Scenario Design for Verification and Validation, Hardware Testing and Operations Research

S. Waugh [a]

[a] *Aerospace Simulation Experimentation and Wargaming, Aerospace Capability Analysis, Joint and Operations Research Division, Defence Science and Technology Group.*
Email: sam.waugh@dst.defence.gov.au

Abstract: Modern military aircraft are complex with increasingly integrated sensing functionality. As mission system capability is increasingly implemented in software rather than hardware — particularly with the development of electronically steered antennas and increasing processing power for software implementations of former hardware components — the scope for specific data requirements and software loads that are tailored for missions is increasing. As the complexity of these loads increases, the operational effectiveness of each aircraft is increasingly dependent on each mission data load. Put simply: an aircraft with ten million lines of software, much of which is configurable by input data, has a significant part of its capability defined by that data.

How do we ensure that the aircraft mission systems perform as required in a complex operational environment? Further, how do we provide Operations Research (OR) to support experimental testing results? A task often given to OR personnel is to construct scenarios that provide context for all levels of experimental work, both hardware lab testing, and to support any OR experimentation that builds on these results. Any scenarios developed in this process need to provide both operational realism and sufficient levels of complexity to suit a testing program. This is problematic in that large scenarios that are generated by the selection of "important" features are by their very nature biased towards particular solutions. How do we avoid bias in the scenario generation?

A specific problem area is defined, namely testing the application of Intelligence Mission Data (IMD) to aircraft operations. How do you decide, with (1) a particular set of mission system hardware in a lab and a (2) reprogrammed mission data file set (MDFS) that provides the IMD in an operational form, that a fighter aircraft will be effective and survivable in a specific operational context. In testing IMD there is a need to consider all levels of the operational context as defined by a scenario:

- the strategic context gives purpose and the broader physical and electromagnetic environment via the specification of an area of operation,
- the operational context gives the goals or missions to be pursued, and
- the tactical flow can affect the specific mission performance as this describes the specific interplay of actors.

A scenario development process is defined. We use the term scenario to refer to the entire context description, including the strategic, operational and tactical levels. Strategic statements give the broader context of actors and reasoning as to the conflict, the operational level relates to the specific mission, and the tactical level refers to potential decision making within the mission. At any stage, there may be limitations on the scenario development: for example verification and validation (V&V) testing may prioritise particular threats, a time frame, the location of the operation, and so on.

Design of Experiments (DOE) approaches are employed to within experimental testing and supporting OR to ensure that the IMD process meet operational needs, dependent on the scenario context. We believe that a similar DOE approach should be employed to V&V the scenarios. This can provide a rigorously verifiable method for determining the extent of scenario coverage but, furthermore, techniques from the DOE literature might be applied to the task of generating scenarios in an effort to trade-off and estimate V&V effort against scenario requirement coverage. This could be combined in future with the experimental design process, but more attractively the development of an independent scenario design method that provides some level of orthogonality within the scenario construction could greatly expand the applicability of all levels of testing, and provide a more comprehensive coverage of the potential scenario space.

Keywords: Scenario design, design of experiments

1.   INTRODUCTION

Modern military aircraft are complex with increasingly integrated sensing functionality. As mission system capability is increasingly implemented in software rather than hardware — particularly with the development of electronically steered antennas and increasing processing power for software implementations of former hardware components — the scope for specific data requirements and software loads that are tailored for missions is increasing. As the complexity of these loads increases, the operational effectiveness of each aircraft is increasingly dependent on each mission data load. Put simply: an aircraft with ten million lines of software, much of which is configurable by input data, has a significant part of its capability defined by that data.

How do we ensure that the aircraft mission systems perform as required in a complex operational environment? Further, how do we provide Operations Research (OR) to support experimental testing results? A task often given to OR personnel is to construct scenarios that provide context for all levels of experimental work, both hardware lab testing, and to support any OR experimentation that builds on these results. Any scenarios developed in this process need to provide both operational realism and sufficient levels of complexity to suit a testing program. This is problematic in that large scenarios that are generated by the selection of "important" features are by their very nature biased towards particular solutions. How do we avoid bias in the scenario generation?

This paper frames the need for complex scenario development by considering a specific problem set: the requirement for Intelligence Mission Data (IMD). Firstly, an overview of IMD is given, followed by a description of the where scenarios are required. An outline of a scenario generation process is given. Challenges and issues with this process are outlined along with a proposed solution.

