Design of Experiments

"Managing Expectations"

James "JD" Carpenter
And
Chris Hauser

AVW Technologies, INC
www.avwtech.com

Agenda

"View from the trenches"

- Why test, Why learn?
- Why DOE makes sense
- Manage Expectations - What works (for us)
- Questions?

AVW TECHNOLOGIES

Why Test?

- Why Test?
    - To learn and bound capabilities
    - To answer some basic questions
        - -Does system meet capability requirements?
        - What is actual system performance?
        - -How is system best employed? (Tactics, Techniques and Procedures)

AVW TECHNOLOGIES

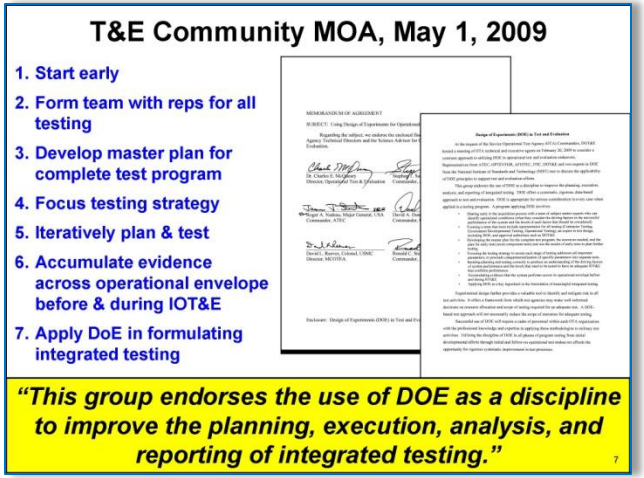
Why Learn?

- Why learn?
  - To discover the “truth” as best we can know it
  - To enable knowledgeable program decisions

Guidance

- Mandated use in Gov't T&E
  - DOT&E requires DOE in Operational Testing
  - Recent DDT&E guidance on Developmental Testing
  - Service OTAs have Joint MOA naming DOE as a best practice

DOT&E rejected TEMPS based on inadequate DOE



We don't need more guidance. We need incentives for PMs/Developers

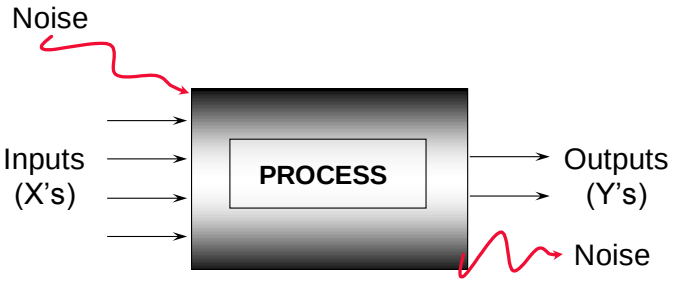
Why DOE?

Scientific Answers to Four Fundamental Test Challenges

Four Challenges faced by any test

1. *How Many?* A: Sufficient samples to control our twin errors – false positives & negatives
2. *Which Points and What's Good?* A: Span the battle-space with orthogonal run matrices using continuous measures tied to the test objectives
3. *How to Execute?* A: Randomize and block runs to exclude effects of the lurking, uncontrollable nuisance variation
4. *What Conclusions?* A: Build math-models* of input/output relations (transfer function), quantifying noise, controlling error

Design of Experiments effectively addresses all these challenges!



* Many model choices: regression, ANOVA, etc.

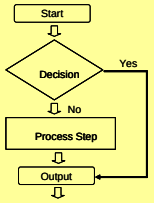
Tester's Challenge

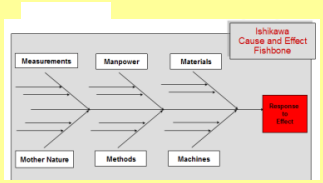
-Time to execute the test
- Resources to support the full scope of planned test
- Funding

The best test may go unfunded while the "worst" test gets funding support

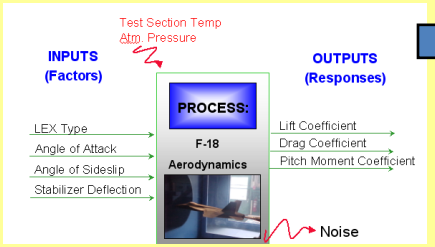
DOE Test Process:
Well-Defined From Blank Paper to Conclusions

