subject populations or measurement repeats. Significant cost.

Multiple platform field trials	NEC field studies (e.g. naval gunfire trials)	Examine interdependencies between systems	Many critical factors not observed or recorded. Little control of independent variables, spurious variables dominant	Complex high cost. Repeats unlikely to provide reproducible results
Fully instrumented exercises	US Army Force XXI experimental series	Gain snapshot of near real system under operational pressure	Complex analysis required. Little control of independent variables, spurious variables dominant	Complex high cost, unique event & unrepeatable.

Field trials are seen as inherently more realistic because they have people in them doing their real tasks in apparently realistic situations, but as these field trials (ground based truth trials) approach some idealised reality there is a corresponding increase in a number of sources of experimental error.

All experiments vary as to whether subjective or objective measurements are taken and the degree to which individual differences are corrected for either by pre-screening for a particular variable or by providing statistically large enough subject populations to level out any significant individual difference.

The common errors found in field trials include:-

- Too few repetitions of the trial or conditions to gain statistical significance in results. This is usually for cost, or other resources limitations (e.g. crew, platform, range, scientist or measurement equipment availability).
- Too small subject populations for statistically reliable trials. The availability of some key operators, platform or crews is finite and severely limited.(e.g. Battle group commanders, Tornado crews, Destroyer crews etc).
- Too few systems to obtain statistically reliable trials. The variability of some platforms or technical systems is finite and severely limited. (e.g. aircraft carriers, nuclear submarines etc).
- Cheating. In an effort to perform at their best, subjects may try to improve their performance, in ways facilitated by poor trials conditions. Alternatively subjects may fake poor performance in an attempt to get withdrawn from the trial.
- Health and Safety limitations. H & S limitations such as range safety areas may limit the area or time frame in which targets can appear, allowing subjects to pre-select areas of interest.

- Lack of battlefield stressors. Welfare considerations during trials will prevent operators and crews during trials from being under the typical battlefield stressors of sleep deprivation, fatigue, fear etc.
- Optimally calibrated/serviced equipment. Equipment specially prepared for trials may be of a higher performance standard than that typically found on operations.
- Limited data capture. Many factors may conspire to prevent the amount of data captured from matching the data capture plan.
- Limited analysis. Large amounts of data may be prohibitively costly or take too long to analyse (typically video data can take 4 times as long to analyse as is taken to capture it).
- Subjective data capture. Data consist of collection of the opinions of SME who are often key stakeholders in the trials outcome and who are observing the conditions 'blind'.
- Data captured only from idealised or easy data points in the systems – not using enough observers to pick up human error.
- Atypical subject populations (using the best personnel available when the trials are considered important or the worst personnel as quality individuals are committed to other duties, and the trials are considered unimportant).
- Capturing subjective data and then analysis to give the feel of objectivity. The use of 'Likert' scales to convert opinion into numbers and then carrying out statistical analysis on the results can give a feeling of objectivity to the data.
- Too many independent variables (too many conditions). Frequently when availability of assets or crews is limited there may be a tendency to combine conditions.

Thus we can see that when it is considered as an isolated technique the use of ground based truth trials has many failings. Such trials can provide many useful insights and data when these limitations are understood. However if we also consider modelling and simulation techniques and focus on key M & S techniques we can see they have essential contributions to make.

Table 2: Spectrum of Modelling & Simulation Techniques

Technique	Example	Benefit	Disadvantage	Notes
Mathematical models	Bridge stress calculations	Tabular output or simple graphics. Deterministic relationship between variables.	Examines small part of system in isolation.	Low cost, simple, & repeatable
Independent simulation	Tank kill probability model	Graphical output. Stochastic & deterministic data	Isolated system or subsystem	Low cost and repeatable.

Constructive simulation	JANUS war game	Data driven medium complexity model to examine complex controller behaviour	Mainly aggregate level interactions. Simplistic behaviours underpinning apparently sophisticated interactions	Low cost. Not repeatable. Output depends significantly on operator's actions
Synthetic environment	STOW	Highly visible outputs. High face validity. Multiple platforms interacting in near real time.	High face fidelity hiding poor behavioural modelling. Many critical factors not observed or recorded	Complex medium/high cost. Repeats unlikely to provide reproducible results

Table 2 examines the modelling & simulation spectrum, in a similar fashion to that used for man-in-the loop trials. It is worth noting that the divisions between examples of modelling and simulation are no less clear that for man the loop trials and many examples of hybrid and interim categories can be found.