1.1.   Specific Problem Example: Intelligence Mission Data

An area at the forefront of the need for comprehensive scenarios is the testing of IMD requirements. IMD is information that is collected by intelligence agencies and used to "reprogram" aircraft; the data is tailored or "coloured" to each aircraft type and each mission set (DoD, 2013). IMD includes, but is not limited to, five specific areas. These are:

- characteristics and performance (C&P) covering physical capabilities of systems,
- emissions covering the non-communications electromagnetic emissions that systems such as radar produce — traditionally this area is known as Electronic Warfare Integrated Reprogramming (EWIR),
- signatures covering how entities can be viewed (non-emissions) across a number of spectrums,
- geospatial intelligence (GEOINT) including maps and imagery, and the
- order of battle (ORBAT) describing the entities and equipment that are likely to be within the area.

Knowing what entities are within an environment, where they are, what they look like, what they emit, and how they move is important for effective operations. Limiting IMD to those entities likely to be an area of operation, rather than searching a database of the whole world, improves the likelihood of your own systems finding them. In a nutshell, this is the *reprogramming problem*.

Consider the issues of preparing IMD for modern fighter aircraft: these aircraft will require information from all IMD areas identified above, and will subsequently fuse the information from multiple sensors and networks together into a consistent picture of the operating environment. How do differing mission data loads then impact the operational performance of these aircraft?

1.2.   Previous Work

While there is a range of literature on scenario design, much of this relates to very specific problem sets. This work is no different: IMD presents a complicated case as not are the specific operations important down to the exact physical locations of aircraft and other entities, but the broad context is need to provide the environmental conditions. The majority of scenario decomposition work undertaken in military settings involves the Joint Military Appreciation Process (ADFWC, 1998) of which scenario decomposition is just a small component of the JMAP process, or the application of strategy to task method (Thaler, 1993) which is aimed at decomposing scenarios at the strategic level. In this situation the strategic context is implied and we needing to link that to a range of specific operations that are tactically instantiated. Neither referenced work is strong on developing operational missions within that context. Regardless, outside requirements often guide scenario construction more than a specific method, bringing with those requirements associated biases (Dobson-Keefe et al, 2015).

2. IMD SCENARIO APPLICATIONS

This gets to the crux of one particular issue: how do you decide, with (1) a particular set of mission system hardware in a lab and a (2) reprogrammed mission data file set (MDFS) that provides the IMD in an operational form, that a fighter aircraft will be effective and survivable in a specific operational context? In testing IMD there is a need to consider all levels of the operational context as defined by a scenario:

- the strategic context gives purpose to the scenario and the broader physical and electromagnetic environment via the specification of an area of operation,
- the operational context gives the goals or missions to be pursued, and
- the tactical flow can affect the specific mission performance as this describes the specific interplay of actors.

2.1. Hardware and IMD Verification and Validation

A question the war fighter will have, and thus a question the reprogrammers will have, will this impact my ability to complete a mission, and this is the aim of the verification and validation (V&V) process. The stated operational requirements are used to verify if the system meets those requirements, and to validate that the systems meet the operational needs. In testing the V&V of the hardware, and IMD loads and the processes that produce them, a range of scenarios need to be provided. The operational requirements partially define the detail required to construct a scenario.

Emitter density is one kind of scenario complexity. But what are the emitters in the scenario? Do you pick those based on the most common set of emitters, or most advanced set of emitters to detect? At what point do reprogrammers say "we believe this aircraft is good to fly based on the requirements", and what caveats are included in that statement? How do represent the emitters and modes that we do not know about? Surrogate data — data that may represent a system that little is directly known about but may be inferred from knowledge of other systems — may have to be used due to the limits of the intelligence process.

In terms of scenario design, V&V will have very specific requirements, as this has implications as to whether the aircraft can be accepted as functional. Those requirements are specifically identified and limit the structure of the acceptable scenarios.

2.2. Scientific Hardware and IMD Testing

The aim of a scientific hardware test program is different from V&V: V&V ensures the hardware meets specific outcome metrics, whereas a broader scientific test program is designed to test the limits of system capability. Both are limited in scope by the physical hardware test lab capabilities, but a design of experiments (DOE) approach (Box et al, 2013; Wass, 2010) may be employed to obtain the greatest insight into the systems. A range of potential issues may be examined in relation to IMD availability: inaccuracy of data fields, missing data fields, threat representation by a surrogate system, and previously unobserved threats. (Missing data fields is the least likely situation: if a threat system is known of, but without the requisite detail, then surrogate information would be used to best represent that system.)