Planning: Factors Desirable and Nuisance

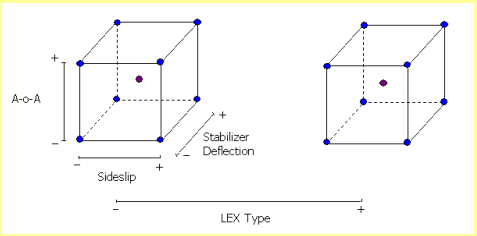




Desired Factors and Responses



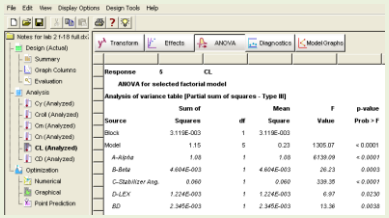
Design Points



Test Matrix

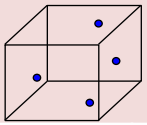
A-o-A	Sideslip	Stabilizer	LEX Type
2	0	5	-1
10	0	-5	1
10	8	5	-1
2	8	5	-1
2	8	-5	-1
2	0	-5	-1
10	8	-5	1
2	0	5	1
2	8	5	1
10	8	5	1
10	8	-5	-1
10	0	5	-1
10	0	-5	-1
2	8	-5	1
10	0	5	1
2	0	-5	1

Analysis and Model



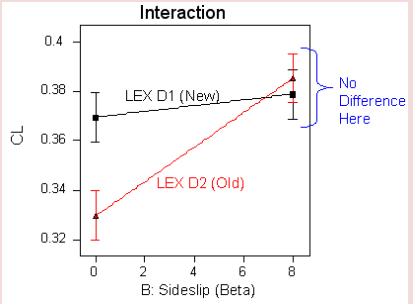
$$C_L = +0.38 + 0.26 \times \text{A-o-A} + 0.017 \times \text{Sideslip} + 0.061 \times \text{Stabilizer Deflection} - .00875 \times \text{LEX Type} + 0.012 \times \text{Sideslip} \times \text{LEX Type}$$

Validation



Actual	Predicted	Valid
0.315	(0.30 , .33)	✓

Discovery, Prediction

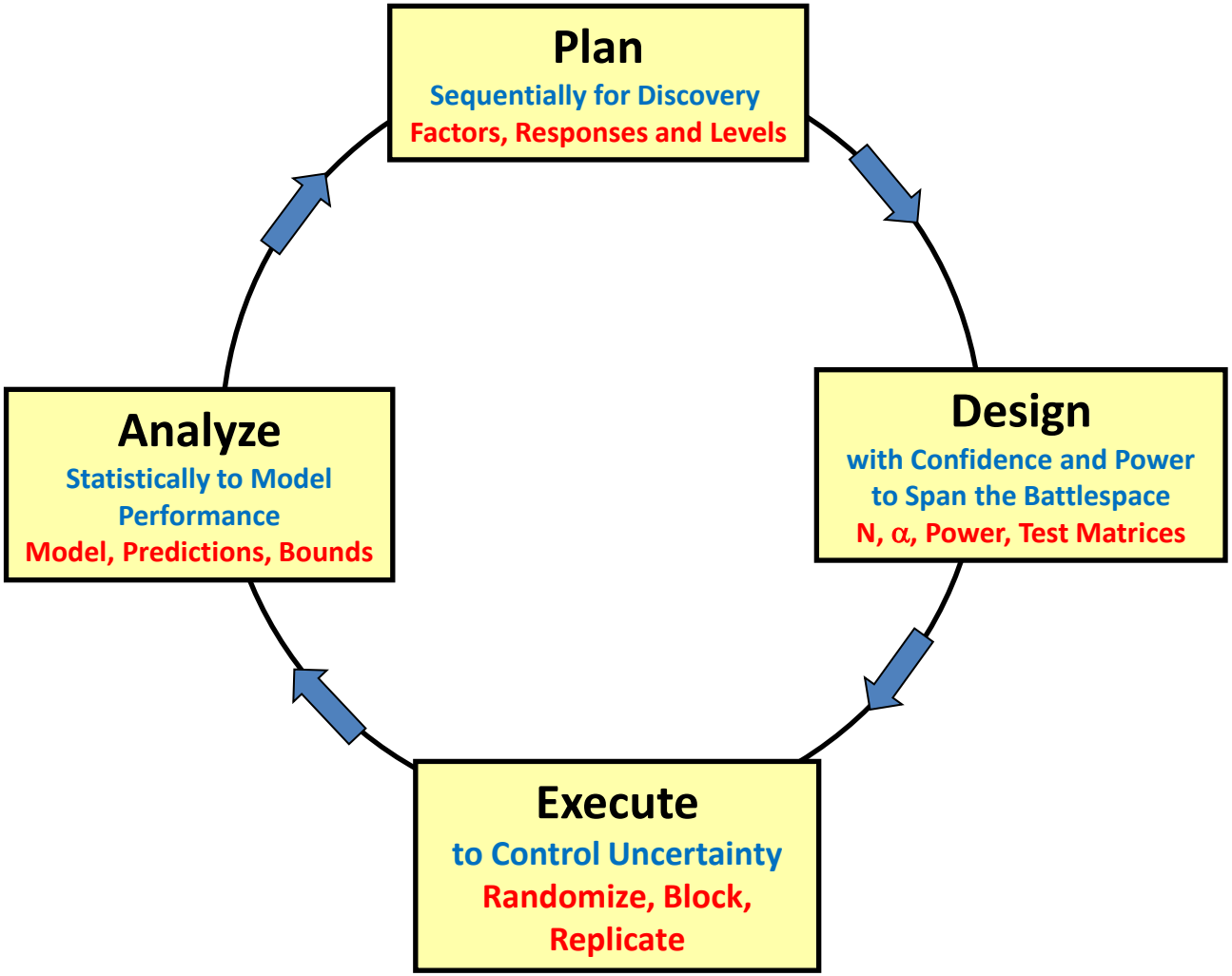


Not simple but doable with this systematic approach.