5.0 SYNTHETIC ENVIRONMENTS

The focus of this paper is to high light the errors found in field trials and the surprising lack of rigour they may offer. These field trial errors must be contrasted with the more commonly understood modelling errors. The errors and assumptions in modelling and simulation are perhaps more widely understood and those specific to synthetic environments have been widely discussed leading to a recognised need to have a formal Verification, Validation and Accreditation process for such confederations (e.g. Moulding 1999 – [2]). VV&A is defined as; Verification - did we build the right thing?; Validation - did we build it right or is it doing what it is suppose to do?; and Accreditation; is it appropriate for its intended use? VV&A provides methodologies, procedures, tools and techniques for establishing the credibility of M&S alone or in federations and their contribution to trial results.

The apparent face validity of a well constructed synthetic environment with all its models and figures interacting in real time with one another can hide a variety of sources of error. However in general the typical errors often found in synthetic environments are:-

- Database correlation errors – flying tanks and buildings floating / below surface / road inconsistencies, road through building or is it building incorrectly placed; holes in terrain.
- Fair fight – inter-visibility; where an object is hidden in one database but visible in another.
- Unrealistic /simplistic behaviour modelling of computer generated opposing forces and/or allies, attritional warfare – last man standing.
- Inaccurate or simplistic hit and kill modelling.
- Scene overload, model's not drawn or drawn slowly due to processing limitations.

- Co-planar polygons leading to alternate walls being drawn.
- Incorrect models; i.e. a tank in one simulation is displayed as an armoured personnel carrier another (e.g. stealth view).
- Poor fidelity and update rates reducing psychological immersion to unacceptable levels.
- Failure to engage in any VV& A (incorporates many of the above issues).

Despite these limitations synthetic environments have been used successfully to support a number of research programmes and procurements. Some notable examples of complex SEs used in the UK include[1]:

- UK Synthetic Theatre Of War
- Exercise Purple Sound (1998- SE supporting Permanent Joint HQ planning)
- National Capability Demonstrators – SECO
  - Virtual Ship
  - Virtual Cockpit
  - Air defence
- Soft Vertical Launch – SE support to Matra BAe Dynamics missile programme
- Raven Explosive Ordnance Disposal - Support to DPA in the development of a bomb disposal robot concept
- Niteworks; UK SE experimental programmes

However a 1999 conference (RMCS 11th International SE Symposium) on SE presented a number of papers [1] proposing the use of SE in support of procurement (the term SEBA – Synthetic Environment Based Acquisition, has been coined). Many tool processes and pilot studies were discussed, but today the use of SE in procurement is still not widespread, although the approach is endorsed by the UK MoD and the Defence Procurement Agency.

The face validity provided by seeing entities moving through a visual environment in a synthetic environment provide an apparent credibility that symbols moving over a constructive map display do not afford, none the less, the parameters driving constructive models are often no more sophisticated in the SE.

Modelling and simulation, and in particular SEs produce an idealised model of a system. They are best suited to examining the benefits of a concept that is assumed to be operating in a perfect way. SEs show how good a system could be. Field trials with their imperfect human operators interacting with the system will stress the system and, if the correct data is collected show up the faults and any potential for inappropriate operation within a system. If their limitations are understood field trials can show you where systems can go wrong.

6.0 MODELLING TO SUPPORT THE CADMID CYCLE

Figure 1 shows the current ratio of modelling and simulation (SEs) to field trials at different stages of process.

- Concept - In the concept phase there is no hardware to trial so the most suitable techniques are modelling and simulation. It is also a low cost method to address a wide range of options.

[1] Further expansions on the aim and effectiveness of these SE's are given in the accompanying presentation.

- Assessment – In this phase there may be some prototype hardware to assess but M& S provides a suitable environment for considerable assessment. Competing SE models will often be idealised representations to favour a commercial contender's position during the assessment phase. Ideally UK MoD should own the base line SE used for assessment.
- Demonstration – Demonstration requires the existence of real equipment but trials may be supported by M & S.
- Manufacture – Though SEBA techniques can support manufacture this activity is essentially focused on real hardware.
- In-Service – Real hardware is typically used for experimentation and doctrine development.
- Disposal – M& S has not been used effectively in the evolution of disposal (including platform upgrades) however it an effective technique.

The ratios of M&S to field trials for software heavy procurements will differ as M&S techniques will favour most phases of the CADMID cycle.

While there have been attempts to introduce SEBA in more formal ways into the business process the use of synthetic environments in MOD business process has not changed significantly in the last decade. Modelling and simulation form a large part of the research underpinning the concept phase and are still employed in other phases of procurement, but the balance remains largely unchanged.

