An Information-Theoretic Metric of System Complexity with Application to Engineering System Design

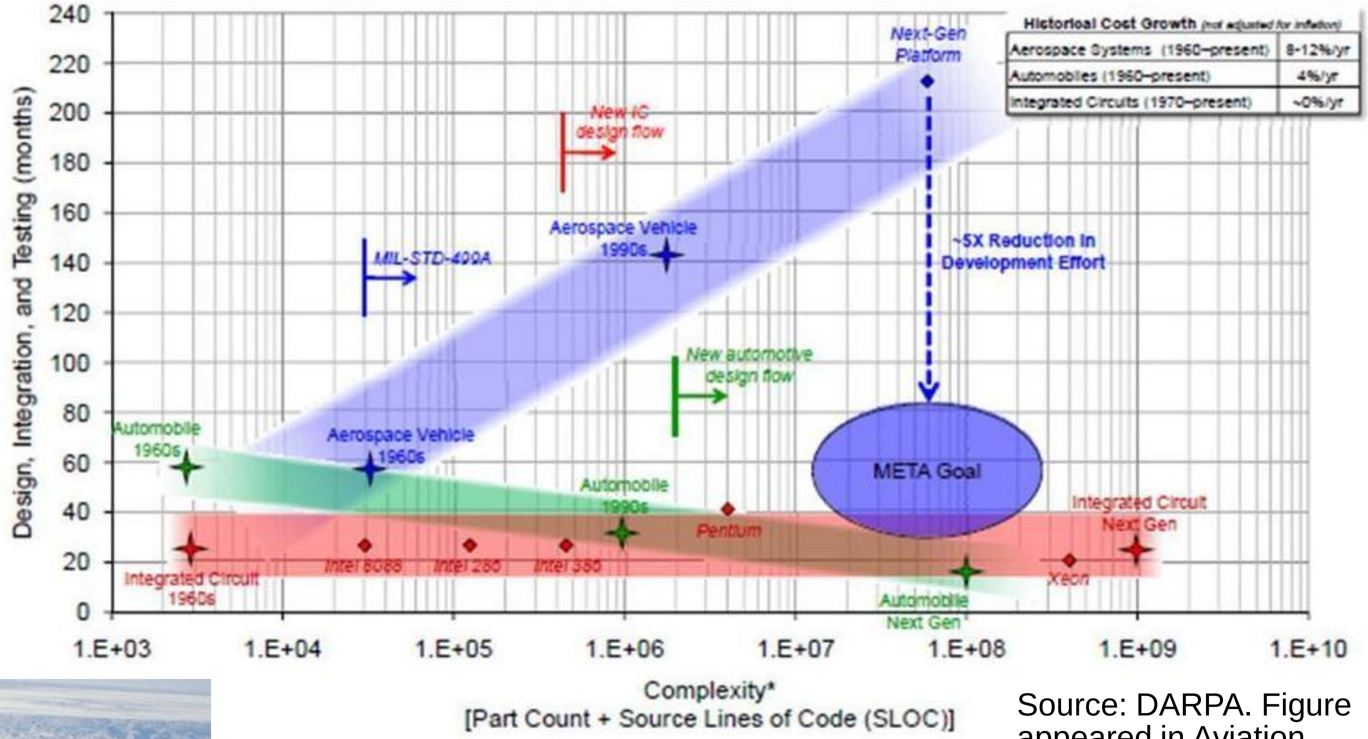
Douglas Allaire, Chelsea He, John Deyst, and Karen Willcox
Department of Aeronautics and Astronautics
Massachusetts Institute of Technology

7th Research Consortium for Multidisciplinary System Design
July 20, 2012
Purdue University, West Lafayette, IN

This work was funded in part by the DARPA META program through AFRL Contract FA8650-10-C-7083 and Vanderbilt University Contract VU-DSR #21807-S7, and by the Singapore University of Technology and Design.

Motivation

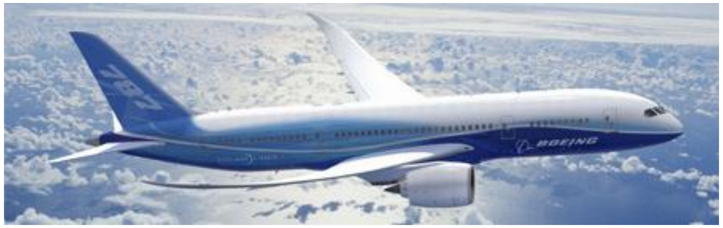
Development times and costs of aerospace systems have reached unsustainable levels – and are getting worse.



Historical Cost Growth *(not adjusted for inflation)*

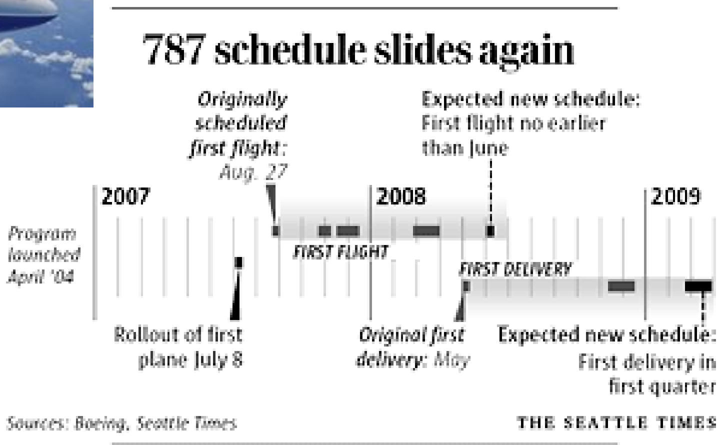
Aerospace Systems (1960–present)	8-12%/yr
Automobiles (1960–present)	4%/yr
Integrated Circuits (1970–present)	~0%/yr

Source: DARPA. Figure appeared in Aviation Week & Space Tech., Nov. 1-8 2011.



Source: www.boeing.com.

The Boeing 787 program has incurred significant cost and schedule overruns due to unexpected integration issues.



787 schedule slides again

Sources: Boeing, Seattle Times

THE SEATTLE TIMES

Outline

- Background
- Definition of system complexity
- Total uncertainty quantification for computer models
- Derivation of sensitivity indices
- Demonstration problems
- Conclusions

Background – Complexity

- Complexity in system design is an elusive concept…
  - Qualitatively:
    - Nebulous middle ground between order and chaos (Weaver 1948)
    - "I know it when I see it" (Johnson 1997)
  - Quantitatively:
    - Structure-based: source lines of code, number of parts, etc. (Griffin 1997)
    - Process-based: algorithmic complexity, computational complexity, etc. (Kolmogorov 1965, Chaitin 1969)
    - Information-based: information entropy, thermodynamic depth (Lloyd 1988)

Many metrics. The usefulness of each depends on context.

Background – Our context

- Generally agreed upon properties of a complex system
  - Consist of many parts
  - Parts interact
  - Difficult to model and understand
- Consider the design of a next generation infantry fighting vehicle
  - What are the quantities we truly care about when designing the vehicle?
    - Range
    - Acceleration
    - Quiet time duration
    - Armor capabilities
    - Cost
    - Development time
    - …

Source: www.inetres.com

Background – Information Entropy

Consider a random variable Y with probability mass function $p(y)$

The entropy of Y is defined as :

$$H(Y) = -\sum_i p(y_i)\log p(y_i),$$

where $y_1, y_2, \ldots$ are the values of y such that $p(y) \neq 0$

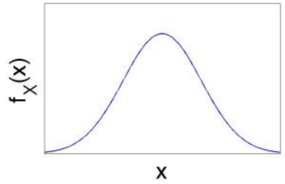
Consider a random variable X with probability density function $f_X(x)$

Differential entropy of X is defined as :

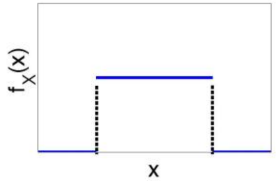
$$h(X) = -\int_{\mathbb{X}} f_X(x)\log f_X(x)dx$$

Examples :

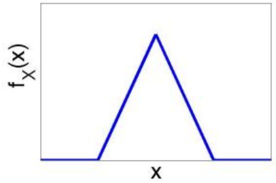
$$h(\mathcal{N}(\mu,\sigma^2)) = \frac{1}{2}\ln(2\pi e\sigma^2)$$



$$h(U[a,b]) = \ln(b-a)$$



$$h(T(a,b,c)) = \frac{1}{2} + \ln\left(\frac{b-a}{2}\right)$$



System Complexity

Proposed Definition: System Complexity

The potential of a system to exhibit unexpected behavior in the quantities of interest.