How to Execute

Four Stages

- Plan deliberately: problem, objective(s), outputs, inputs, background variables, phases
- Design for power in spanning battlespace: many choices of designs, depends on your system
- Execute with insurance against lurking variables and unknown-unknowns
- Objectively analyze with statistical methods (ANOVA/Regression) to determine what matters, direction, magnitude



Why DOE Makes Sense

DT&E: Science & Engineering are Vital to Success of our Tests

We already have good science in our DT&E!

We understand sys-engineering, guidance, aero, mechanics, materials, physics, electromagnetics …

DOE introduces the *Science of Test*

Why DOE Makes Sense

OT&E: Operations Skills are Vital to the Success of Test

Similarly: we already have good ops in our OT&E!

We understand attack, defense, tactics, ISR, mass, unity of command, artillery, CAS, ASW, AAW, armored cav…

DOE adds the *Science of Test*

We make decisions too important to be left to professional opinion alone…our decisions should be based on *mathematical fact*

Greg Hutto

Managing Expectations

Observation by a Practitioner

- At this point in history, (for OT) using DOE simply means laying out the primary factors that affect the response variable in at worst a notional design (and at best a design that one could readily use with proper resources and leadership support)

Dr. R. McIntyre Feb 2011

What Works (for us)

- *DOE provides for efficient testing and more useful results – but not necessarily at a reduced up front cost*
- *DOE is most effectively applied early in the development process where build a little, test little is cost effective*
- *Know your process; know the tool*
  - *Investing the time up front for process decomposition (MBTD/E) will pay great dividends in developing the experimental design*
  - *Use a DOE practitioner to assist in the actual design development (then execute the design)*
  - *Clearly articulate the pros and cons of each design (metrics scorecard)*
- *Ask better questions ;get better answers*
- *Even when DOE is not the correct tool to use for a particular application, it will at least aid you in discovering the most useful demonstrations to observe (May need to use other DOE-like tools – HTT)*

Design of Experiments

"Managing Expectations"

James "JD" Carpenter
carpenter@avwtech.com
(757) 361-5830

Chris Hauser
hauser@avwtech.com
(757) 361-9011

AVW Technologies INC
860 Greenbrier Circle
Chesapeake, VA 23320
www.avwtech.com

Design of Experiments

"Managing Expectations"

QUESTIONS?

BACK-UPS

DOE Metrics Scorecard

	Basic Report Card - Designed Experiments				
Wheel	Design Alternative	0	1	2	3
	Design Name	Baseline	CCD x3Cat	2^5+4cp	2^5-1+4cp
Plan	Number of Factors				
	Levels ea Factor				
	Num Responses (MOPS)				
	Real-values?				
	Objective?				
Design	Test Events (N)				
	Savings (-Incr)				
	Aliasing/Res/Ortho/Confound				
	comparisons)				
	2 σ Power				
	Name Design Strategy				
Execute	Randomized?				
	Blocked or calibrated?				
	Replicates? True?				
Analy	Pred Model Supported				
	FDS Pred Err @50/95%				
	Leverage Avg/Max				
	VIF Avg/Max				

DOE expert assistance recommended

Aerial Tgts Example

	Aerial Target Report Card - Designed Experiments				
Wheel	Design Alternative	0	1	2	3
	Design Name	Baseline	Factorial	2^(6-1)x3	7v 2/3 D Opt
Plan	Number of Factors	3	3	7	7
	Levels ea Factor	2x2x3	2x2x3	2,3	2,3
	Num Responses (MOPS)	1	1	1	1
	Real-values?	no	no	no	no
	Objective?	no	no	no	no
Design	Test Events (N)	13	12	96 (12)	46 (6)
	Savings (-Incr)	--	8%	8%	54%
	Aliasing/Orthogonality	Res II (A=B)	Full Res	RV+	
	comparisons)	5%	5%	5%	5%
	2 σ Power	5-65%	50-82%	99.90%	99%
	Name Design Strategy	??	Factorial	FractionxCat	Dopt Fract
Execute	Randomized?	--	--	--	--
	Blocked or calibrated?	--	--	--	--
	Replicates? True?	--	--	--	--
Analy	Pred Model Supported	Main Eff	3 FI	3FI	2FI
	FDS Pred Err @50/95%	.72/1.1	.71/.71	.33/.42	.66/.77
	Leverage Avg/Max	.38/1	.5/.5	.375/.375	.37/.47
	VIF Avg/Max	2/2.5	1/1	1/1	1.2/1.3

- Summary thoughts … avoid binary, define test event, max events per sortie/mission, create design alternatives, exploit sequential experimentation