Applying Design of Experiments (DOE) methodology to Sortie Generation Rate (SGR) Evaluation

Josh Tribble
MILITARY ANALYST
AVW TECHNOLOGIES
Agenda

- Introduction
  - Acquisition humor
  - The Integrated T&E Challenge
- Intro to Design of Experiments
- SGR Assessment Methodology
  - Overview of SGR Assessment to date
  - SGR Assessment objectives, MOEs, factors
  - SGR Testbed Assessment Design Factors / Run Matrix
  - SGR Live Testing Validation
- Benefits of DOE over single scenario based analysis
- Conclusion / Q&A

NOTE: My remarks are intended to spur thought on improving how we as testers can do business better to support the warfighter. While I hope this aligns well with DoD and Services T&E initiatives, I am not representing any government agencies’ official position.
Acquisition 101?

How the user described it
How the requirement was understood
How the contractor designed it
How the programmer wrote it
How the PM/sponsor described it

How the project was documented
What was actually installed
How the Government was billed
How the helpdesk supported it
What the user really needed

How do we avoid this?
Integrated T&E Challenge

- Coordinated planning and development of individual test objectives
- DT / CT / OT / LFT&E remain separate but leverage data and resources whenever possible
- Potential for significant cost savings and earlier risk reduction
- Requires buy-in from all orgs + strong T&E Working IPT
- Requires strong, up-front, test planning and data analysis methodology – *Design of Experiments (DOE!)*

\[
T&E_{\text{integrated}} = \int f(CT, DT, OT, LFT&E, Joint Exp, M&S, Analysis, etc.) \, dt
\]

Program Conception
Intro to DOE
Background of DOE

• DOE originated in the field of agricultural studies in the 1930s by R. Fisher, building on W.T. Gossett’s work at Guinness Brewery—Brilliant!

• Used throughout industry in industrial experiments, process improvement, statistical process control

• USAF has significant experience in use of DOE across numerous programs; Navy is beginning to implement

• DOE methodology is used to interrogate a process, improve knowledge of how the process works, and identify factors and interactions affecting variability of performance outcomes.
DOE Process Goal / Benefits

• Compared to other systematic methods DOE designs:

  • Yield better process understanding
  • Can be planned and analyzed faster
  • Cheaper – using between 20-80% of usual runs/tests/resources
  • Better exploration across range of performance—depth and breadth of testing
  • Challenge assumptions and demonstrate real performance
  • Better way to design and test complex systems
**DOE Process Outline—4 Basic Steps**

- **Project description and decomposition**
  - Problem statement and objective of experiment (test)
  - Response variables, and potential causal variables – Ishikawa fish bone.

- **Plan test matrix**
  - Determine constraints, prioritize factors, and select statistical design ($2^K$ vs. $3^K$ vs. mixed, Taguchi vs. classical arrays, full vs. fractional, non-linear effects?, replications?, blocking?)
  - Write the test plan with sample matrices, profiles, and sample output; run sample analysis.

- **Produce observations**—random run order & blocked against unknown effects
  - Block runs to guard against uncontrollable unknown effects as needed.

- **Ponder the results**
  - Analyze and project data; draw conclusions, redesign test as necessary and assess results.
  - Perform “salvo testing” (test-analyze-test); screen large # of factors then model
SGR Assessment Methodology
## SGR Assessment Requirements

### SGR Key Performance Parameter

<table>
<thead>
<tr>
<th></th>
<th><strong>THRESHOLD</strong></th>
<th><strong>OBJECTIVE</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sustained SGR</strong></td>
<td>Average of 160 operational combat equivalent aircraft sorties in 12 hours of launching per day over 30 days (26 Flying and 4 Non-Flying Days as specified in the Design Reference Mission (DRM) – total cycle of 4160.</td>
<td>Average of 220 operational combat equivalent aircraft sorties with 12 hours of launching per day sustained over 30 days (26 Flying and 4 Non-Flying Days as specified in the DRM) – total cycle of 5720.</td>
</tr>
<tr>
<td><strong>Surge SGR</strong></td>
<td>Average of 270 operational combat equivalent aircraft sorties generated during each successive 24-hour period over 4 continuous days.</td>
<td>Surge: average of 310 operational combat equivalent aircraft sorties generated during each successive 24-hour period over 4 continuous days.</td>
</tr>
</tbody>
</table>

### Other Measures of Performance: cycle times, task timing, launch and recovery cycles, resource usage, crew fatigue levels, fuel states/rates, etc.
SGR Assessment Testbed

- M&S testbed captures times and actions associated with preparing, launching, and recovering sorties per the DRM