- Captures qualitative aspects of system complexity
  - notion of emergent behavior
  - Lack of understanding
- Can be quantitatively measured

Proposed Metric: System Complexity

Let $f_Q(q)$ be the probability density function of a quantity of interest. Then

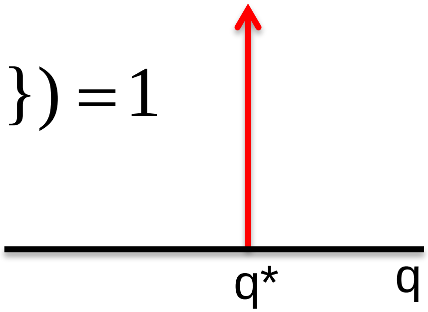
$$C(Q) \triangleq \exp\left\{-\int_{-\infty}^{\infty} f_Q(q)\ln f_Q(q)dq\right\} = \exp\{h(Q)\}$$

is a metric of system complexity as defined above.

Complexity Metric

- For the case where we have perfect knowledge, complexity $C(Q) = 0$

$$\Pr(\{Q = q^*\}) = 1$$



- For all other cases, $C(Q) \in [0, \infty)$
- For the case of a uniform random variable

$$Q \sim U[a, b]$$

$$C(Q) = \exp(\{h(Q)\} = \exp(\{\ln(b - a)\}) = b - a$$

Total UQ for Computer Models

- **Parametric uncertainty** – refers to uncertain inputs or parameters of a model
- **Parametric variability** – uncontrolled or unspecified conditions in inputs or parameters
- **Model discrepancy** – no model is perfect…
- **Code uncertainty** – uncertainty associated with not knowing the output of a computer model given any particular input configuration until the code is run

Model Discrepancy

- We <u>must</u> quantify model discrepancy
- From Kennedy and O'Hagan, 2001: "No model is perfect. Even if there is no parameter uncertainty, so that we know the true values of all the inputs required to make a particular prediction of the process being modeled, the predicted value will not equal the true value of the process. The discrepancy is model inadequacy."

$$\boxed{\text{Model Fidelity} \propto (\text{Model Discrepancy})^{-1}}$$

$$\overset{\text{Quantity of interest}}{q} = \overset{\text{Simulator}}{g(\mathbf{x})} + \overset{\text{Model Discrepancy}}{\epsilon(\mathbf{x})}$$

Code Uncertainty

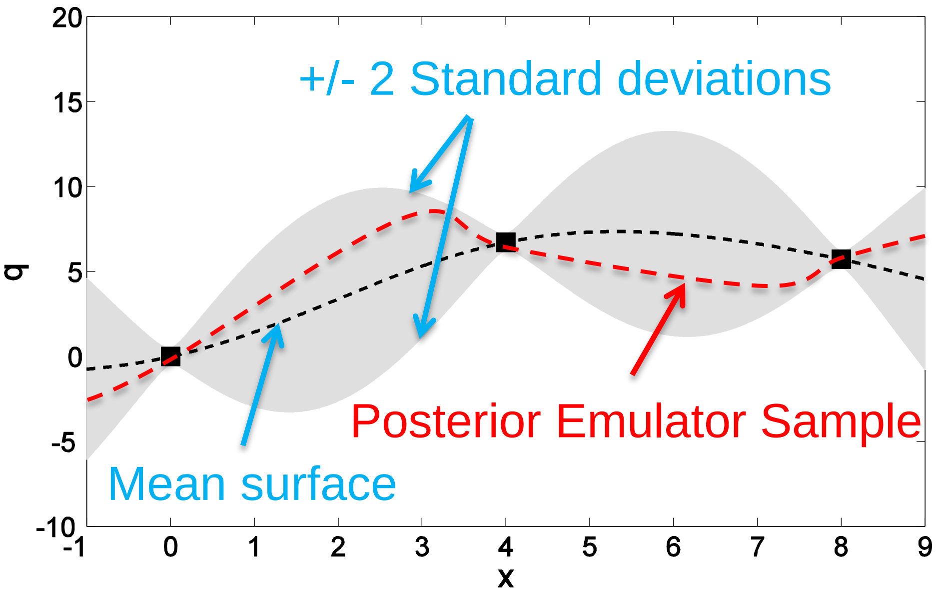
- Code uncertainty – uncertainty associated with not knowing the output of a computer model given any particular configuration until the code is run

Quantity of interest

Model Discrepancy

Emulator

$$q = \mathcal{G}(\mathbf{x}) + \epsilon(\mathbf{x})$$



<u>Gaussian Process Emulator</u>

$$\mathcal{G}(\mathbf{x}) \sim \mathcal{GP}(\mathrm{m}(\mathbf{x}), \mathrm{k}(\mathbf{x}, \mathbf{x}'))$$

$\mathrm{m}(\mathbf{x})$ is a mean function

$\mathrm{k}(\mathbf{x}, \mathbf{x}')$ is a covariance kernel

Complexity Metric Estimation

Must incorporate all sources of uncertainty

$f_Q(q)$ [Quantity of interest density]

$$h(Q^{\Delta}) = -\sum_{j=1}^{N} [\, f_Q(q^j)\Delta] \log[\, f_Q(q^j)\Delta + \log \Delta$$ [Entropy Estimate]

$$\hat{C}(Q) = \exp\{h(Q^{\Delta})\}$$ [Complexity estimate]

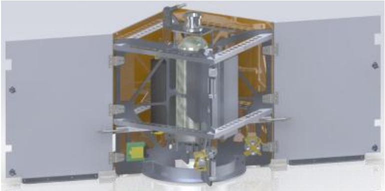
$$\hat{C}(Q \mid \mathcal{G} = G)$$ [Complexity estimate conditioned on emulator sample]

$$\tilde{C}(Q) = \max_{G}(\hat{C}(Q \mid \mathcal{G} = G))$$ [Complexity estimator]

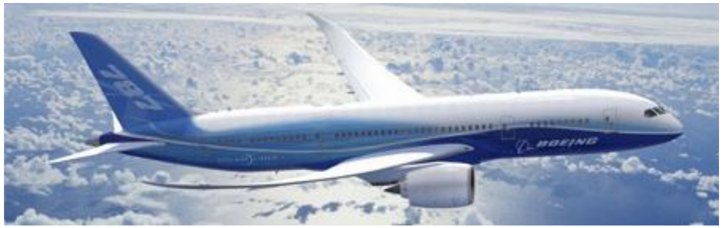
Identification of Key Contributors to Complexity

Complex System Design

Source: www.inetres.com

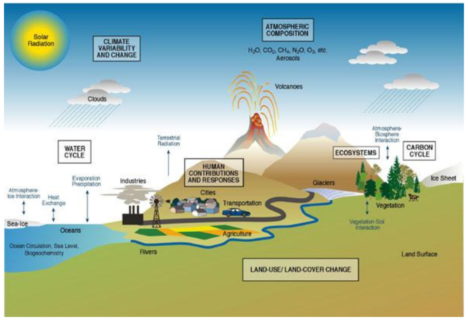


Cathode **A**node **S**pace **T**hruster for **O**rbital **R**epositioning Satellite



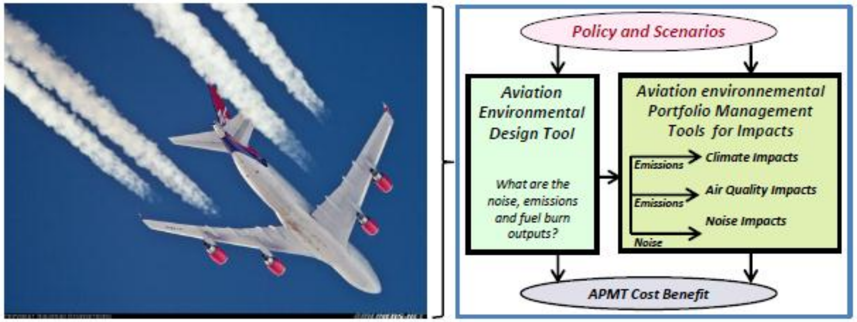
Source: www.boeing.com.

Complex System Analysis



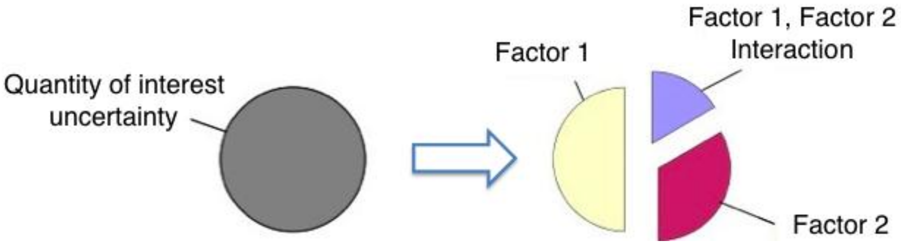
Source: blog.cunysustainablecities.org.

