Design of Experiments in Highly Constrained Design Spaces

2011 Army Conference on Applied Statistics

Laura Freeman
Research Staff Member
lfreeman@ida.org

Approved for public release; distribution is unlimited.

IDA

Outline

- **Design of Experiments (DOE) in Operational Test and Evaluation**
  - Policy & guidance
  - Challenges
- **Two representative case studies**
  - Jammer
  - Chemical agent detector
- **DOE selection criteria**
  - Statistical & other
- **Summary**

Guidance

OFFICE OF THE SECRETARY OF DEFENSE
1700 DEFENSE PENTAGON
WASHINGTON, DC 20301-1700

OPERATIONAL TEST AND EVALUATION

OCT 1 9 2010

MEMORANDUM FOR COMMANDER, ARMY TEST AND EVALUATION COMMAND
COMMANDER, OPERATIONAL TEST AND EVALUATION FORCE
COMMANDER, AIR FORCE OPERATIONAL TEST AND EVALUATION CENTER
DIRECTOR, MARINE CORPS OPERATIONAL TEST AND EVALUATION ACTIVITY
COMMANDER, JOINT INTEROPERABILITY TEST COMMAND
DEPUTY UNDER SECRETARY OF THE ARMY, TEST & EVALUATION COMMAND
DEPUTY, DEPARTMENT OF THE NAVY TEST & EVALUATION EXECUTIVE
DIRECTOR, TEST & EVALUATION, HEADQUARTERS, U.S. AIR FORCE
TEST AND EVALUATION EXECUTIVE, DEFENSE INFORMATION SYSTEMS AGENCY
DOT&E STAFF

SUBJECT: Guidance on the use of Design of Experiments (DOE) in Operational Test and Evaluation

This memorandum provides further guidance on my initiative to increase the use of scientific and statistical methods in developing rigorous, defensible test plans and in evaluating their results. As I review Test and Evaluation Master Plans (TEMPs) and Test Plans, I am looking for specific information. In general, I am looking for substance vice a 'cookbook' or template approach - each program is unique and will require thoughtful tradeoffs in how this guidance is applied.

A "designed" experiment is a test or test program, planned specifically to determine the effect of a factor or several factors (also called independent variables) on one or more measured responses (also called dependent variables). The purpose is to ensure that the right type of data and enough of it are available to answer the questions of interest. Those questions, and the associated factors and levels, should be determined by subject matter experts -- including both operators and engineers -- at the outset of test planning.

early as possible, but each TEMP revision can be tailored as more information becomes available. That content can either be explicitly made part of TEMPs and Test Plans, or referenced in those documents and provided separately to DOT&E for review.

J. Michael Gilmore
Director

cc:
DDT&E

2

- The goal of the experiment. This should reflect evaluation of end-to-end mission effectiveness in an operationally realistic environment.
- Quantitative mission-oriented response variables for effectiveness and suitability. (These could be Key Performance Parameters but most likely there will be others.)
- Factors that affect those measures of effectiveness and suitability. Systematically, in a rigorous and structured way, develop a test plan that provides good breadth of coverage of those factors across the applicable levels of the factors, taking into account known information in order to concentrate on the factors of most interest.
- A method for strategically varying factors across both developmental and operational testing with respect to responses of interest.
- Statistical measures of merit (power and confidence) on the relevant response variables for which it makes sense. These statistical measures are important to understand "how much testing is enough?" and can be evaluated by decision makers on a quantitative basis so they can trade off test resources for desired confidence in results.

Operational Testing and DOE Principles

- **Traditional Principles of Design of Experiments**
  - Randomization
    - » Are you kidding?
  - Replication
    - » Costs are often prohibitive
  - Local Control of Error
    - » Operational Tests intentionally introduce variability (i.e. the human factor) into the testing process!

