Measuring the Value of Test and Evaluation Using Uncertainty Reduction

Eileen Bjorkman
Air Force Flight Test Center
Shahram Sarkani, Ph.D., P.E.
Thomas A. Mazzuchi, D.Sc.
George Washington University

This presentation is based on work leading to a dissertation submitted to the George Washington University in partial fulfillment of the Doctor of Philosophy Degree.

Overview

- Motivation – The Need for an Objective Test Value Metric
- Technical Uncertainty as a Value Metric
- Decision Analysis Framework
- Technical Uncertainty Framework
- Measuring Technical Uncertainty
- Example Application to Test & Evaluation Problems
- Conclusions and Future Work

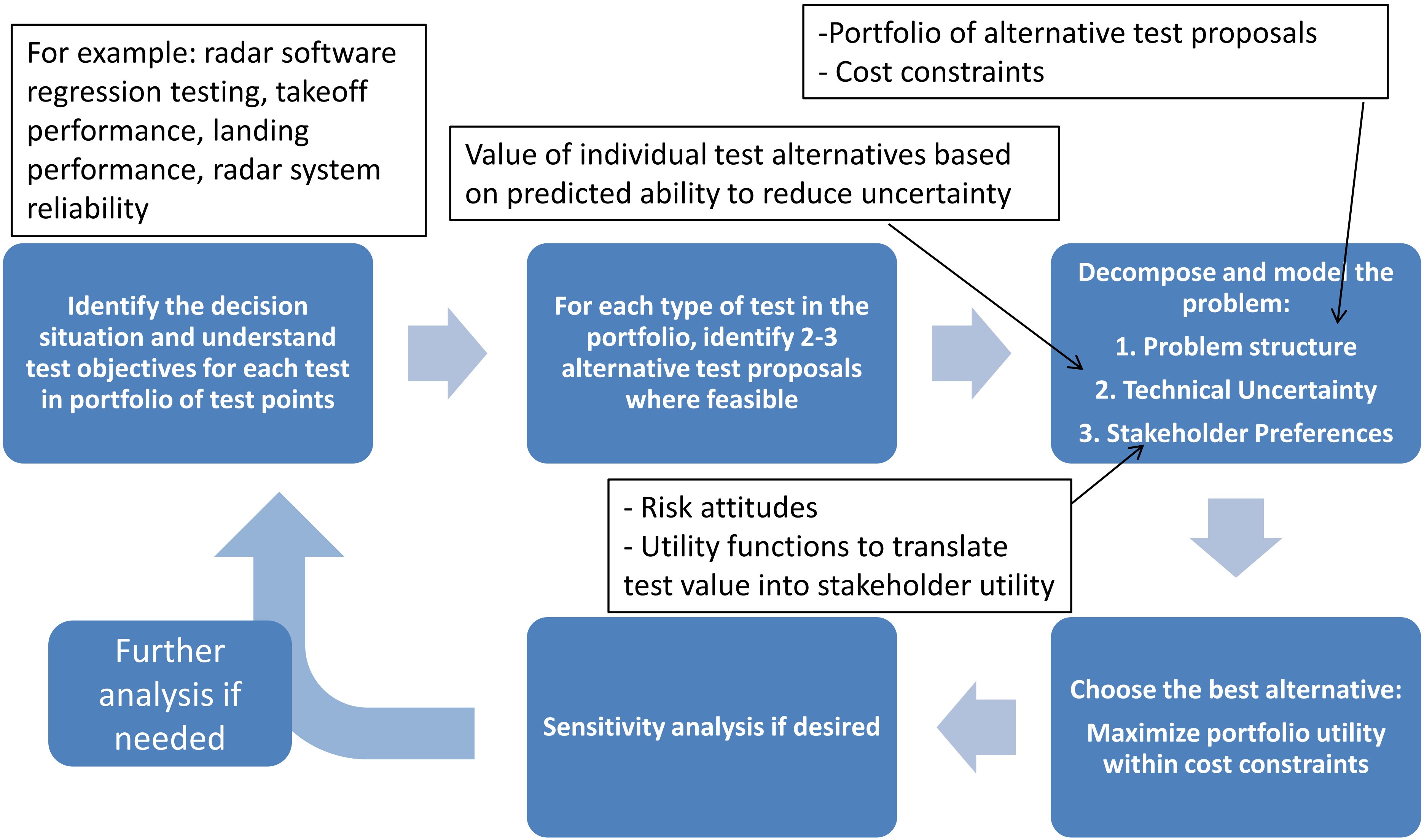
Motivation: The Need for an Objective Test Value Metric

- Test value is generally proposed as:
  - Early identification of problems
  - Cost avoidance
  - Cost of rework
- All must be estimated – we can never know what would not be found if a test was not conducted
- All are valid reasons to test, but none can be measured absolutely to provide a common reference point

Technical Uncertainty as a Value Metric

- Cost avoidance and rework are not good value metrics because they cannot be defined or measured during a weapon system life cycle
- Technical uncertainty can be estimated prior to a test and evaluated after the test using statistical techniques
  - Provides a common metric of comparison
  - Can be measured
  - Is easily related to risk
- Technical uncertainty as a value metric, along with cost, schedule, and other parameters of interest can be related to stakeholder utility as a basis of comparison