Global climate change



Source: www.airliners.net

Coupled Aviation-Environmental System



Variance-based approach (Homma 1996)

$$\operatorname{var}(Q) = \operatorname{var}(\mathbb{E}[Q \mid X_i]) + \mathbb{E}[\operatorname{var}(Q \mid X_i)]$$

$$S_i = \frac{\operatorname{var}(\mathbb{E}[Q \mid X_i])}{\operatorname{var}(Q)} = \frac{\operatorname{var}(Q) - \mathbb{E}[\operatorname{var}(Q \mid X_i)]}{\operatorname{var}(Q)}$$

Sensitivity Indices

For parametric uncertainty and variability

$$\hat{\eta}_i(\mathcal{G}) = \frac{\overbrace{\hat{C}(Q\,|\,\mathcal{G}=G)}^{\text{Initial complexity conditioned on emulator sample}} - \overbrace{\mathbb{E}_{X_i}[\hat{C}(Q\,|\,\mathcal{G}=G, X_i = x_i^j)]}^{\text{Expected complexity remaining once factor i is known}}}{\hat{C}(Q\,|\,\mathcal{G}=G)}$$

Average over the emulator samples to obtain sensitivity indices

$$\boxed{\tilde{\eta}_i = \mathbb{E}_{\mathcal{G}}[\hat{\eta}_i(\mathcal{G})]}$$

Sensitivity Indices

For model discrepancy

Recall:

$$\tilde{C}(Q) = \max_{G}(\hat{C}(Q \mid \mathcal{G} = G))$$

$$\tilde{\eta}_{\mathrm{MD}} = \frac{\overbrace{\tilde{C}(Q)}^{\text{Initial complexity}} - \overbrace{\tilde{C}(Q \mid \epsilon(\mathbf{x}) = 0)}^{\text{Complexity remaining if the model is assumed to be perfect}}}{\tilde{C}(Q)}$$

For code uncertainty

$$\tilde{\eta}_{\mathrm{CU}} = \frac{\overbrace{\tilde{C}(Q)}^{\text{Initial complexity}} - \overbrace{\mathbb{E}_{\mathcal{G}}[\hat{C}(Q \mid \mathcal{G} = G)]}^{\text{Expected complexity remaining if simulator was known at every point}}}{\tilde{C}(Q)}$$

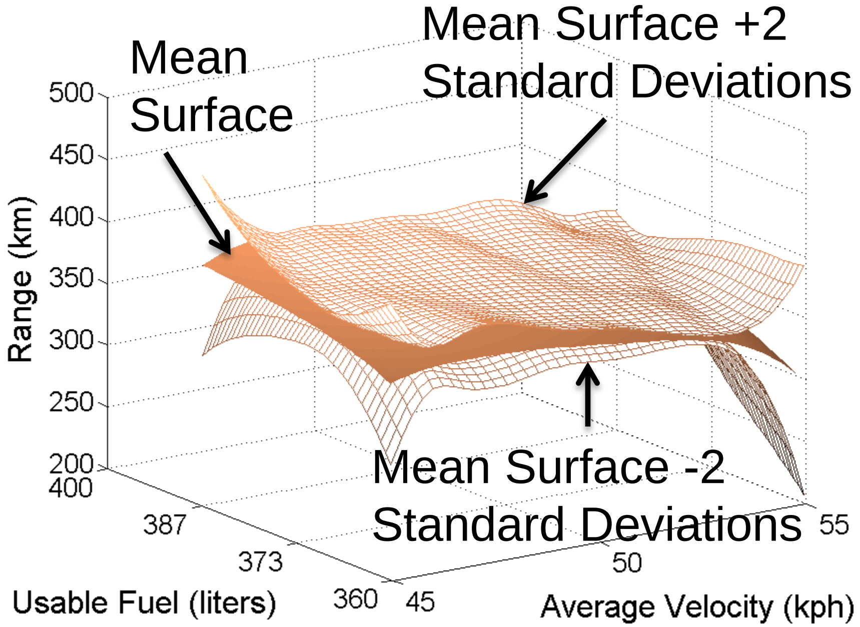
Demonstration 1

- Estimate the complexity of an IFV design with respect to **range** as the quantity of interest
- Parameters:
  - – Usable fuel ~ U[360,400] liters (parametric uncertainty)
  - – Average velocity ~ U[45,55] kph (parametric variability)
- Model discrepancy

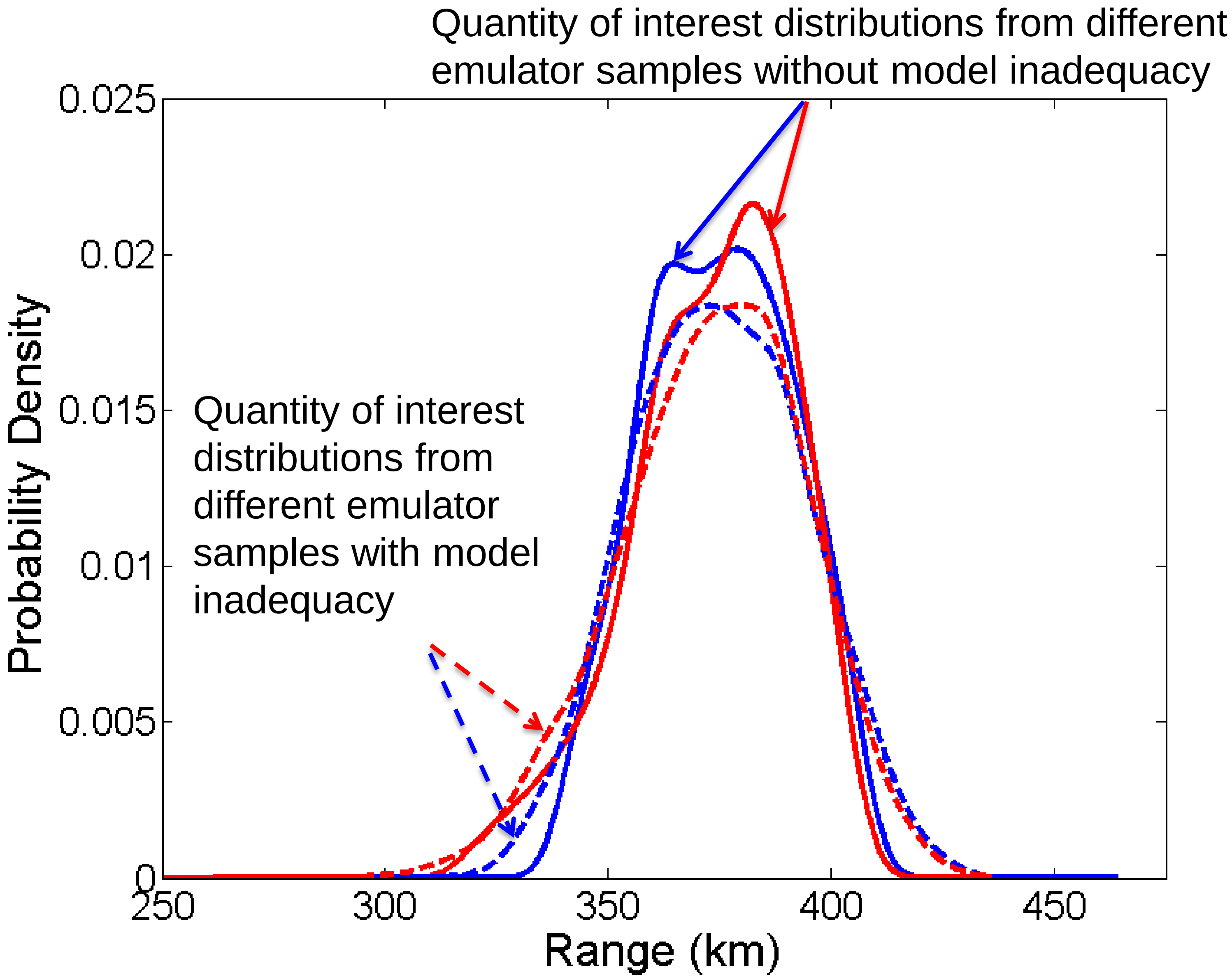
$$q = g(\mathbf{x}) + \epsilon(\mathbf{x})$$
$$\epsilon(\mathbf{x}) \sim N(0, 100\text{km}^2)$$