- **How do we preserve the benefits of DOE in the operational test environment?**

Steps for Designing an Experiment

1. **Define the problem**
2. **Select appropriate response variables**
3. **Choose factors, levels**
4. **Choose experimental design**
5. **Perform the experiment**
6. **Statistically analyze the data**
7. **Draw conclusions**

Two additional steps are needed:
3a. Identify constraints
3b. Identify statistical model for analysis

Jammer Case Study

- **Goals of the Test**
  - Characterize performance of a new jammer
  - New jammer is required to be a measurable improvement over the legacy jammer
  - Screen factors for future testing
- **Response variables**
  - Miss distance of simulated missile shots
- **Factors and levels**
  - Aircraft variant: 2 variants (A1, A2)
  - Threat: 4 different type of threats (T1, T2, T3, T4)
  - Jammer type: legacy and new
  - Counter Measures: dry, wet Non-maneuvering, or wet with one of three maneuvers
  - Number of sorties per mission: 1 ship or 2 ship

Approved for public release; distribution is unlimited.

Jammer Case Study: DOE Solution

- **DOE Challenges**
  - Complete randomization is not possible
    - » Each mission allowed for up to 8 potential engagements but aircraft and threat could not be easily varied from run to run
  - Disallowed combinations of factors
    - » The legacy system can only be used on one type of aircraft
    - » The legacy system will only be flown in a subset of the operational envelope
      - Dry and wet non-maneuvering
      - Single ship missions
    - » The second aircraft variant can only do a subset of the three maneuvers
  - Limited sample size
    - » 11 operational sorties

- **DOE Solution**
  - D-optimal Split-Plot Design
    - » Allows for restrictions in randomization
  - Creation of new “factors”
    - » Combine original factors into allowed cases for design generation
    - » Accounts for disallowed combinations of factors