Decision Analysis Framework*



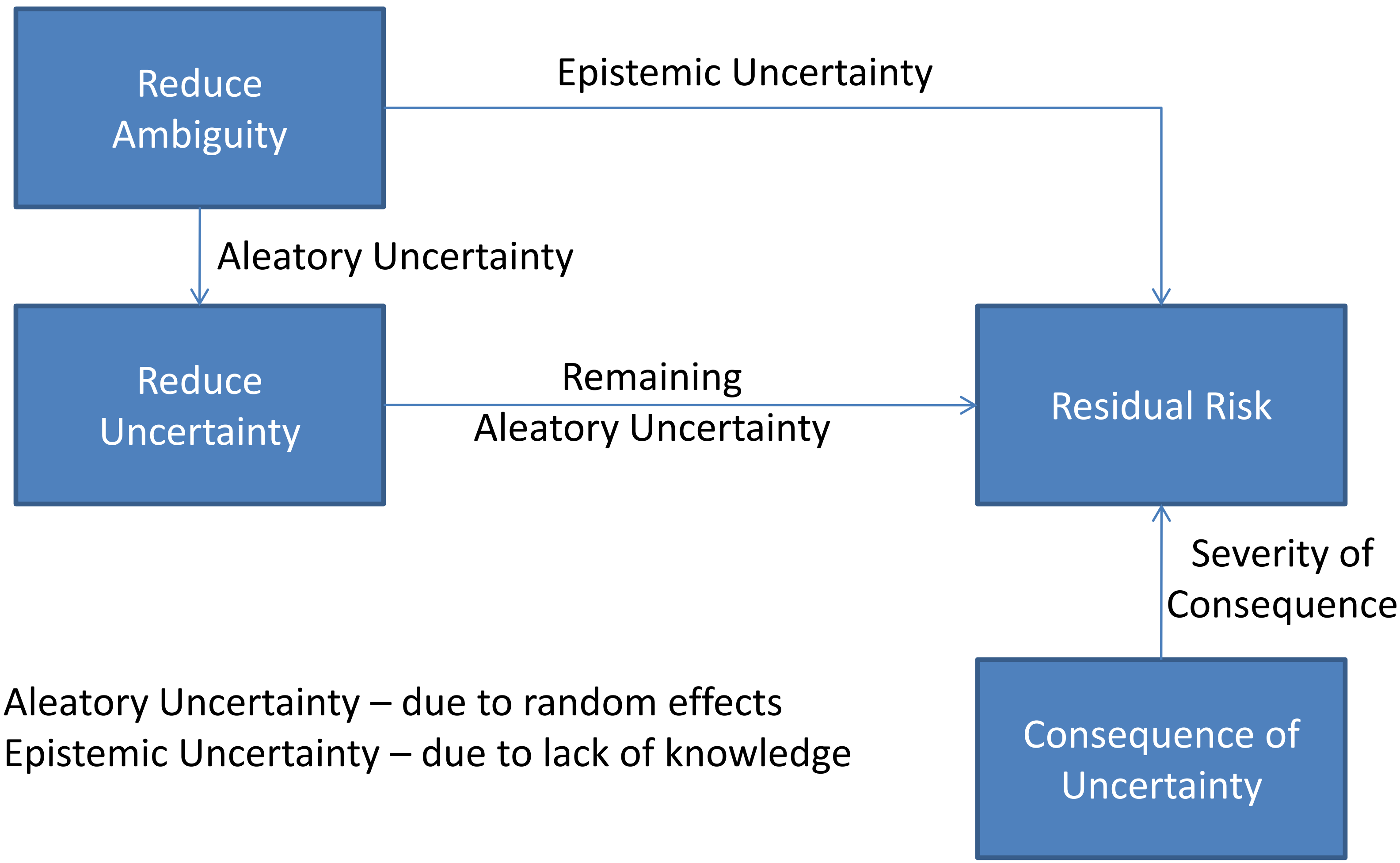
*Adapted from Clemen, R. T., & Reilly, T. (2001). *Making Hard Decisions with DecisionTools(R)* (Second ed.). Pacific Grove, California: Duxbury Thomson Learning.

Technical Uncertainty Framework Developed From Literature Review*

	Unknowable Uncertainty	Knowable Uncertainty (Ambiguity)
Essential Elements of Uncertainty:		
Components of Uncertainty	Aleatory	Epistemic
Sources of Technical Uncertainty	Measurement (input/output), model structure, model selection, prediction error, inference uncertainty	
Application to Test and Evaluation:		
Test Goal	Reduce Uncertainty	Characterize Uncertainty
Type of Model Available	Physics-based	None or limited
	Empirical	
Characterization of Uncertainty:		
Uncertainty Evaluation	Entropy and entropy-based measures	
Uncertainty Reduction	Model Using and Updating: Using data to reduce uncertainty and validate/update model	Model Building: Using data to build model and estimate uncertainty
Uncertainty Depiction (not an exhaustive list)	Probability Distribution/Summary Statistics Confidence, Prediction, Tolerance, or Credible Intervals Akaike Information Criterion, Deviance Information Criterion	

*References in backup slides

Technical Uncertainty Relationships



Aleatory Uncertainty – due to random effects
Epistemic Uncertainty – due to lack of knowledge

Measuring Technical Uncertainty: What to Use?