- Code uncertainty

Gaussian process emulator



Results – Quantity of Interest Densities



Quantity of interest distributions from different emulator samples without model inadequacy

Quantity of interest distributions from different emulator samples with model inadequacy

Results

Complexity	$\tilde{C}(Q) = 104\ \mathrm{km}$

Average velocity sensitivity	$\tilde{\eta}_{\mathrm{AV}} = 0.46$
Usable fuel sensitivity	$\tilde{\eta}_{\mathrm{F}} = 0.44$
Model discrepancy sensitivity	$\tilde{\eta}_{\mathrm{MD}} = 0.15$
Code uncertainty	$\tilde{\eta}_{\mathbf{CU}} = 0.16$

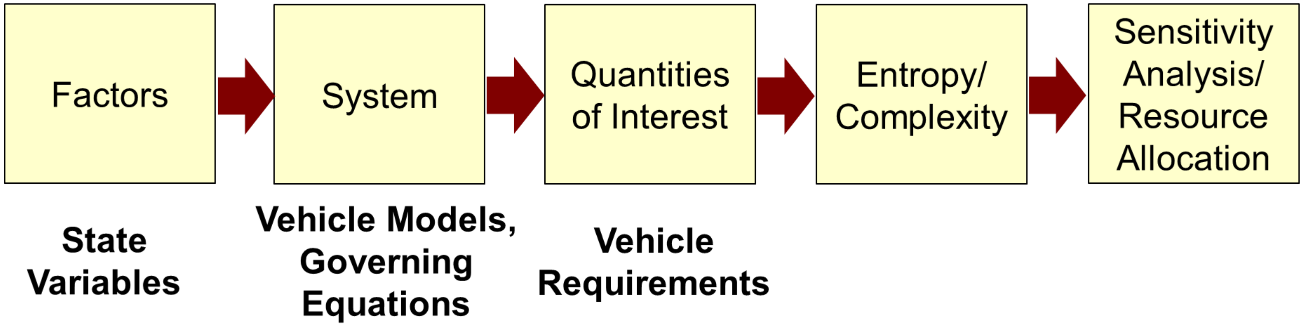
Allocate resources to learning more about the target velocity and fuel level

Demonstration 2

- Notional design process for a hybrid IFV
- Purpose is to demonstrate at a high level the role of sensitivity analysis and feedback
- Primarily the work of John Deyst and Chelsea He

Approach

- Specify the quantities of interest → **vehicle requirements**
- Specify the factors that influence the quantities of interest → **state variables**
- Decompose the vehicle design in terms of systems, subsystems, and components, and identify linking variables
- Use available vehicle design models to compute the quantities of interest
- Compute the probability of failure with respect to the quantities of interest
- Estimate complexity in the quantities of interest using Monte Carlo simulation
- Perform sensitivity analysis to identify sources of complexity
- Perform resource allocation to reduce complexity
- Iterate until feasible design is achieved that exhibits acceptable complexity in the quantities of interest



Factors

System

Quantities of Interest

Entropy/ Complexity

Sensitivity Analysis/ Resource Allocation

State Variables

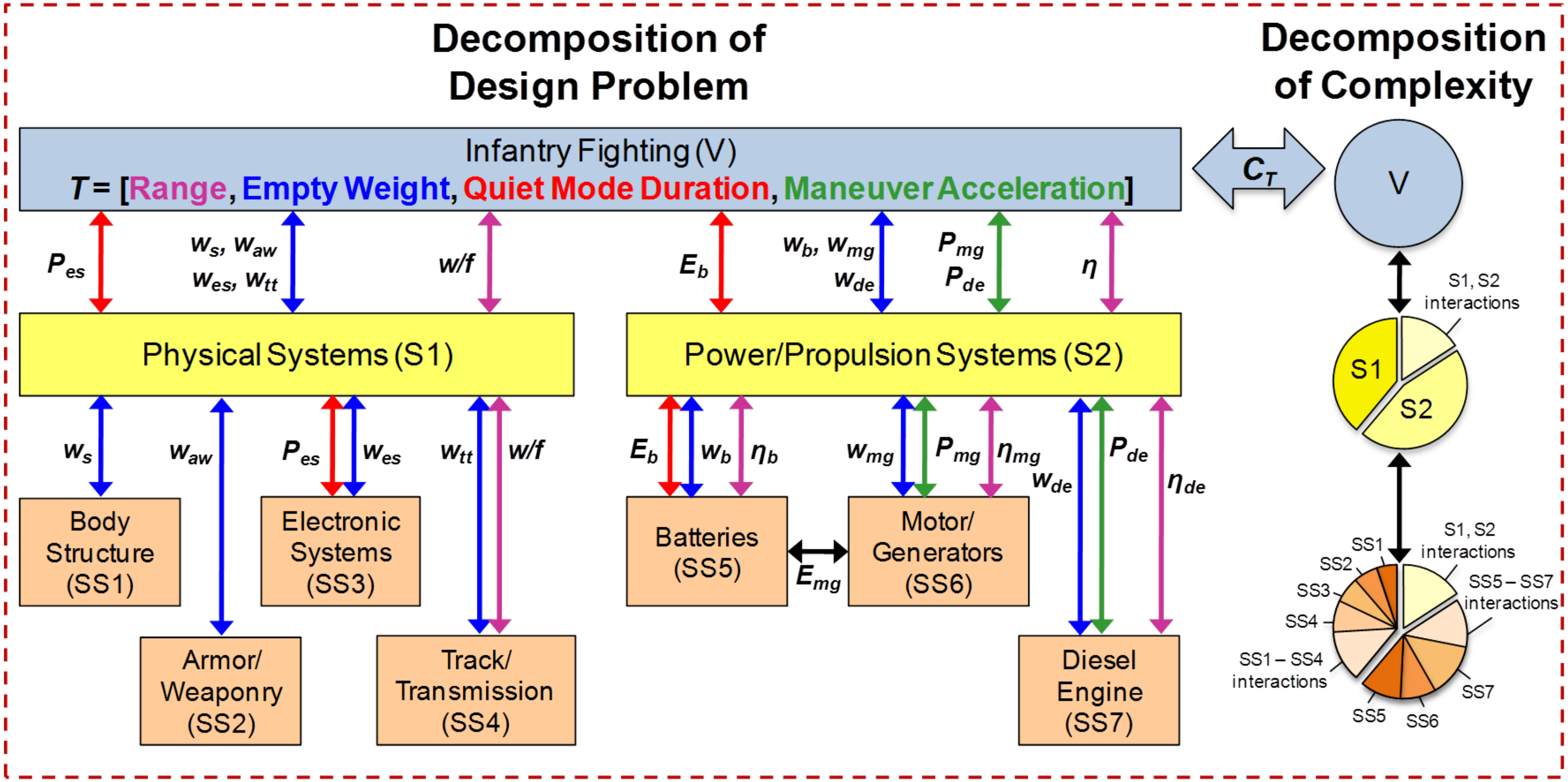
Vehicle Models, Governing Equations

Vehicle Requirements

Requirements and System Decomposition

Vehicle Requirements

- The hybrid IFV must have a range of at least 500 kilometers.
- The empty weight of the hybrid IFV must not exceed 25,000 kilograms.
- The hybrid IFV must operate in quiet mode for at least 8 hours.
- The hybrid IFV must achieve a maneuver acceleration of 0 to 10 m/s in 5 sec.