7.0 BROAD SPECTRUM OF TECHNIQUES OR TWO SETS OF TECHNIQUES?

In examining the "either or" argument between field trials or modelling and simulation techniques such as SE we are perhaps setting up an artificial dichotomy. Many data gathering exercises are a mix of manned systems and modelled components. Trials that use 'synthetic wrap' produce the deep battle enemy force picture through simulations to stimulate the command information systems whilst the close battle platforms are manned. It therefore may be more appropriate to see the whole experimental domain as a continuous spectrum from rigid mathematical modelling through synthetic environments to man-in-the-loop immersive SE, lab based man- in-the-loop trials and fully instrumented complex exercises, and beyond to collection of data in, or post, operations. Therefore it would be inappropriate to conclude that SEs are bad and Ground truth trials are good as both techniques have advantages and disadvantages as has been outlined in this paper. Rather it is better to conclude that a greater understanding of the relative merits of each approach need to be applied when developing the experimental construct and devising the test and evaluation criteria.

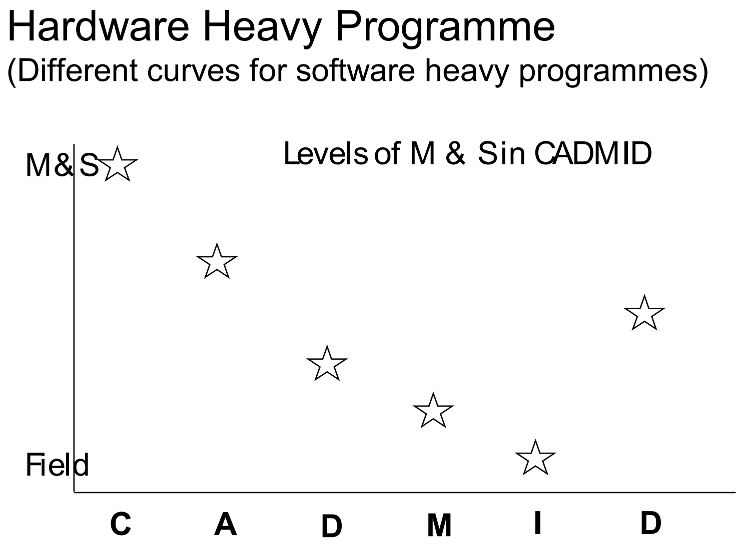

Figure 1. Ratio of Modelling & Simulation to Field Trials in supporting the CADMID Cycle

8.0 CONCLUSIONS

Since Synthetic Environments (SEs) are likely to play an ever increasing role in the attainment of a future NEC, it is important to understand and consider the issues, interactions and trade-offs available between the use of SEs and the measurement of real world system behaviour. Previous research has at times created a somewhat arbitrary dichotomy between field trials and M&S techniques which, we have argued, sit along a broad continuum of assessment techniques. Attempts to utilise existing SEs to examine detailed geo-spatial and temporal issues also highlighted that, of those assessed, the fidelity of current SEs is insufficient to support such activities. This does not reflect on the ability of these SEs to fulfil their original design purpose but instead reflects the fact that such SEs have not typically been created to adequately represent geo-spatial and temporal issues. In this respect it needs to be acknowledged that SEs have limitations and must be treated with a certain degree of caution and that these limitations need to be fully understood when using them for this purpose. If SEs are to be used it is important that due consideration is given to ensuring that they are populated with realistic error information and assumptions and that they are validated against real world data.

However as this paper seeks to make clear there are equally many reasons for caution in respect of using independent ground truth trials techniques. What is commonly less well appreciated is the severe limitation or scientific rigour that large scale and complex field trials impose.

Lastly the case for changing the MoD business process to rely more on using independent ground truth trials at all stages is not made. Many efforts have been underway to increase the uses of modelling and simulation in this process but the balance at various stages in the process remains largely unchanged over the decade with perhaps the use SEs increasing as our ability to produce suitably rigorous models increases and their limitations become better understood.

9.0 ACKNOWLEDGEMENTS

The author would like to acknowledge the help and support of Mr Gareth Pugh of QinetiQ in preparing this paper.

10.0 REFERENCES

[1] Parfitt, A., "Simulation based acquisition; future plans for defence procurement", Conference on Simulation Based Acquisition, Mar 2000, SMI Conference, London.

[2] 11th International Synthetic Environment Symposium. Royal Military College of Science. Shrivenham. UK. Oct 1999