Manipulation of that experimental setup, assuming a static set of hardware, then takes two broad forms: variation of the scenario and context, or variation of the IMD through the MDFS. Both approaches can be effective and both are employed. Fixing an MDFS and manipulating the models and environment by adding entities can replicate the issues described above, and is particularly inviting should the MDFS generation and associated data be costly. An issue is whether the scenarios are deemed "operationally valid" or realistic tests. Fixing the scenario and context through the models and environment, and then manipulating the MDFS to omit platforms, maintains the scenario realism at the potential cost of generating a larger number of MDFS. This form of hardware testing as compared to V&V then involves trading off what might be termed operational "validity" for variance in testing, noting of course the limits of "truth" in the data represented. The fundamental requirement is to understand the impact of data variance on system performance.

In terms of scenario requirements, the scenarios can be looser in structure but still limited in scope. Considering two or three scenarios with variations as described for IMD and/or the environment provides a large scope of potential testing, given that usually hardware labs have limited availability.

2.3. Supporting IMD Analysis Program

An analysis, or OR, program supports IMD hardware testing in two ways. Firstly is to provide the operational context to support appreciation of the operational implications of the outcomes of hardware

testing. Hardware testing may provide a range of measures of system performance: and the analysis program would provide an examination of the measures of operational effectiveness with the likely assistance simulation or modelling. In terms of scenario requirements it is likely to mirror the hardware testing needs; as noted above an OR specialist is critical for scenario specifications.

Secondly, an analysis program may extrapolate from models developed by the hardware testing to a range of scenarios that were not previously tested. Like a DOE approach, "data farming" is analogous to supporting the best outcome from a simulation test environment (Sanchez et al, 2015):

> The statistical DOE literature focuses, in large part, on comparing designs in terms of the number of design points or the precision of specific factor effect estimates (e.g., main effects) based on assumed response behavior. This means there is a tendency to limit the investigation to a very small number of factors and/or limit the number of levels for each factor. This mindset is counterproductive for simulation experiments, particularly given the availability of computing clusters and the relative time required to create the model.

Maximising the number of factors that can be examined through a set number of simulations provides an extrapolation that allows scope of more varied testing with greater excursions from a limited set of scenarios.

In this second situation the question is how do we ensure operational validity while controlling the scope of the scenarios? This is a similar question to whether an operationally realistic threat is required or a more complex threat; it may be generalised to whether there are sufficient edge cases with the scenario set. In terms of our EWIR example set, five collinear threats are unrealistic but potentially interesting from a system capacity point of view. It may be possible to develop a range of more general use cases, especially with simply providing sets of entities, however this does not avoid issues of making this operationally relevant: operators are unlikely to see value in providing tactical advice on an operationally implausible test.

3. CURRENT SCENARIO DEVELOPMENT PROCESS

Given that the subject matter relates to military operations, many examples are classified in nature, hence the discussion below relates to the process that is undertaken, and considers an artificial problem. This section aims to provide an understanding of the scope of scenario requirements that are needed to support testing at all levels. The design of experiments issues we encounter in this problem are applicable to other problems where a system must be proposed, assessed, and tested against an unknown (and in some cases unknowable) set of conditions.

We use the term scenario to refer to the entire context description, including the strategic, operational and tactical levels. Strategic statements give the broader context of actors and reasoning as to the conflict, the operational level relates to the specific mission, and the tactical level refers to potential decision making within the mission. At any stage, there may be limitations on the scenario development: for example V&V testing may prioritise particular threats, a time frame, the location of the operation, and so on.

3.1. Strategic Level

The first stage is to identify the strategic setting:

- What roles are being examined? For example, defensive counter air may be the role under consideration.
- Who are the protagonists: who is "blue" (own force), "red" (enemy), "grey" (other military forces) and "white" (civilian) entities? This may include a description of the potential entities that are likely within the area, otherwise known as the ORBAT for military entities.
- How are going to measure operational effectiveness are you wanting to test? Common Measures of Effectiveness (MOEs) for combat include lethality, survivability, concurrency, and combined operations.
- Provide narrative that adds flesh to the scenario. When and where is the scenario happening? The year influences likely equipment that could be employed, and the area of operation (AO) influences factors like weather.