Iterate until feasible design achieved

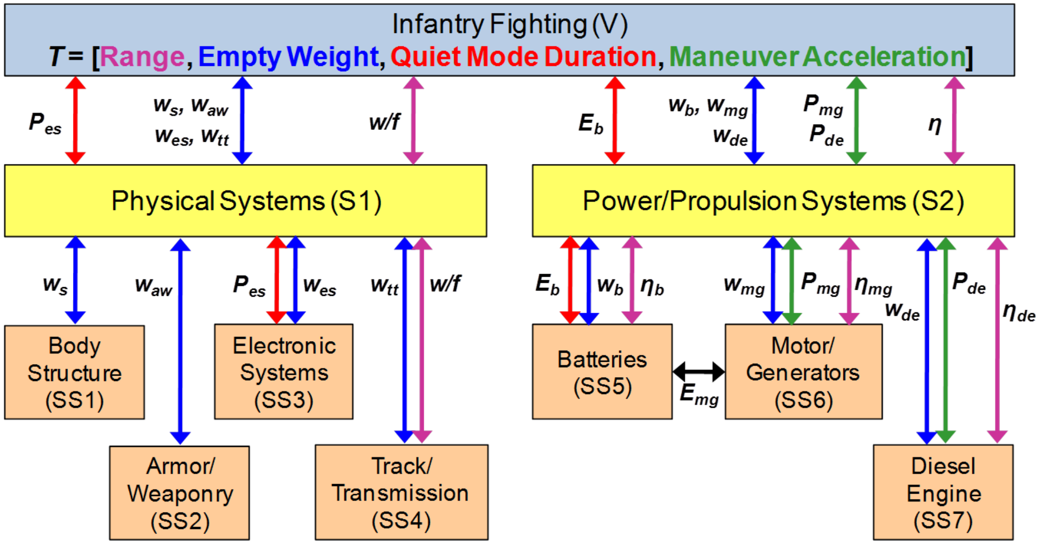
System Factors

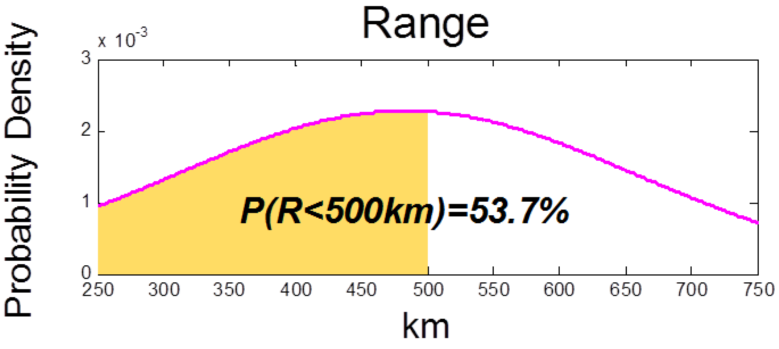
State Variable			Initial Value	Std. Dev.
1.	w/f	Weight to thrust ratio [–]	15	3
2.	η_{de}	Diesel engine efficiency [–]	0.3	0.06
3.	η_{b}	Battery charging efficiency [–]	0.9	0.18
4.	η_{mg}	Motor/generator efficiency [–]	0.8	0.16
5.	W_{fuel}	Fuel weight [kg]	400	80
6.	W_{p}	Payload weight [kg]	500	100
7.	W_{e}	Empty weight [kg]	24,000	4,800
8.	P_{es}	Quiet mode power [kW]	10	2
9.	E_{b}	Battery energy capacity [kWh]	80	16
10.	P_{de}	Diesel engine power [kW]	200	40
11.	P_{mg}	Motor/generator power [kW]	275	55

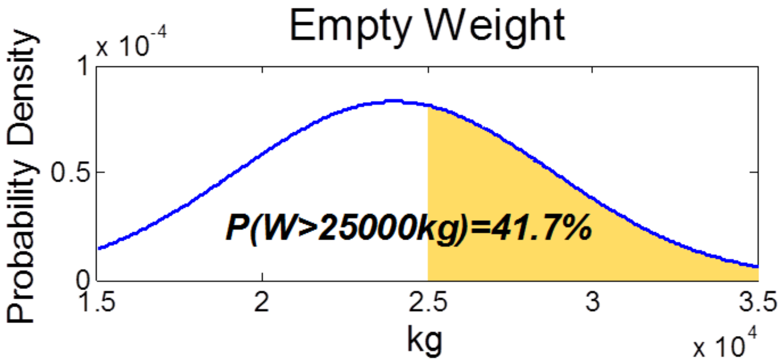
System Quantities of Interest

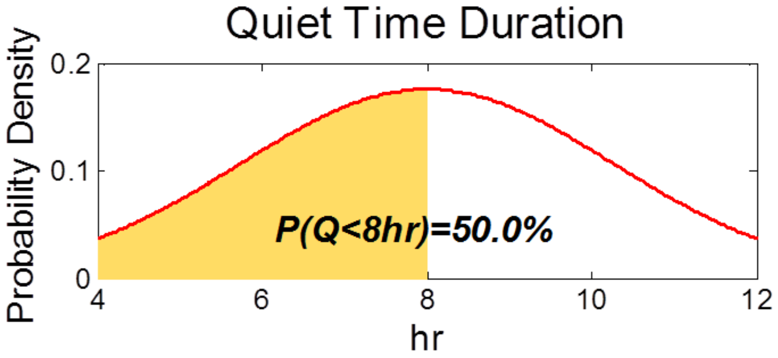
Quantity of Interest (QOI)		
1.	**R**	Range [km]
2.	**W**	Empty weight [kg]
3.	**Q**	Quiet time duration [hr]
4.	**A**	Maneuver Acceleration [# pulses]