IDA

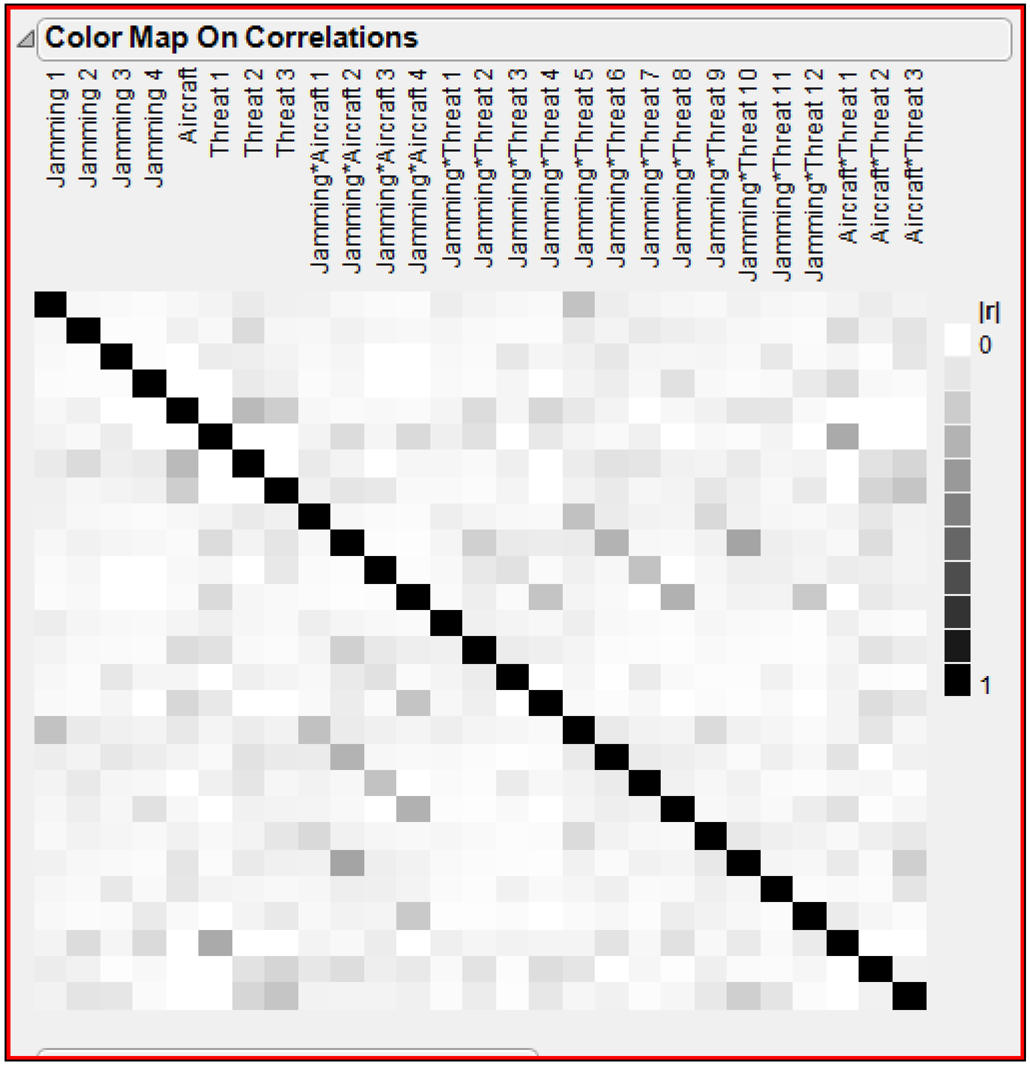
Jammer Case Study: Run Table

- **Design approach (a.k.a tricks of the trade)**
  - Use a generation variable to appropriately weight runs and eliminate some disallowed combinations
  - Practice counting to make sure the right number of whole-plots and sub-plots are selected
  - Customize design & import into software to check properties

Whole Plot	Sub-Plot	Generation Variable	Threat	Jamming	Aircraft	Variant	Number Aircraft
1	1	New A2 1 ship	T4	Wet NM	A2	New	1
1	1	New A2 1 ship	T4	Dry	A2	New	1
1	1	New A2 1 ship	T4	Wet M2	A2	New	1
1	1	New A2 1 ship	T4	Wet M2	A2	New	1
1	2	New A2 1 ship	T1	Wet NM	A2	New	1
1	2	New A2 1 ship	T1	Wet NM	A2	New	1
1	2	New A2 1 ship	T1	Wet M2	A2	New	1
1	2	New A2 1 ship	T1	Dry	A2	New	1
2	3	New A1 1 ship	T4	Wet M2	A1	New	1
2	3	New A1 1 ship	T4	Wet M1	A1	New	1
2	3	New A1 1 ship	T4	Wet NM	A1	New	1
2	3	New A1 1 ship	T4	Dry	A1	New	1
2	4	New A1 1 ship	T2	Wet M2	A1	New	1
2	4	New A1 1 ship	T2	Dry	A1	New	1
2	4	New A1 1 ship	T2	Wet NM	A1	New	1
2	4	New A1 1 ship	T2	Wet M1	A1	New	1
3	5	Legacy A1	T4	Dry	A1	Legacy	1
3	5	Legacy A1	T4	Wet NM	A1	Legacy	1
3	5	Legacy A1	T4	Wet NM	A1	Legacy	1
3	5	Legacy A1	T4	Dry	A1	Legacy	1

Jammer Case Study: Design Properties

Power Numbers		
Factor	S:N = 1	S:N = 2
Aircraft	0.258	0.745
Variant	0.258	0.745
Jamming	0.975	0.999
Threat	0.388	0.844
Wingman	0.258	0.745