This description is rather abstract. Let us consider a fictitious scenario, which is in effect the required narrative:

> Tensions have risen within Australia and New Zealand in 2025 following the breakdown in societal order that has plagued the west following the decline of the body politic. In both countries internal order has fractured along societal flash points. Tasmania, led by a dictatorial premier, has seceded from Australia using the lack of the establishment of a Tasmanian-based AFL team as a point of contention. Similarly in New Zealand, Maori tribal differences and perceived biased selections for

the All Blacks have splintered the country between the north and south islands. Both Tasmania and Te Waipounamu (the south island of the former New Zealand) have been courted by a third country, with the aim of further driving discord within Western societies. That country has deployed a small force of fighters and SAM systems to Hobart and Christchurch to support the local militia. A carrier force is rumoured to be deploying to the Tasman Sea.

The overarching requirement is for Australia to re-assert sovereignty over Tasmania, with the first mission being an operation to neutralise any deployed SAM systems. The blue force is Australian air assets deployed to Melbourne, and Australian naval assets (two Air Warfare Destroyers). Grey includes former New Zealand North Island shipping including naval vessels and maritime patrol. Red assets include six fighters and two high-end SAM systems. White entities include minimal shipping (fishing and freight), and commercial flights between the Te Waipounamu and Tasmania.

This is included simply to give an indication of the components that a narrative could contain. This particular narrative is patently ludicrous — except for the strength of feeling regarding local sporting teams — but sets the scene for scenario development, both strategically and operationally. The narrative has essential components that influence the cohesion of a scenario, with the time frame and facts providing the AO and some of the eventual entity set (including military ORBAT). The final entity set may include additional fleet elements and specific platforms that may need identifying.

3.2. Operational Level

Next we develop the specific operational setting:

- Define the specific mission requirements, particularly for combat missions, which may involve and rules of engagement (ROE). Also specify the missions and operation of third parties.
- Identify the specific ORBATs and entities involved.
- What sub-MOEs are needed: such as mission phases that need to be completed (JCS, 2013).
- Provide any additional narrative.

The operational and strategic level information could overlap depending on the size of the scenario. An operation, though, has a short action timeframe, and a strategic description could involve several operations. For the example described above, specification of the operations is left as an exercise for the reader.

3.3. Tactical Level

Next we develop the tactical settings and an example tactical flow for all sides in the operation. This includes identifying specific environmental conditions such as weather, white entities, and grey ORBAT actions. For combat operations this needs to be identified for blue and for red:

- The initial states: specific objective, risks, specific ORBAT, basing, weapon load-outs, specific AO, specific ROE, support assets, communication methods.
- The tactical flow requirements: initiation, ingress, presentation of assets, route and fuel planning, game plan and specific tactics, assumptions about enemy.
- Identify where knowledge and understanding about the opposition is wrong.
- Develop an example of the tactical flow interactions.

At the strategic and operational level we include MOEs to measure the operational process or performance. At the tactical level when we are considering a specific set of interactions, the measures can be instantiated: this is where the process and interaction occur. In OR simulations, the tactical flow would be an instance of the simulations emergent behaviour. For hardware testing, one or several examples of this flow need to be described. The development of a specific tactical flow is essential, at a minimum as a check of consistency.

This scenario development needs to be completed for each specific tactical scenario. The complexity of each scenario is mitigated if they relate to each other, but the process can be elaborate depending on the requirement. In the case of IMD testing the process is elaborate: the strategic level defines the potential systems within the environment (ORBAT), even if they are not within the operation specification.

4. DEVELOPING MULTIPLE SCENARIOS

A range of scenarios with a range of complexity are required, as the range of operational missions varies in complexity. While maintaining realism, low complexity scenarios allow for more isolated testing whereas

high complexity scenarios are needed to stress the aircraft's capabilities. Obviously in the fighter aircraft space, the range of operation types are well known, and the requirements could be split into, for example:

1. Defensive Counter Air (DCA)
2. Close Air Support (CAS)
3. Offensive Counter Air (OCA)
4. Maritime Strike (MS)

For each operation type a range of specific operations can be developed with increasing complexity. If the simpler ones are subsets of the more complicated operations then this may reduce the development overhead. The prosecution of each operation is then worked with input from subject matter experts (SMEs) to give each consistent scenario.

This utilizes the operational level as the point of branching to multiple scenarios, as this allows us to test the inclusion of different scenarios without modifying the broader strategic description. Arguably we could have chosen multiple strategic descriptions applying a single operational mission in each, however there is the overhead of generating enough relevant strategic descriptions. Providing multiple variations through different tactical level interactions is simpler to construct, but that does not provide us with the required range of IMD variation.