Decomposition of Design Problem



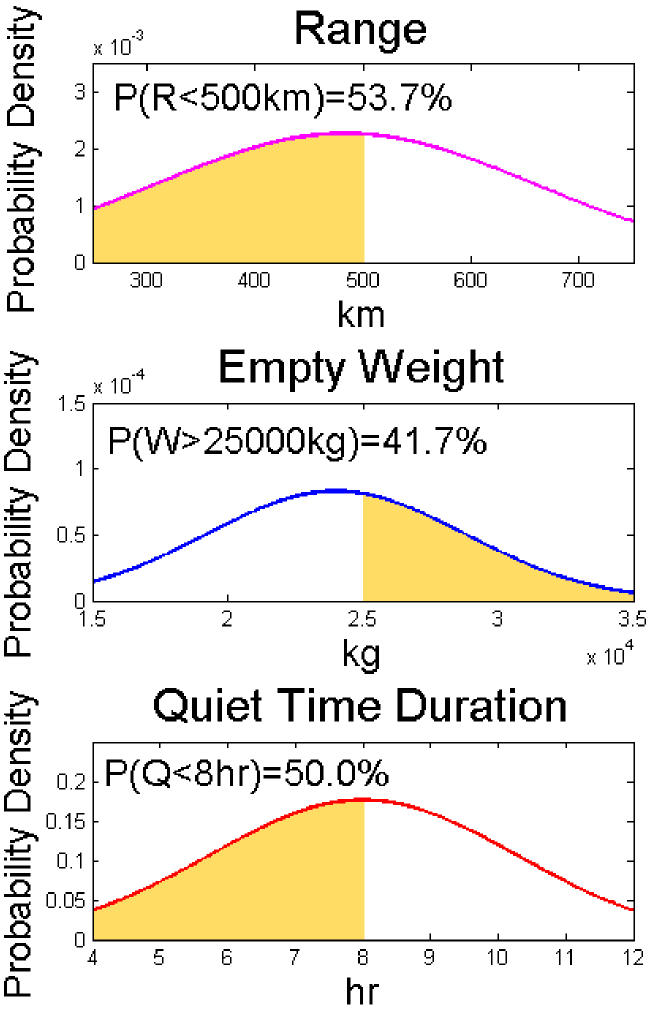


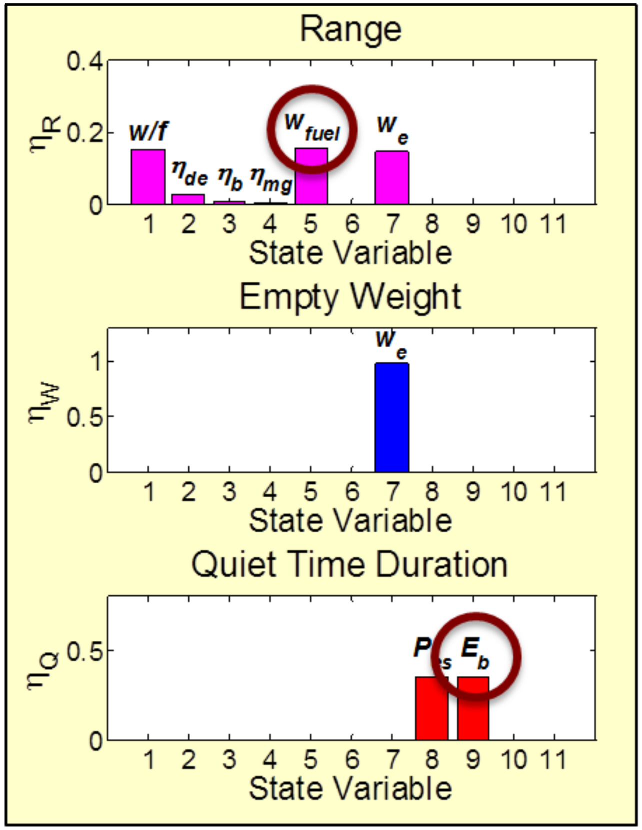


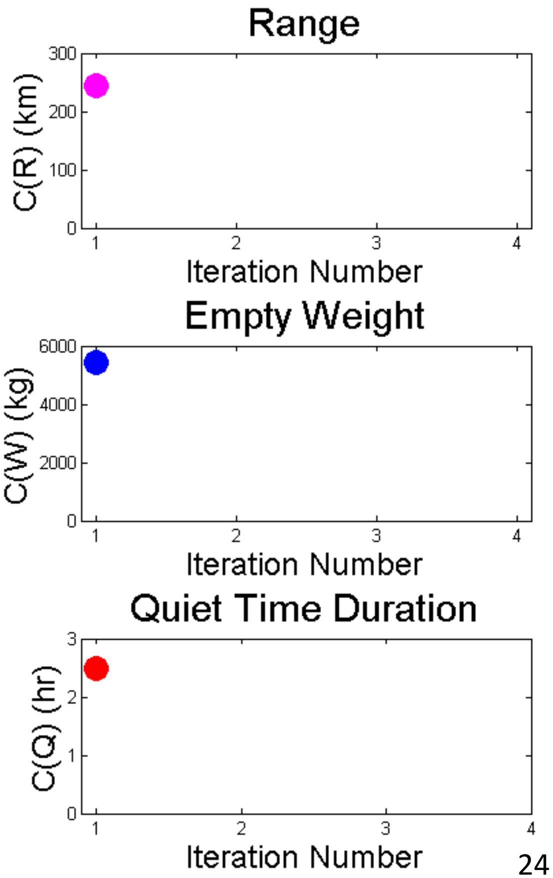


Iteration 1

- Initialize values and standard deviations
- Identify W_{fuel} and E_b as targets for resource allocation

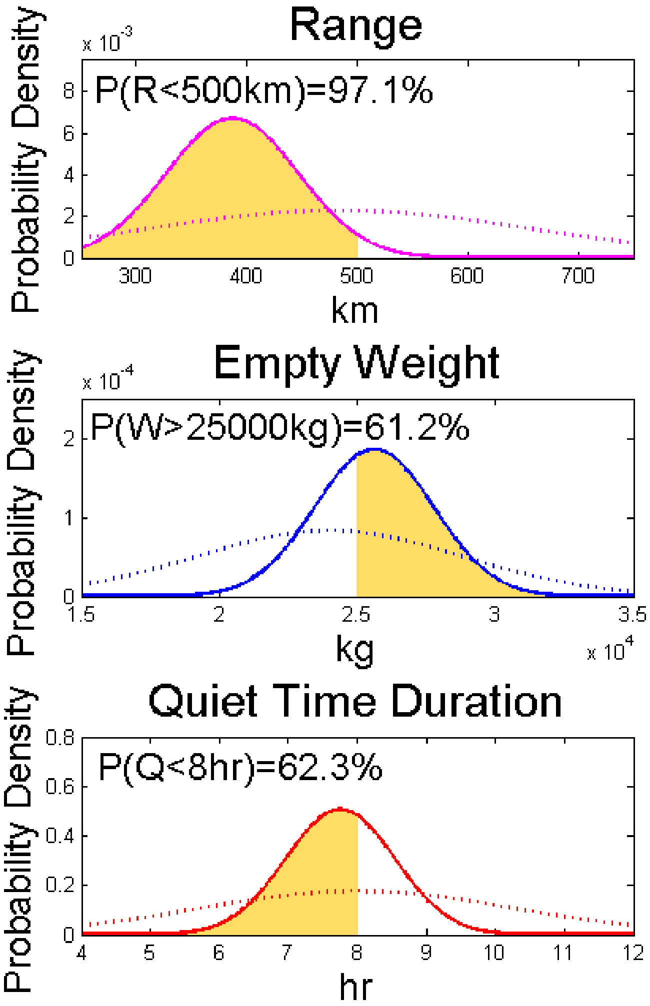


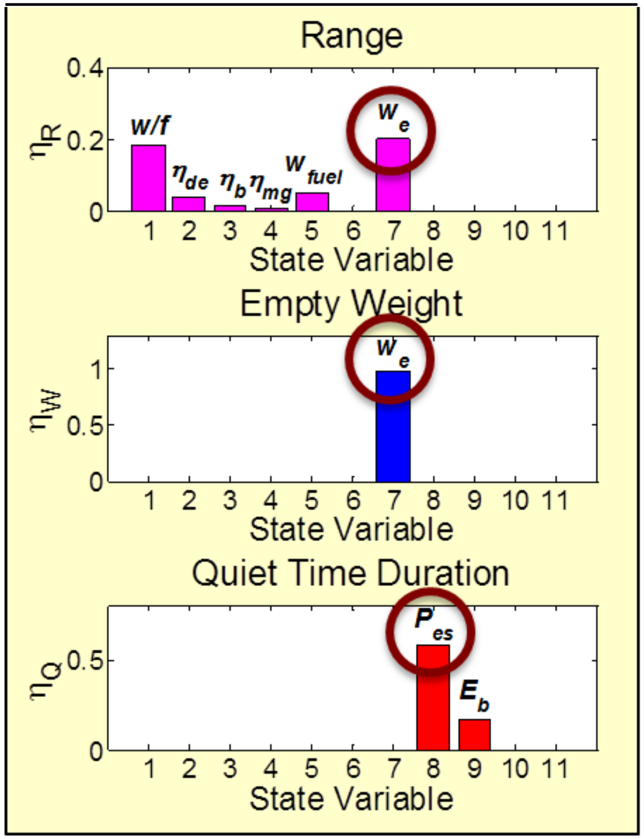


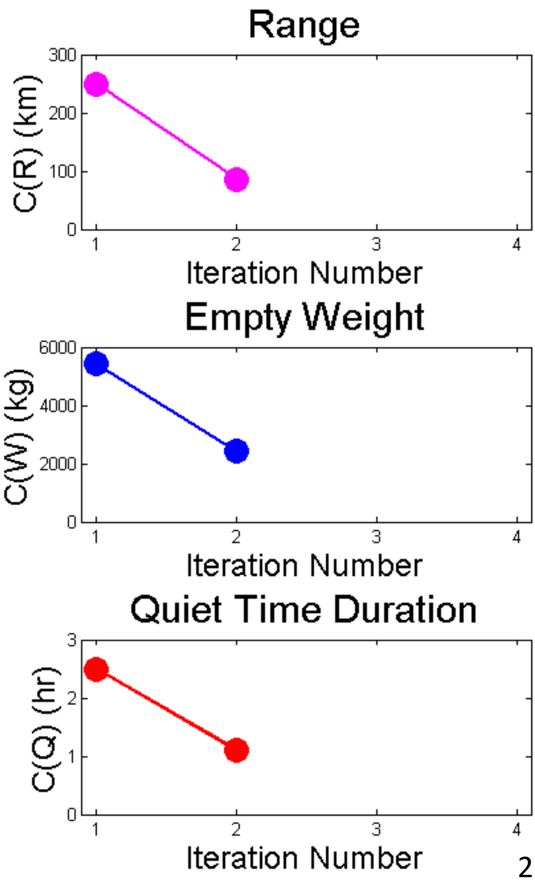


Iteration 2

- Reduce std. dev. of W_{fuel} and E_b by 75%; all others by 50%
- Identify W_e and P_{es} as targets for resource allocation