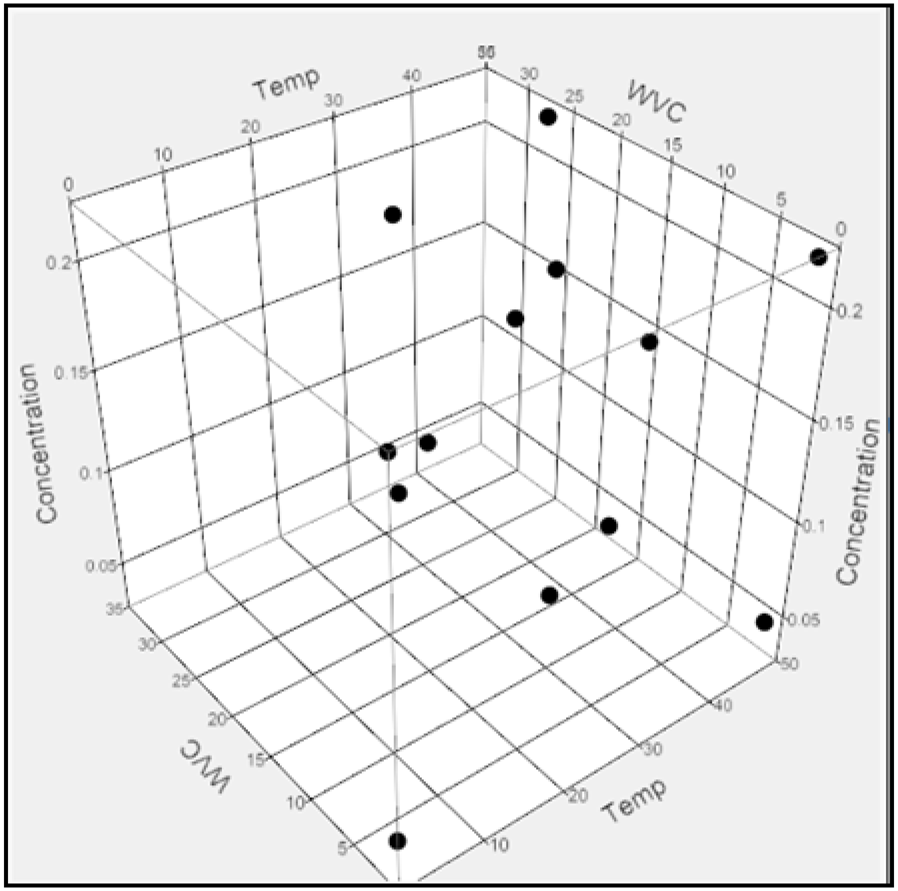
IDA

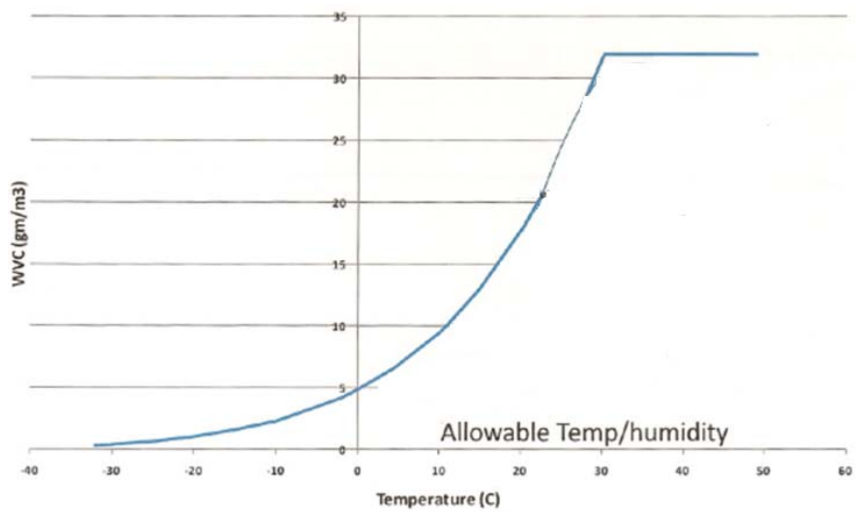
Chemical Agent Detector

- **Goals of the Test**
  - Characterize probability of detection for a range of chemical agents
  - Characterize performance as a function of time
- **Response variables**
  - Time until detection
  - Censored data for non-detects
- **Factors and levels**
  - Agent (9 agents and 2 simulants)
  - Temperature
  - Water vapor concentration
  - Agent concentration
  - Operation mode (monitor or survey)
- **DOE Challenges**
  - Constrained design space

Approved for public release; distribution is unlimited.