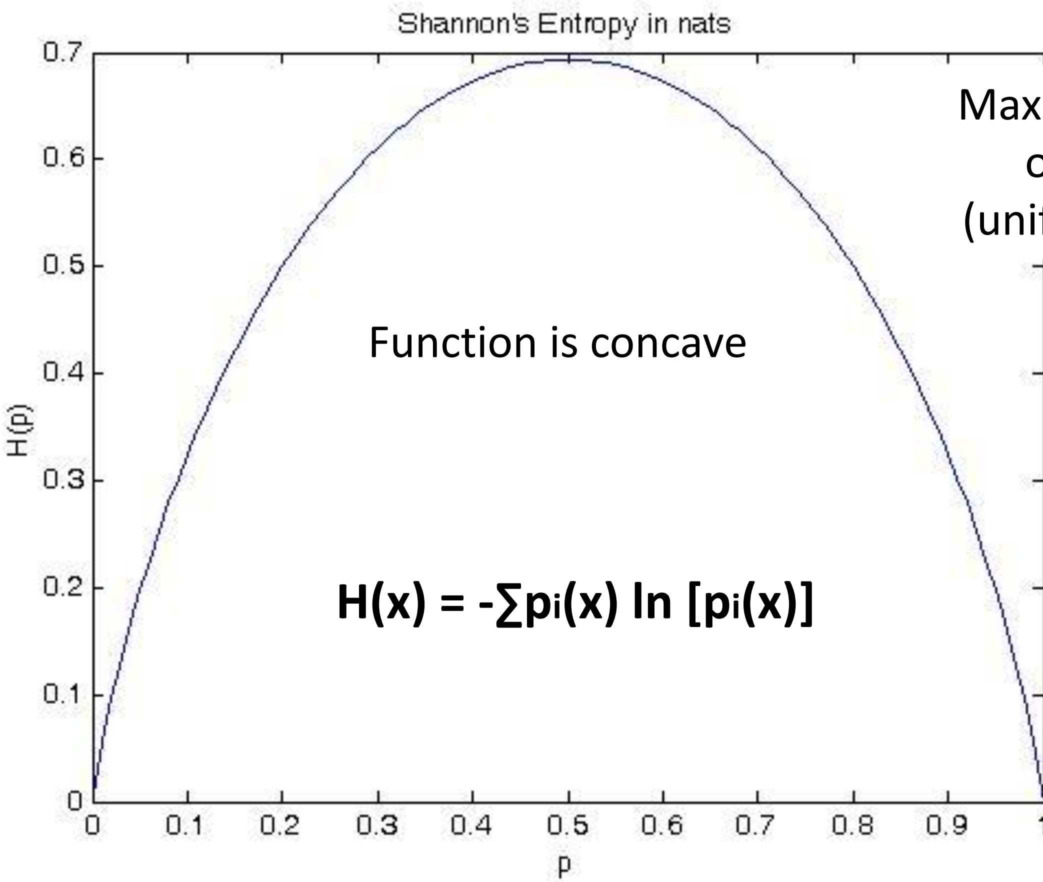
- Desirable properties*:
  - Concavity
  - Attaining global maximum at the uniform distribution (all values are equally likely)
- Shannon's entropy – meets above properties and is easy to measure
- Variance for normal random variables does not meet above properties, but can be considered for the case of normal random variables

Variance in general is not the best measure of uncertainty

* Ebrahimi, N., Soofi, E. S., & Soyer, R. (2010). Information Measures in Perspective. *International Statistical Review, 78(3), 383-412. doi: 10.1111/j.1751-5823.2010.00105.x*

Measuring Uncertainty: Shannon's Entropy

Shannon's Entropy for a Binary Variable



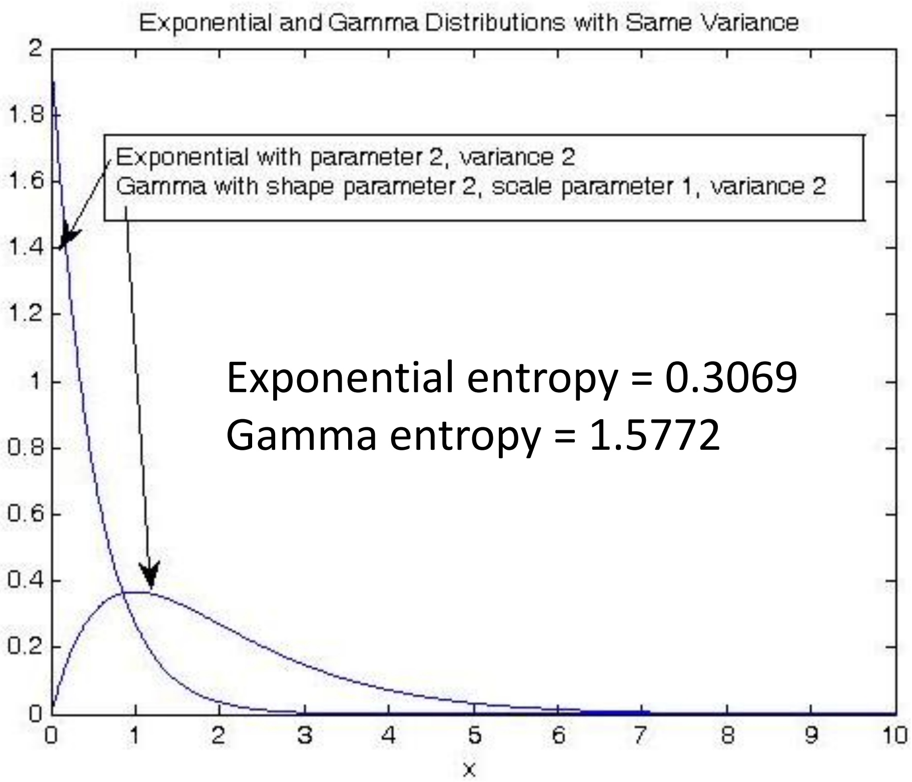
Maximum uncertainty occurs at p=0.5 (uniform distribution)