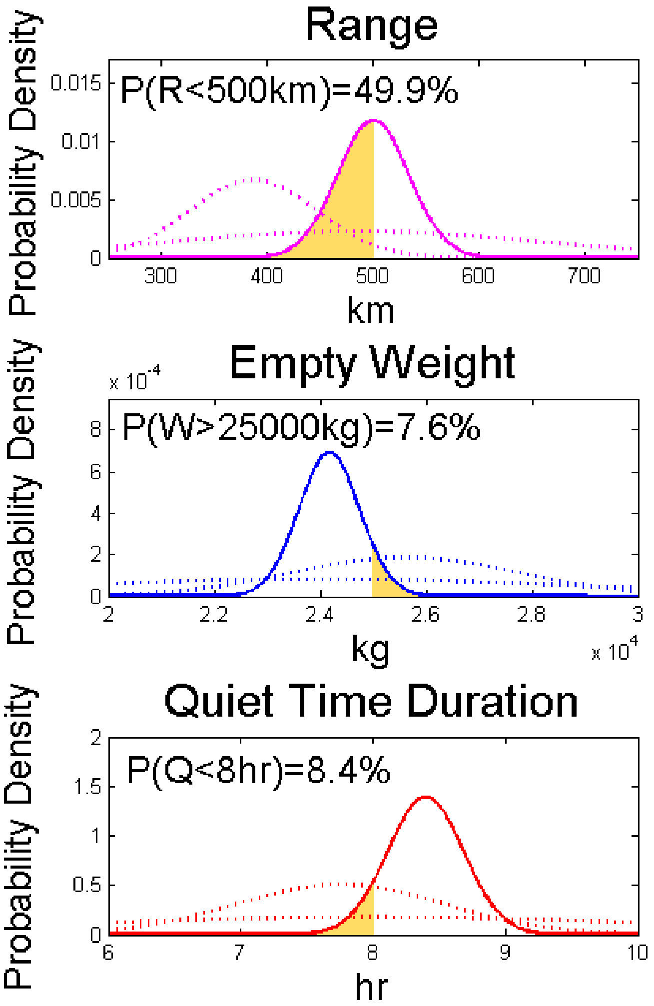


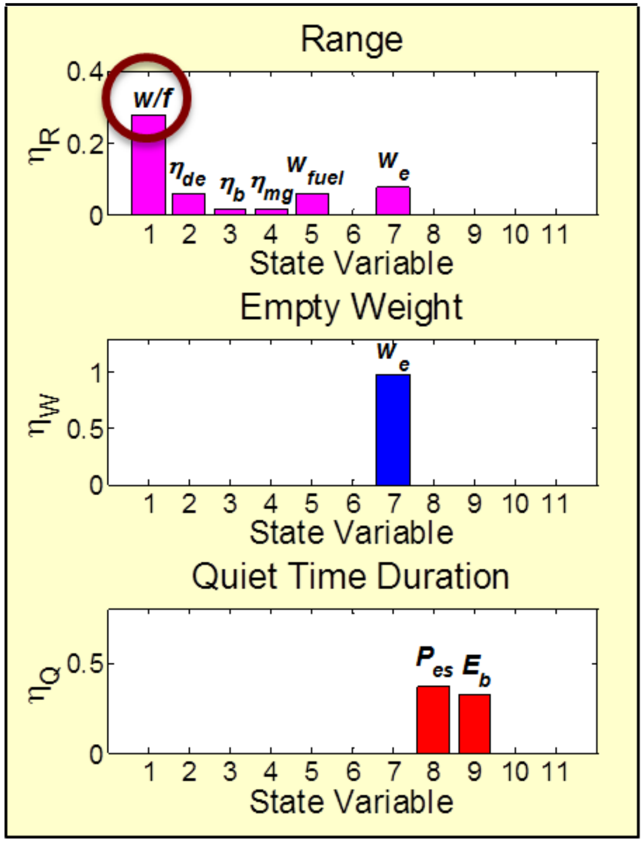


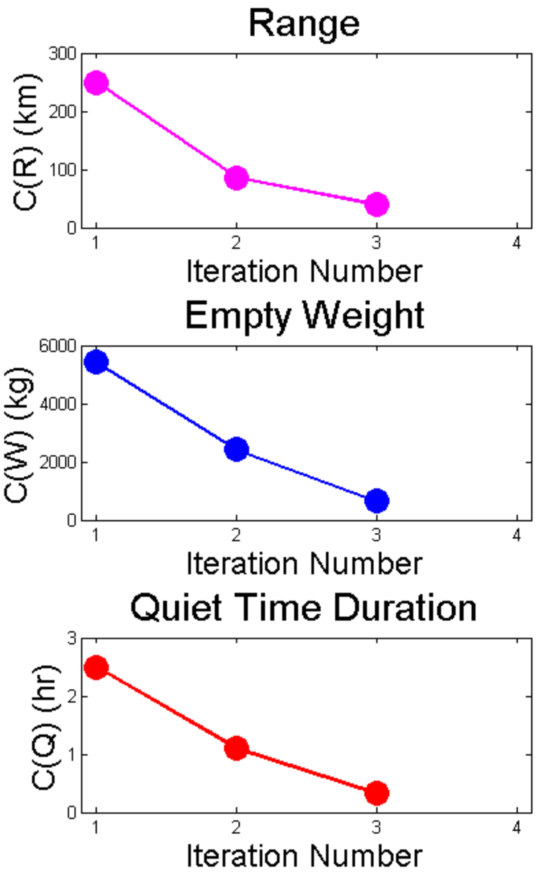


Iteration 3

- Reduce std. dev. of W_e and P_{es} by 75%; all others by 50%
- Identify w/f for resource allocation