- M&S matured and validated over time prior to runs for score
- Live test used for validation once ship is delivered and aviation certified
SGR is a function of
- Launch Cycle/Interval Timing
- Recovery Times/Intervals
- Mission Planning Timing
- Aircraft Recovery Time Which Encompasses:
  - Fueling Time
  - Ordnance Handling Times
  - Aircraft Movement/Spotting Times On The Flight Deck
  - Aircraft Movement/Spotting Times In The Hangar Bay
  - Aircraft Availability
SGR Assessment Analysis

Objectives

- Determine average SGR over DRM to meet KPP requirement
- Determine active factors influencing the variability & overall outcome
  - Measure % sorties completion rather than binomial pass/fail
  - Each day in the DRM treated as a single design point due to interdependencies of events within that day
- Provide the fleet with an analytical model showing probability of meeting a given airplan based on its size, mission composition, environment, and any other active factors

\[
\% \text{Airplan _ Sorties _ Completed} = \frac{\text{Daily _ sorties _ completed _ successfully}}{\text{Airplan _ sorties}} \times 100\% 
\]

- Allows equal comparison of the 4 T/O surge/sustained requirements across all factors
- Continuous dependent variable provides more statistical power than pass/fail
- Supports more robust assessment of capes and lims
**SGR Factor Selection**

*Experimental control factors:*

- **Environmental**
  - Sea/Winds: state 1 vs. 3
  - Visibility/Sky Cover: Clear Skies (Case I) or Cloudy/Night (Case III)
  - Time of day: midday or midnight (for 12 hour ops, N/A for 24 hour ops)

- **Systems:**
  - Availability: 100% & actual (for CVN-21 systems and aircraft)—allows for analysis of impact of equipment failures

- **Mission**
  - Sortie Size: Threshold & Objective levels from the DRM
  - Sustained and Surge Mission (12 vs. 24 hr ops (with augmented crew))
  - Operation day: early and late in ship on-station operational period; expect to interact with availability for system failures and also translates to possible crew fatigue
  - Airplan mission mix: early/late DRM days representing different ordnance mix;
  - Mission mix and operation day
**SGR Factor Selection (cont’)**

**Controllable Factors held constant:**
- Underway Replenishment
  - Not a factor of SGR but presumed to occur on assigned days or fuel and ordnance will not be available for the planned flight days
- Aircrew augmentation
  - Confounded with mission type – assumed normal crew for sustained operations and augmented crew for surge missions

**Measurable Noise Factors**
- Other environmental factors not controlled (if in test / model)
  - Temperature extremes
- Specific metrics in the subordinate models driven by the main inputs, such as:
  - Crew fatigue (driven by the mission day)
  - Resource availability
  - Number of aircraft available
  - Weapon skids available
  - Timing for critical tasks, etc.
SGR Factor Selection (cont’)

- **Design factors:**
  - Factors with highest expected influence listed first
    - Important when setting up fractional factorial matrices—usually easier to resolve factors and interactions
  - Setup for M&S only; cannot test all of these in live testing
  - Requires M&S improvements
  - Need buy-in for “excursions” above threshold
    - High levels force the “system” towards a higher failure rate to see more variation in response

<table>
<thead>
<tr>
<th>Setting Factor</th>
<th>(Low) -1</th>
<th>(Center Point) 0</th>
<th>(High) +1</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Surge/ Sustained Operations</td>
<td>Sustained (12 Hr ops)</td>
<td>N/A</td>
</tr>
<tr>
<td>B</td>
<td>Sortie Size (T/O)</td>
<td>Threshold</td>
<td>Halfway btwn</td>
</tr>
<tr>
<td>C</td>
<td>operational day</td>
<td>Early (1/4 or 5/30)</td>
<td>Mid (2/4 or 15/30)</td>
</tr>
<tr>
<td>D</td>
<td>Availability</td>
<td>100%</td>
<td>Halfway btwn</td>
</tr>
<tr>
<td>E</td>
<td>Visibility/ Cloud Cover:</td>
<td>Clear/ Case I</td>
<td>Partly Cloudy/ Case II?</td>
</tr>
<tr>
<td>F</td>
<td>Seakeeping motion effects</td>
<td>5 kts/SS1</td>
<td>12 kts/SS2</td>
</tr>
<tr>
<td>G</td>
<td>Time of day</td>
<td>Day</td>
<td>Dusk?</td>
</tr>
<tr>
<td>H</td>
<td>Mission Day</td>
<td>Early</td>
<td>Mid</td>
</tr>
</tbody>
</table>
SGR Testbed Run
Assessment Design