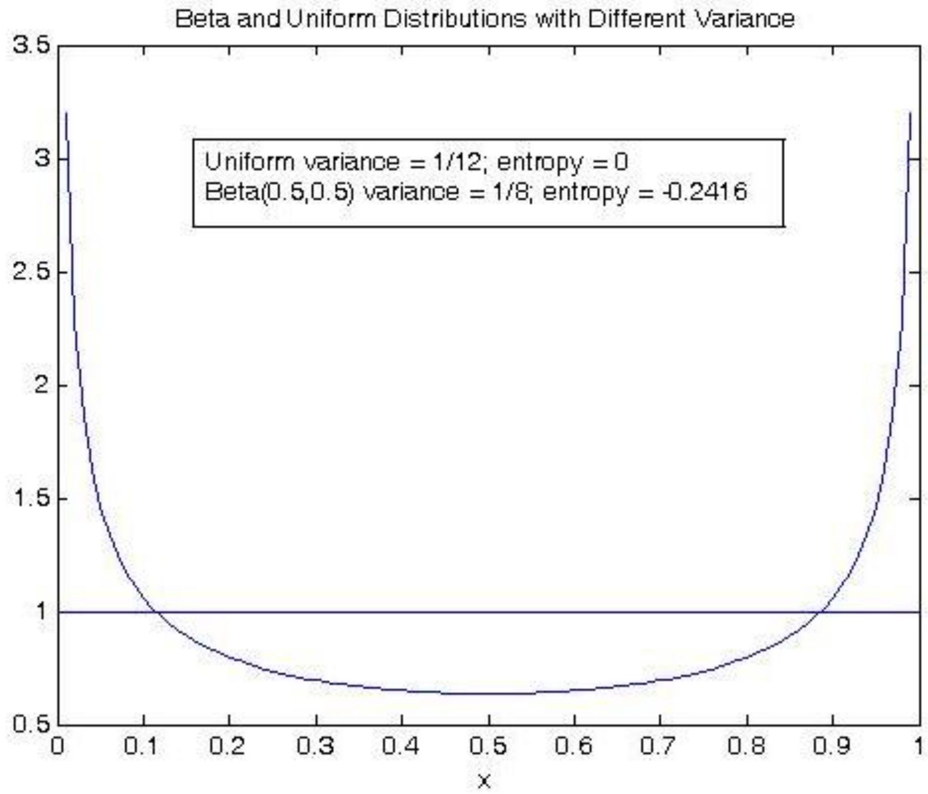
Function is concave

$$\mathbf{H(x) = -\sum p_i(x) \ln [p_i(x)]}$$

Example adapted from Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory (Second ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.*

Examples Where Variance Is Not a Good Uncertainty Measure*



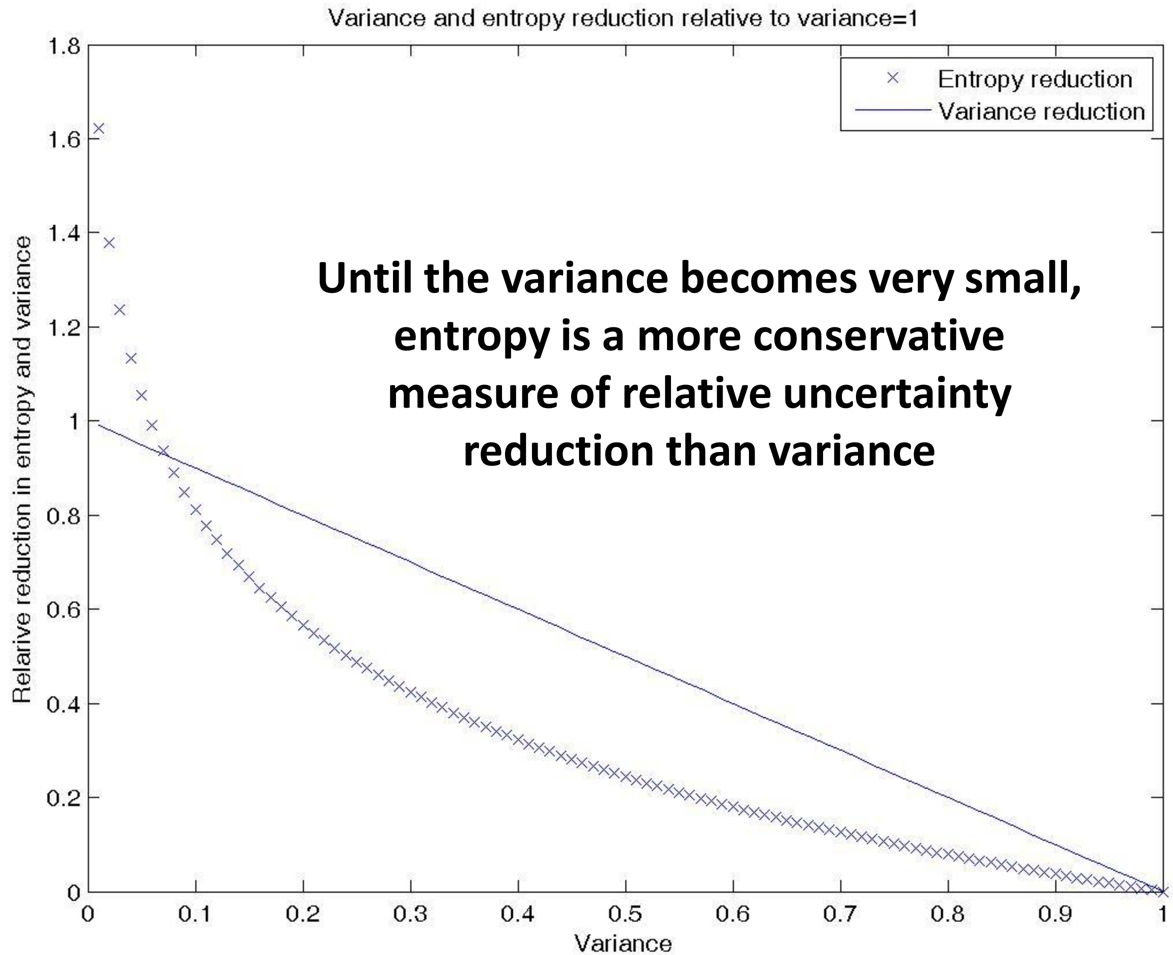


Equal variance, but very different entropy; exponential entropy is significantly lower

Uniform distribution has smaller variance, but beta distribution has lower entropy

* Examples adapted from Ebrahimi, N., Soofi, E. S., & Soyer, R. (2010). Information Measures in Perspective. *International Statistical Review, 78*(3), 383-412. doi: 10.1111/j.1751-5823.2010.00105.x