Detector: Constrained Design Space



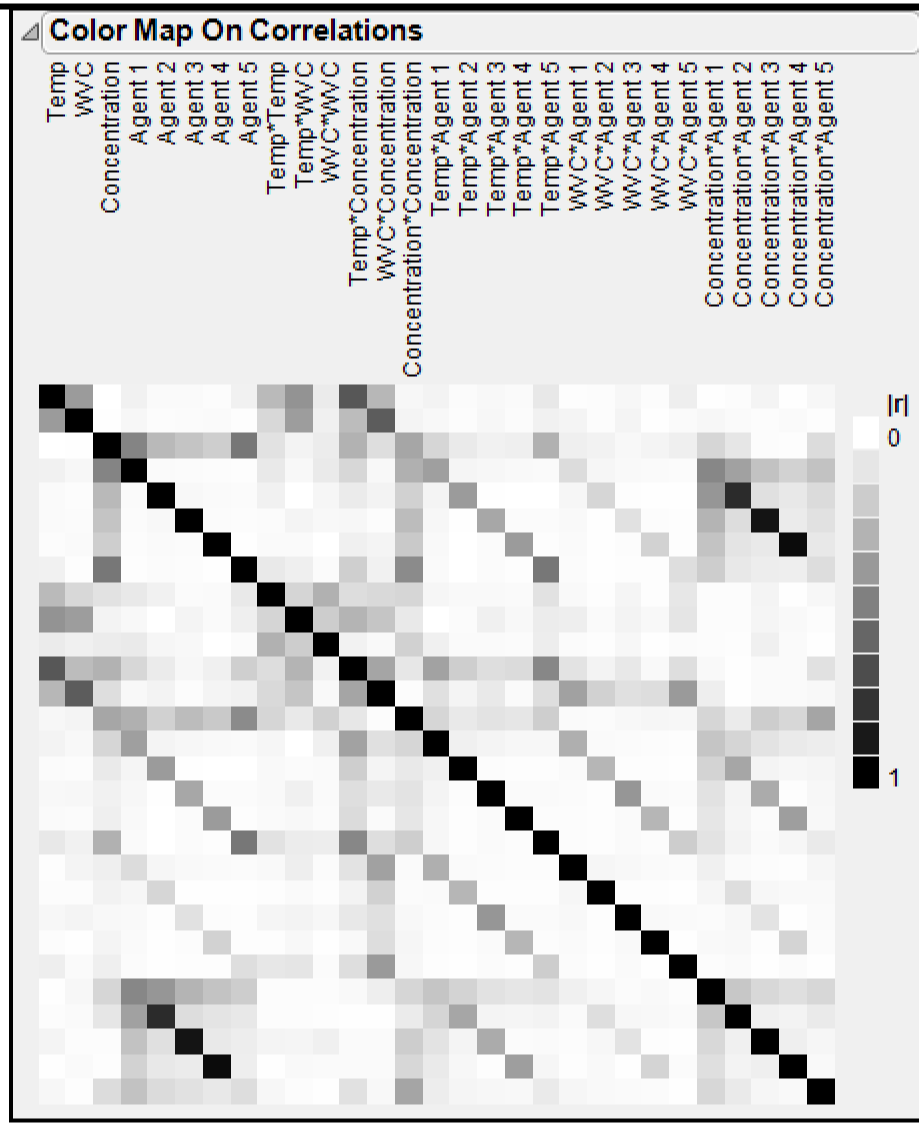


Detector: DOE Solution

- **DOE Solution**
  - I-optimal space filling design
    - Fills the feasible design space
  - Supports response surface modeling
- **Design approach (a.k.a tricks of the trade)**
  - Generate a larger I-optimal design candidate set of points
  - Eliminate and/or scale infeasible points
  - Select final design from candidate set using I-optimality criteria & check design properties
- **Design supports a 2nd order response surface model**
  - Limited correlation between the factors

Design Evaluation

Power Numbers	
Factor	S:N = 1
Temperature	1
Water Vapor Content (WVC)	1
Concentration	0.731
Agent	0.46
Temperature^2	1
WVC^2	1
Concentration^2	1



DOE Selection Criteria

- **Statistical**
  - Correlation between factors is near zero
  - Maximize the number of estimable two factor interactions and other higher order terms (depending on the goal of the test)
  - Minimize correlation between two-factor interactions and main effects
  - Robustness to missing data
  - Adequate statistical power for important effects

- **Other**
  - How easily can the design be modified if things fail to go as planned?
  - Is run order executable?
  - Are there buffer points built into the design?

Conclusions

- **Traditional design strategies often result in designs that are not executable in operational testing.**
- **Advanced techniques in design can overcome many of these limitations**
  - Split-plot designs
  - Optimal design selection
  - Optimal point selection from a candidate set
- **A little creativeness is needed as well**