The major issue with this process is that decisions on the scenario construction can have major influences on the subsequent scenario development and complexity. This is the fundamental issue if a scenario or scenarios were to be used to determine acquisition decisions that influence the force structure. In the case of IMD studies the intent would be to provide priorities to influence the existing IMD enterprise. The second issue is that the development of scenarios is ad hoc and potentially biased by developers and maybe overly constrained by the requirements, and yet testing against every possible scenario is infeasible.

A side issue that must be considered in summarising this complexity of scenario development is to note the large number of stakeholders involved. Given the broad scope of the scenarios, this work impacts combat pilots, system maintainers, logistics personnel, the intelligence organisations, and strategic policy wonks: almost everyone cares in some way. While approaches such as the Delphi Method (Dalkey et al, 1963) provide a structured approach to eliciting expert opinion, the need for an articulated scenario to minimise confusion cannot be underestimated. By combining expert judgement – however that is obtained – with a strong analytical base, a negotiation between stakeholders can build towards a reasonable consensus view. The risk is that too many cooks can make a dank broth.

4.1. DOE Approaches to Scenario Development

DOE approaches are employed to within experimental testing and supporting OR to ensure that the IMD process meet operational needs, dependent on the scenario context. We believe that a similar DOE approach should be employed to V&V the scenarios. This can provide a rigorously verifiable method for determining the extent of scenario coverage but, furthermore, techniques from the DOE literature might be applied to the task of generating scenarios in an effort to trade-off and estimate V&V effort against scenario requirement coverage. This could be combined in future with the experimental design process, but more attractively the development of an independent scenario design method that provides a more comprehensive coverage of the potential scenario space within the scenario construction could greatly expand the applicability of all levels of testing.

Let us consider how this may develop:

1. The requirements of the scenario become the independent factors.
2. A set of rules will define scenario consistency. For example, the distances travelled by aircraft in the scenario may require a tanker to be included within the ORBAT. These interactions ensure that the scenario is consistent.
3. An orthogonal design sets the levels of input factors, and hence the conditions of each scenario. These scenarios may not result in combat outcomes, and include the shape and deter phases prior to combat.
4. The response may be a measure of the scenario complexity, for example the total ORBAT numbers in the specific scenario.
5. Guided by the design, specific scenario laydowns may be developed.

By undertaking a similar process, a DOE design may guide us towards scenarios that are relevant, and give confidence that the scenario set provides coverage of the space of potential outcomes.

5.   CONCLUSION AND FURTHER WORK

This paper has described a complex problem space in undertaking testing of hardware systems whose performance is dependent on completely separate data inputs. It considers the development of scenarios for testing such systems, in three different environments: directed hardware testing, exploratory hardware testing, and OR simulation to support the hardware testing. A scenario development process is outlined. The issues of multiple scenario development are then discussed, and a proposal for the application of DOE and data mining methods to independent scenario development is proposed to ensure that the scenario set covers a full range of complicating factors. This paper does not attempt to solve the issues in the scenario development process, but to highlight the complexity of the problem space.

Further work includes the application of methods to develop a set of scenarios, both for a toy example and for a more involved example. Consideration could then be given to expanding the problem space to future systems to be handled by IMD. Rapidly adjusting systems will require rapid adjustment of mission data. Future scenarios will need to be designed to include such systems within their consideration.

REFERENCES

Box, G.E.P., W.G. Hunter, and J.S. Hunter (2005). *Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building*. 2nd ed. New York: Wiley.

Australian Defence Force Warfare Centre (1998). *Joint military appreciation process: a guide to planning at the operational level*.

Dalkey, N.; Helmer, O. (1963). "An Experimental Application of the Delphi Method to the use of experts". *Management Science*. 9 (3): 458–467.

Dobson-Keefe, N., and W. Coaker (2015). "Thinking More Rationally: cognitive biases and the Joint Military Appreciation Process", Australian Defence Force Journal, No. 197.

DoD (2013). Management of Intelligence Mission Data (IMD) in DoD Acquisition, Department of Defense Directive, Number 5250.01, January 22, 2013.

Joint Chiefs of Staff (2013). Joint targeting, *Joint Publication* 3-60, 31 January 2013.

Sanchez, S.M. and H. Wan (2015). Work smarter, not harder: A tutorial on designing and conducting simulation experiments. *Proceedings of the 2015 Winter Simulation Conference*, 1795-1809.

Thaler, DE. (1993). *Strategies to Tasks: A Framework for Linking Means and Ends*, RAND report MR-330. ISBN 0-8330-1461-7.

Wass, J.A. (2010). First Steps in Experimental Design—The Screening Experiment. *The Journal of Validation Technology*, Spring 2010.