Entropy and Gaussian Variance:
Entropy is the more conservative measure



Based on Matlab simulation of 10000 replications

Intermediate Conclusion

- Shannon’s entropy is a better technical uncertainty measure than variance
  - Works for all distributions and is simple to compute
  - Variance is sometimes misleading; probability distribution function with smallest variance does not necessarily have least uncertainty
  - Is more conservative than variance for a normal random variable
  - Can compare all uncertainties in a test portfolio with same units (nats)

Simple Examples: Using the Technical Uncertainty Measure

- Maximizing test utility to two decision makers (one test, no constraints)
- Portfolio optimization based on maximizing overall test value/utility subject to cost constraint

Maximizing Test Utility to Multiple Decision Makers: The "U-100"

- Overall test objective: Obtain quantitative performance data for incorporation into the U-100 Flight Manual (the U-100 is a notional aircraft)
- Two flight test programs are proposed:
  – A: Use hand held data; test cost $25000, 1 month test
  – B: Use full instrumentation; test cost $50000+2 month schedule delay to install instrumentation
- Two decision makers, one risk tolerant and one risk averse
- Which test has the greater value: A or B?
- Which test has greater utility, given the two decision makers?

Decision Analysis Framework for the U-100 Landing Test

Overall test objective: Obtain quantitative performance data for incorporation into the U-100 flight manual

Test Alternatives:
0: No test
A: $25K, one month
B: $50K, three months

-Two test alternatives
- No explicit cost or schedule constraints

Identify the decision situation and understand test objectives for each test in portfolio of test points

For each type of test in the portfolio, identify 2-3 alternative test proposals where feasible

Decompose and model the problem:

1. Problem structure

2. Technical Uncertainty

3. Stakeholder Preferences

Derived test objectives:
- With 99% confidence, determine the maximum rollout distance for various U-100 configurations
- Determine best braking technique
- Validate physics-based model

Value of individual test alternatives based on predicted ability to reduce uncertainty

- One risk-averse decision maker, one risk-tolerant decision maker
- Elicit test utilities from DMs based on test value

Choose the best alternative:

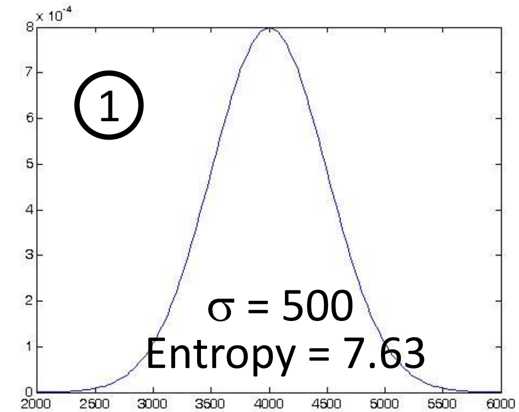
Maximize portfolio utility within cost constraints

Use simple weights combined with test utility to maximize utility

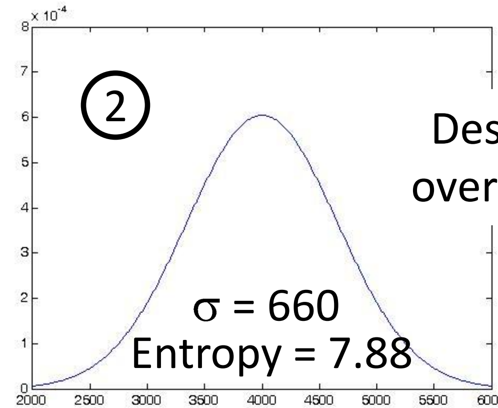
Technical Uncertainty Framework for the U-100 Landing Test

	Unknowable Uncertainty	Knowable Uncertainty (Ambiguity)
Essential Elements of Uncertainty:		
Components of Uncertainty	Aleatory – dominant	Epistemic -- small (good model)
Sources of Technical Uncertainty	Variability in landing/braking process, instrumentation precision and accuracy, using model to compare to and predict flight manual values	
Application to Test and Evaluation:		
Test Goal	Reduce Uncertainty	
Type of Model Available	Physics-based	
Characterization of Uncertainty:		
Uncertainty Evaluation	Entropy	
Uncertainty Reduction	Model Using and Updating: Using data to reduce uncertainty and validate/update model	
Uncertainty Depiction (Based on proposed analysis techniques)	Probability Distribution/Summary Statistics Confidence or Prediction Intervals	

Using Uncertainty as a Value Measure

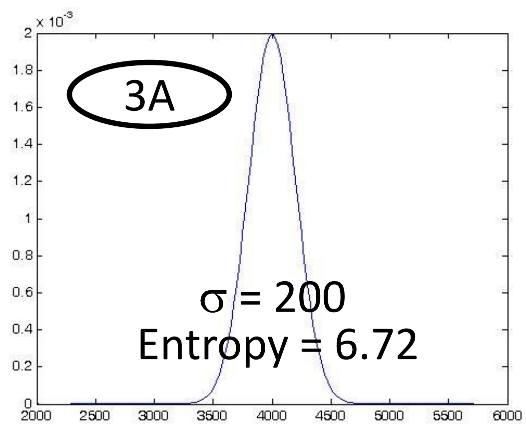


SME Estimate After
Initial Design

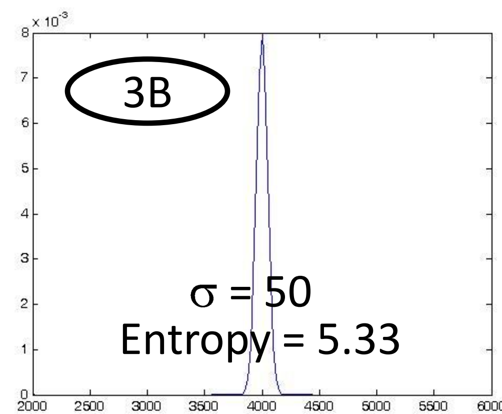


Designers were
overly optimistic!

Initial Model
Estimate



Predicted Flight Test
Result With Hand
Held Data



Predicted Flight Test
Result With Full
Instrumentation

B clearly has more value, but does it have more utility, given the two decision makers?