- Full factorial requires $2^8$ or 256 runs
  - *Unnecessary since many effects are inactive*
- Resulting test matrix is a resolution IV $2^{8-4}$ fractional factorial of 16 runs + 8 additional runs for central composite design
  - *Some interactions are confounded but can be resolved*
- Model DRM days per the assigned settings and evaluate SGR Compl %
- “salvo test”:
  - Runs 1-8, then analyze for effects
  - Runs 9-16, then reanalyze for effects
  - Perform center points to check for linearity
  - If necessary, run CCD (face points) for non-linear effects
SGR Live Testing Validation
Test Design

- Live test conditions and cost (potentially $100M?) limit amount of live test and the conditions
- Focus on validating specific test points of interest and confirm within the M&S runs for score

<table>
<thead>
<tr>
<th>Factor</th>
<th>Factor</th>
<th>-1</th>
<th>0</th>
<th>+1</th>
<th>Rationale</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Surge/ Sust. Ops</td>
<td>Surge</td>
<td>Sustained</td>
<td>N/A</td>
<td>Surge</td>
<td>Both operations can be run</td>
</tr>
<tr>
<td>B Sortie Size (T/O)</td>
<td>Threshold</td>
<td>(T+ O)/ 2 Objective</td>
<td>A mix of sortie sizes can be run</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C Operational day</td>
<td>Early</td>
<td>Mid</td>
<td>Late</td>
<td></td>
<td>No means of imposing a late day due to cost</td>
</tr>
<tr>
<td>D CVN-21/A/C Ao</td>
<td>100%</td>
<td>Halfway</td>
<td>Actual</td>
<td>Actual equipment Ao</td>
<td></td>
</tr>
<tr>
<td>E Cloud Cover</td>
<td>Actual conditions?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F Sea-State</td>
<td>Actual conditions?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G Time of day</td>
<td>Actual conditions?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H DRM Mission mix</td>
<td>Early</td>
<td>Mid</td>
<td>Late</td>
<td></td>
<td>Factor is probably inactive so randomly assign</td>
</tr>
</tbody>
</table>


SGR Live Testing Validation
Test Design (cont’)

• Final Test Matrix with settings:

<table>
<thead>
<tr>
<th>Test Case</th>
<th>A: Ops Type</th>
<th>B: Sortie Level</th>
<th>Actual (# Sorties)</th>
<th>H: DRM Mission Day</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Sustained</td>
<td>Threshold</td>
<td>160</td>
<td>5</td>
<td>Priority</td>
</tr>
<tr>
<td>2</td>
<td>Sustained</td>
<td>Objective</td>
<td>220</td>
<td>26</td>
<td>Priority</td>
</tr>
<tr>
<td>3</td>
<td>Surge</td>
<td>Threshold</td>
<td>270</td>
<td>26</td>
<td>Priority</td>
</tr>
<tr>
<td>4</td>
<td>Surge</td>
<td>Objective</td>
<td>310</td>
<td>5</td>
<td>Priority</td>
</tr>
<tr>
<td>5</td>
<td>Sustained</td>
<td>Halfway btwn</td>
<td>190</td>
<td>15</td>
<td>Additional run for midpoint</td>
</tr>
<tr>
<td>6</td>
<td>Surge</td>
<td>Halfway btwn</td>
<td>290</td>
<td>15</td>
<td>Additional run for midpoint</td>
</tr>
<tr>
<td>7</td>
<td>Sustained</td>
<td>Threshold</td>
<td>160</td>
<td>26</td>
<td>Additional run for alternate mission mix</td>
</tr>
<tr>
<td>8</td>
<td>Sustained</td>
<td>Objective</td>
<td>220</td>
<td>5</td>
<td>Additional run for alternate mission mix</td>
</tr>
</tbody>
</table>

• Recommend run during Joint Task Force Exercise to ensure combat ready crew & systems

• Some analysis of variance can be run directly but main objective is to compare day for day with M&S results (including V&V of lower level measures within the specific process models)

• Runs 1-4 are priority; select additional runs based on M&S results
SGR Testbed Assessment
Sample Data Analysis

• Response surface plot across factors of interest showing response & interactions

• Table of plan vs. predicted actual SGR Completion Rate for factor settings of interest -- shows SGR completion % falling off as too many are sequenced

• demonstrates how analysis can describe ship caps & lims, not just a pass/fail grade for a KPP tested only to threshold
Benefits of DOE
CONCLUSION

• DOE methodology:
  – may significantly **reduce the required runs** for Testbed Assessment and live test validation while...
  – providing a **more robust process** for statistical analysis of variance to determine where the ship design can and cannot support a given air-plan under the other conditions
  – supports robust & efficient **integration of M&S development, testing, VV&A, & evaluation**

• DOE is:
  – a smarter way of doing testing
  – can provides superior knowledge to the systems engineers
  – something all testers & systems engineers should become familiar with!

• QUESTIONS?