Computing Utility Using Simple Weights

Test	Factor	Measure	Risk-Averse Decision Maker			Risk-Tolerant Decision Maker		
			Utility (U)	Weight (W)	Total (U*W)	Utility	Weight	Total
None	Schedule	0	0	5	0	10	10	100
	Cost	0	0	6	0	10	9	90
	Uncertainty reduction	0	0	10	0	0	3	0
UTILITY OF THE OPTION TO NOT TEST					**0**	**Highest utility if there is no requirement to test**		**190**
Test A	Schedule	1	10	5	50	9	10	90
	Cost	25000	10	6	60	7	9	63
	Uncertainty reduction	1.16	8	10	80	6	3	18
UTILITY OF TEST OPTION A					**190**	**Highest utility if there is a requirement to test**		**171**
Test B	Schedule	3	10	5	50	8	10	80
	Cost	50000	9	6	54	7	9	63
	Uncertainty reduction	2.55	10	10	100	7	3	21
UTILITY OF TEST OPTION B				**Highest utility**	**204**			**164**

Computing Utility if Both DMs Have a Vote of Equal Weight

Test	DM 1 Total	DM 2 Total	Total
None	0	190	190
Test A	190	171	361
Test B	204	164	368

Slightly higher by small margin

But, then there was a budget cut. Test A was selected due to the lower cost and the fact that it has only slightly less utility.

And yet another budget cut reduced the number of test points that could be flown …

Test Results: Landing Data*, Pre- and Post-Test Monte-Carlo Analysis

Configuration		σ (ft)		Entropy (nats)	
Flaps	Braking	**Pre-test model**	**Post-test model**	**Pre-test model**	**Post-test model**
45	Moderate	858.4	404.1	8.17	7.42
45	Heavy	755.0	402.8	8.05	7.42
45	Max	723.9	578.0	8.00	7.78
60	Moderate	731.3	352.8	8.01	7.28
60	Heavy	629.7	350.6	7.86	7.28
60	Max	608.7	500.1	7.83	7.63
100	Moderate	554.6	295.7	7.74	7.11
100	Heavy	501.0	285.9	7.64	7.07
100	Max	483.4	398.9	7.60	7.41
	Pooled	**660.4**	**406.3**	**7.88**	**7.38**

Variance/standard deviation decreases significantly post-test, but entropy decreased only slightly → there is still a significant amount of uncertainty!

* Data taken from McNamar, L. F., & Gordon, H. C. (1963). T-38A Category II Performance Test. Edwards AFB: Air Force Flight Test Center.

Notional Test Problem:

Maximizing Overall Test Portfolio Value Subject to Cost Constraint

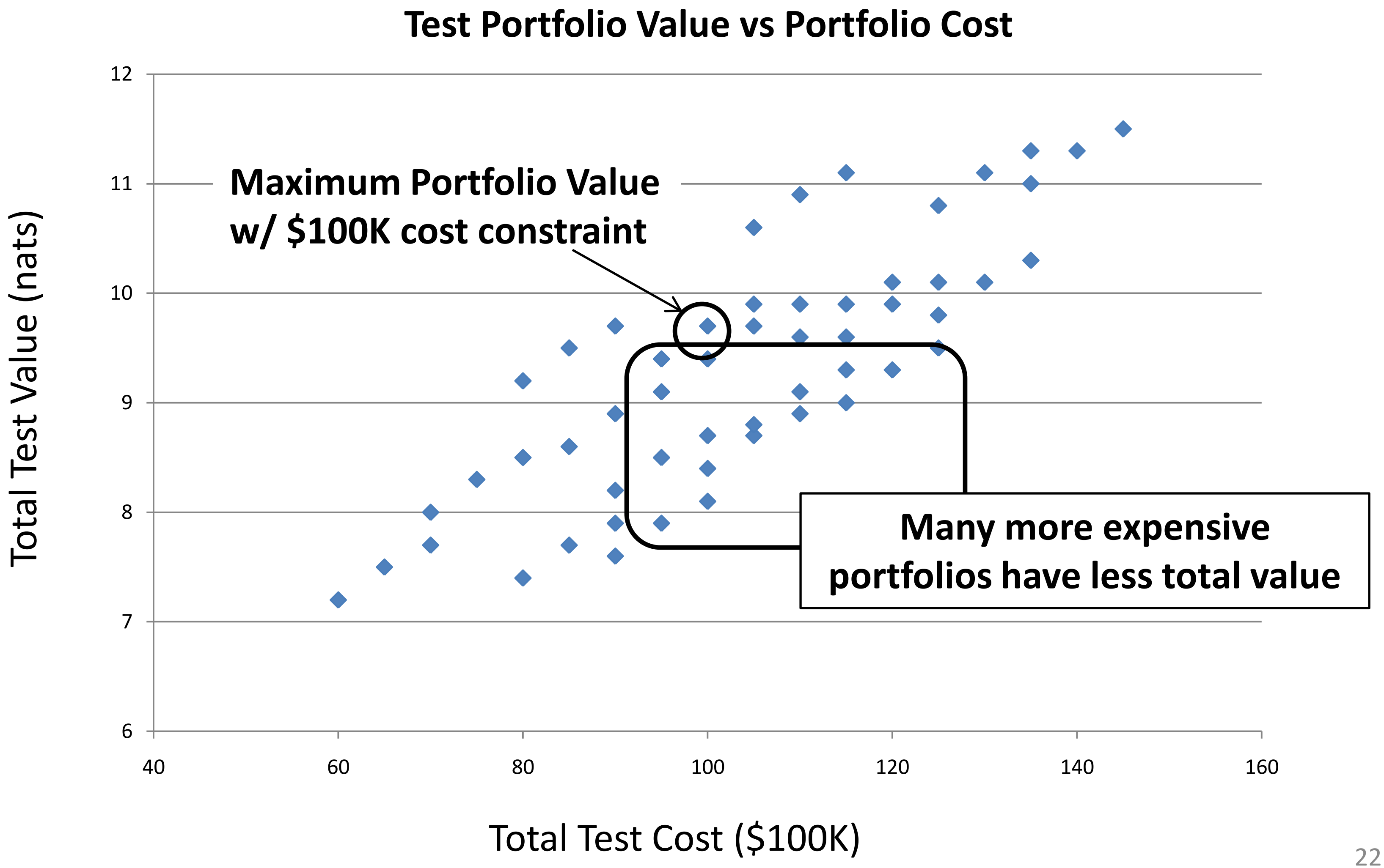
Test Value (Entropy Reduction)			
Test	**Opt A**	**Opt B**	**Opt C**
Landing	1.2	2.6	-
Radar	2.4	2.7	2.9
ADC	1.9	2.1	2.3
HUD	1.7	2.5	3.7

Test Cost (x100K)			
Test	**Opt A**	**Opt B**	**Opt C**
Landing	25	50	-
Radar	5	10	15
ADC	20	40	50
HUD	10	20	30

Maximum test value assuming total test budget of $100K
(test value = 9.7 at $100K)

ADC = Air Data Computer
HUD = Heads-Up Display

Notional Test Problem:
Maximizing Overall Test Portfolio Value Subject to Cost Constraint



Conclusions

- Technical uncertainty reduction measured using entropy provides a consistent way to measure the value of a test
- Utility of test to multiple stakeholders must still be taken into account before selecting actual test option (highest value test may not have greatest utility)
- Although technical uncertainty reduction is an estimate before the test executes, it can be measured once test is complete to determine if test value metric was met (may also lead to decision to terminate test early or conduct additional testing)

Future Work

- Apply full decision analysis framework to a portfolio of test points selected from real-world flight test problems and both single and multiple stakeholders/decision makers:
  - Physics-based models
  - Empirical models
  - Include model uncertainty and multi-model inference
  - Large variance data with small sample sizes
- Examine different utility functions and optimization techniques
- Continue technical uncertainty research and update technical uncertainty framework as required, with focus on analysis techniques currently in use:
  - Traditional statistical approaches
  - Bayesian techniques
  - Information theoretic approaches
  - Results represented via different intervals (e.g., confidence, prediction, tolerance, and credible)

Contact Information

Eileen A. Bjorkman
(661) 275-2074
eileen.bjorkman@edwards.af.mi

Shahram Sarkani, Ph.D., P.E
(888) 694-9627
sarkani@gwu.edu

Thomas A. Mazzuchi, D.Sc.
(202) 994-7541
mazzu@gwu.edu

Backup

References (1/4)

- Abbas, A. E. (2004). Entropy methods for adaptive utility elicitation. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 34*(2), 169-178.
- Abel, P. S., & Singpurwalla, N. D. (1994). To Survive or to Fail: That Is the Question. *The American Statistician, 48*(1), 18-21.
- An, J., Acar, E., Haftka, R. T., Kim, N. H., Ifju, P. G., & Johnson, T. F. (2008). Being Conservative with a Limited Number of Test Results. *Journal of Aircraft, 45*(6), 1969-1975.
- Anderson, D. R., Burnham, K. P., & Thompson, W. L. (2000). Null Hypothesis Testing: Problems, Prevalence, and an Alternative. *The Journal of Wildlife Management, 64*(4), 912-923.
- Aven, T. (2004). On How to Approach Risk and Uncertainty to Support Decision-Making. *Risk Management, 6*(4), 27-39.
- Ben-Haim, Y. (1999). Design certification with information-gap uncertainty. *Structural Safety, 21*(3), 269-289. doi: 10.1016/s0167-4730(99)00023-5
- Browning, T. R. (2003). On Customer Value and Improvement in Product Development Processes. [Regular Paper]. *Systems Engineering, 6*(1), 49-61.
- Browning, T. R., Deyst, J. J., Eppinger, S. D., & Whitney, D. E. (2002). Adding value in product development by creating information and reducing risk. *Engineering Management, IEEE Transactions on, 49*(4), 443-458.
- Burnham, K. P., & Anderson, D. R. (2004). Mutlimodel inference: understanding AIC and BIC in model selection. *Sociol. Methods Res., 33*(2), 261-304. doi: 10.1177/0049124104268644
- Cacuci, D. G. (2003). *Sensitivity and uncertainty analysis* (Vol. 1, Theory).
- Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory* (Second ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.

References (2/4)

- DeLoach, R. (2008). *Bayesian Inference in the Modern Design of Experiments*. Paper presented at the 46th AIAA Aerospace Sciences Meeting and Exhibit, Reno, Nevada.
- De Meyer, A., Loch, C. H., & Pich, M. T. (2002). Managing Project Uncertainty: From Variation to Chaos. [Article]. *MIT Sloan Management Review, 43*(2), 60-67.
- Deyst, J. J. (2002). *The application of estimation theory to managing risk in product developments.* Paper presented at the Digital Avionics Systems Conference, 2002. Proceedings. The 21st.
- Dezfuli, H., Kelly, D., Smith, C., Vedros, K., & Galyean, W. (2009). *Bayesian Inference for NASA Probabilistic Risk and Reliability Analysis*. (NASA/SP-2009-569). Washington, DC.
- Draper, D. (1995). Assessment and Propagation of Model Uncertainty. *Journal of the Royal Statistical Society. Series B (Methodological), 57(1), 45-97.*
- Ebrahimi, N., Soofi, E. S., & Soyer, R. (2010). Information Measures in Perspective. *International Statistical Review, 78*(3), 383-412. doi: 10.1111/j.1751-5823.2010.00105.x
- Emery, A. F. (2009). Estimating deterministic parameters by Bayesian inference with emphasis on estimating the uncertainty of the parameters. [Article]. *Inverse Problems in Science & Engineering, 17*(2), 263-274. doi: 10.1080/17415970802404985
- Ferrero, A., & Salicone, S. (2004). The random-fuzzy variables: a new approach to the expression of uncertainty in measurement. *Instrumentation and Measurement, IEEE Transactions on, 53*(5), 1370-1377.
- Giadrosich, D. L. (1995). *Operations Research Analysis in Test and Evaluation*. Washington, DC: American Institute of Aeronautics and Astronautics.
- Hamada, M., Johnson, V., Moore, L. M., & Wendelberger, J. (2004). Bayesian Prediction Intervals and Their Relationship to Tolerance Intervals. *Technometrics, 46*(4), 452-459.
- Hatfield, A. J., & Hipel, K. W. (1999). *Understanding and managing uncertainty and information.* Paper presented at the Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings. 1999 IEEE International Conference on.
- Hess, J. T., & Valerdi, R. (2010). *Test and evaluation of a SoS using a prescriptive and adaptive testing framework.* Paper presented at the System of Systems Engineering (SoSE), 2010 5th International Conference on.

References (3/4)

- Hodges, J. S. (1987). Uncertainty, Policy Analysis and Statistics. *Statistical Science, 2*(3), 259-275.
- Hoppe, M., Engel, A., & Shachar, S. (2007). SysTest: Improving the Verification, Validation, and Testing Process--Assessing Six Industrial Pilot Projects. *Systems Engineering, 10*(4), 323-347.
- JCGM. (2010). Evaluation of measurement data -- Guide to the expression of uncertainty in measurement.
- Jones, B. L. (1995). Near Real-Time Approach to Statistical Flight Test. *Journal of Aircraft, 32*(4), 782-786.
- Kadvany, J. (1996). Taming Chance: Risk and the Quantification of Uncertainty. *Policy Sciences, 29*(1), 1-27.
- Kraft, E. M. (2010). After 40 Years Why Hasn't the Computer Replaced the Wind Tunnel? *ITEA Journal, 31*(3), 329-346.
- Lenz, R., & Gardner, L. (1997). *Risk in Weapon System Acquisition: A Decision Support Approach to the Economics of Test and Evalaution* Paper presented at the ITEA Conference on the Economics of Test and Evaluation, Atlanta, Georgia.
- McNamar, L. F., & Gordon, H. C. (1963). T-38A Category II Performance Test. Edwards AFB: Air Force Flight Test Center.
- Merrick, J. R. W. (2009). Bayesian Simulation and Decision Analysis: An Expository Survey. [Article]. *Decision Analysis, 6*(4), 222-238. doi: 10.1287/deca.1090.0151
- Ng, S. H., & Chick, S. E. (2006). Reducing parameter uncertainty for stochastic systems. *ACM Trans. Model. Comput. Simul., 16*(1), 26-51. doi: 10.1145/1122012.1122014
- Nilsen, T., & Aven, T. (2003). Models and model uncertainty in the context of risk analysis. *Reliability Engineering & System Safety, 79*(3), 309-317. doi: 10.1016/s0951-8320(02)00239-9
- Park, I., & Grandhi, R. V. (2010). *Quantification of Multiple Types of Uncertainty in Computer Simulation Using Bayesian Model Averaging*. Paper presented at the 51st AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Orlando, Florida.
- Parry, G. W. (1996). The characterization of uncertainty in Probabilistic Risk Assessments of complex systems. *Reliability Engineering & System Safety, 54*(2-3), 119-126. doi: 10.1016/s0951-8320(96)00069-5

References (4/4)

- Paté-Cornell, M. E. (1996). Uncertainties in risk analysis: Six levels of treatment. *Reliability Engineering & System Safety, 54*(2-3), 95-111. doi: 10.1016/s0951-8320(96)00067-1
- Paté-Cornell, M. E., & Dillon, R. L. (2006). The Respective Roles of Risk and Decision Analyses in Decision Support. [Article]. *Decision Analysis, 3*(4), 220-232. doi: 10.1287/deca.1060.0077
- Pich, M. T., Loch, C. H., & Meyer, A. d. (2002). On Uncertainty, Ambiguity, and Complexity in Project Management. *Management Science, 48*(8), 1008-1023.
- Pilch, M., Trucano, T. G., & Helton, J. C. (2011). Ideas underlying the Quantification of Margins and Uncertainties. *Reliability Engineering and System Safety*. doi: 10.1016/j.ress.2011.03.106
- Sankararaman, S., & Mahadevan, S. (2011). *Uncertainty Quantification and Model Validation under Epistemic Uncertainty due to Sparse and Imprecise Data*. Paper presented at the 52nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, Denver, Colorado.
- Schrader, S., Riggs, W. M., & Smith, R. P. (1993). Choice over uncertainty and ambiguity in technical problem solving. *Journal of Engineering and Technology Management, 10*(1-2), 73-99. doi: 10.1016/0923-4748(93)90059-r
- Sheridan, T. B. (1995). Reflections on information and information value. *Systems, Man and Cybernetics, IEEE Transactions on, 25*(1), 194-196.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Linde, A. v. d. (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society. Series B (Statistical Methodology), 64*(4), 583-639.
- Tucker, A. A., & Dagil, C. H. (2009). Design of experiments as a means of lean value delivery to the flight test enterprise. *Systems Engineering, 12*(3), 201-217.
- Ward, S., & Chapman, C. (2008). Stakeholders and uncertainty management in projects. [Article]. *Construction Management & Economics, 26*(6), 563-577. doi: 10.1080/01446190801998708
- Wendelberger, J. R. (2010). Uncertainty in Designed Experiments. [Article]. *Quality Engineering, 22*(2), 88-102. doi: 10.1080/08982110903510420
- Zadeh, L. A. (1995). Discussion: Probability Theory and Fuzzy Logic Are Complementary Rather Than Competitive. *Technometrics, 37*(3), 271-276.