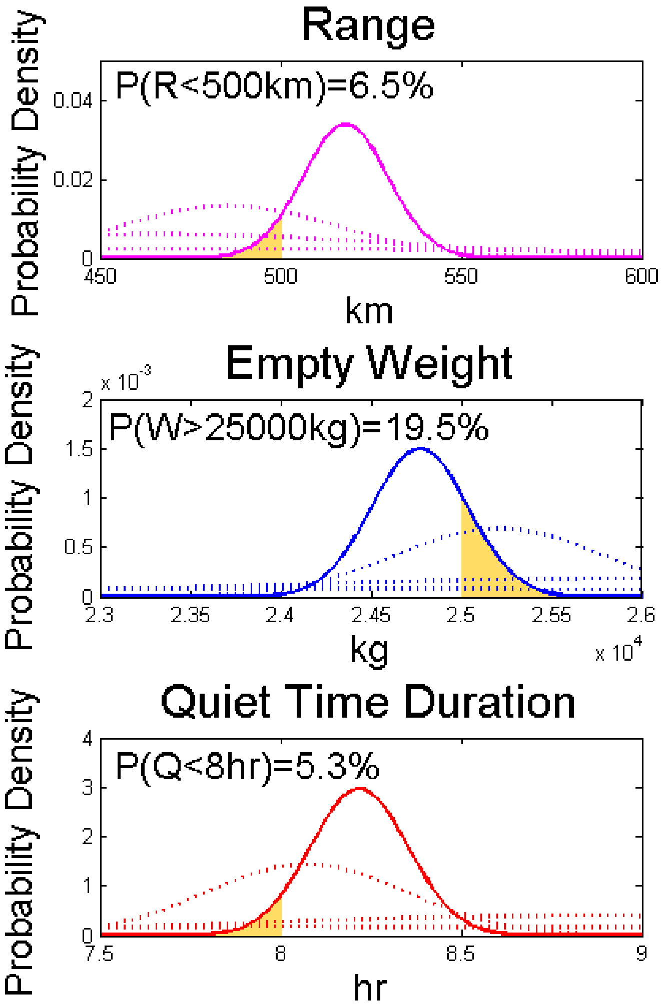


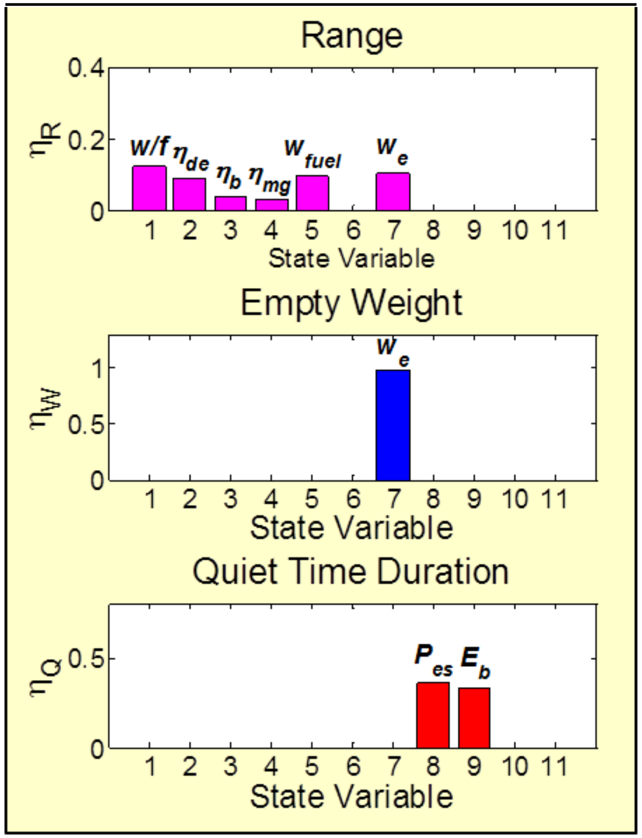


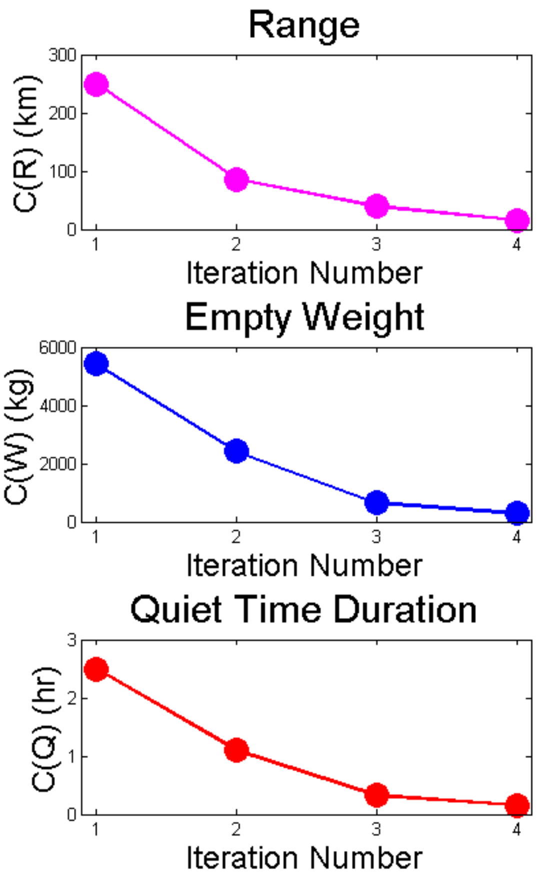


Iteration 4

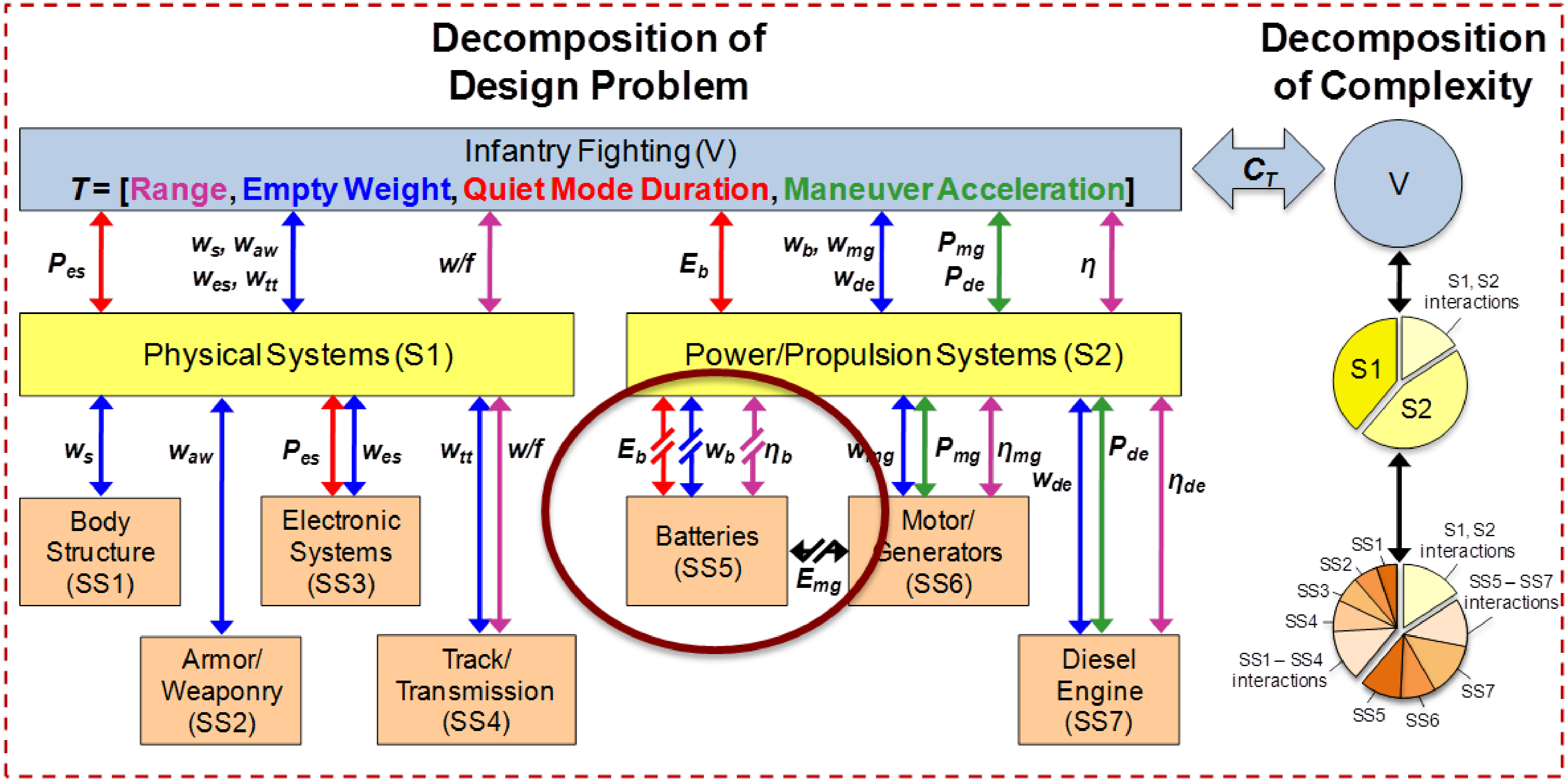
- Reduce std. dev. of w/f by 75%; all others by 50%







The Importance of Feedback



Decomposition of Design Problem

Decomposition of Complexity

Infantry Fighting (V)

T **= [Range, Empty Weight, Quiet Mode Duration, Maneuver Acceleration]**

Physical Systems (S1)

Power/Propulsion Systems (S2)

Body Structure (SS1)

Armor/ Weaponry (SS2)

Electronic Systems (SS3)

Track/ Transmission (SS4)

Batteries (SS5)

Motor/ Generators (SS6)

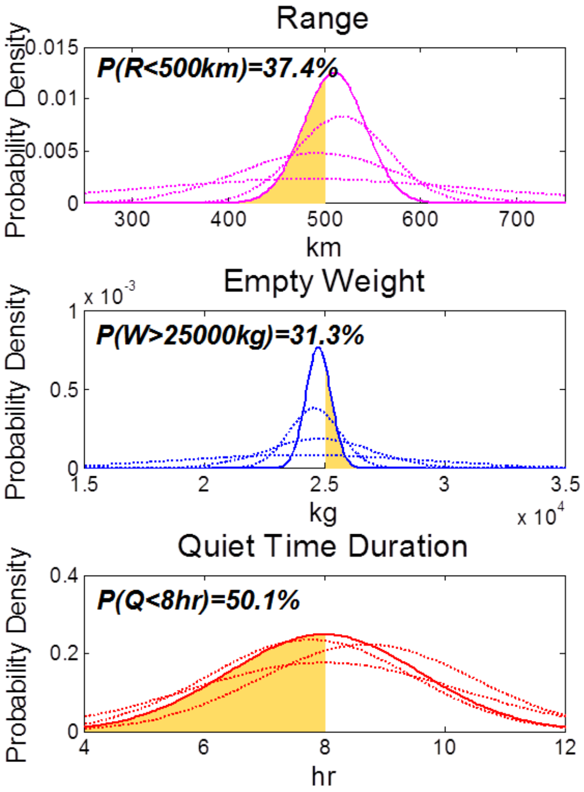
Diesel Engine (SS7)

Iterate until feasible design achieved

Broken Battery Feedback Results

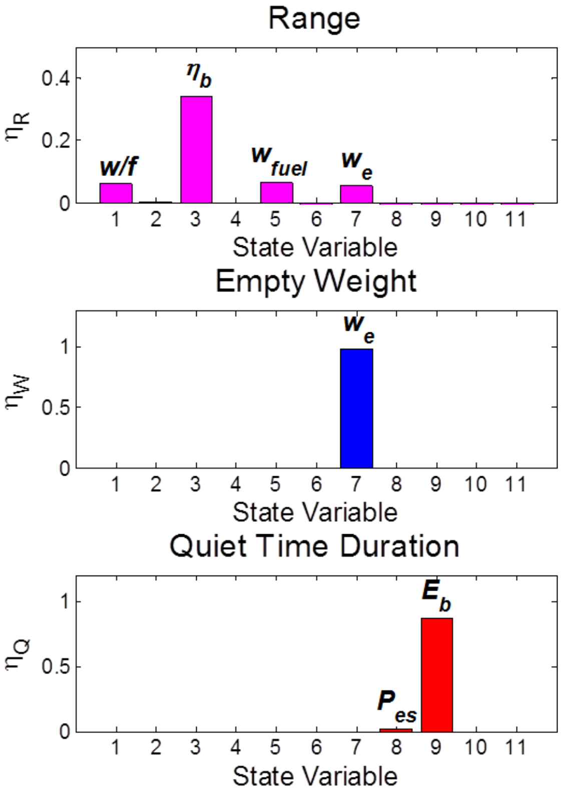
QOI Estimates

Uncertainty in QOI estimates is slower to decrease between iterations



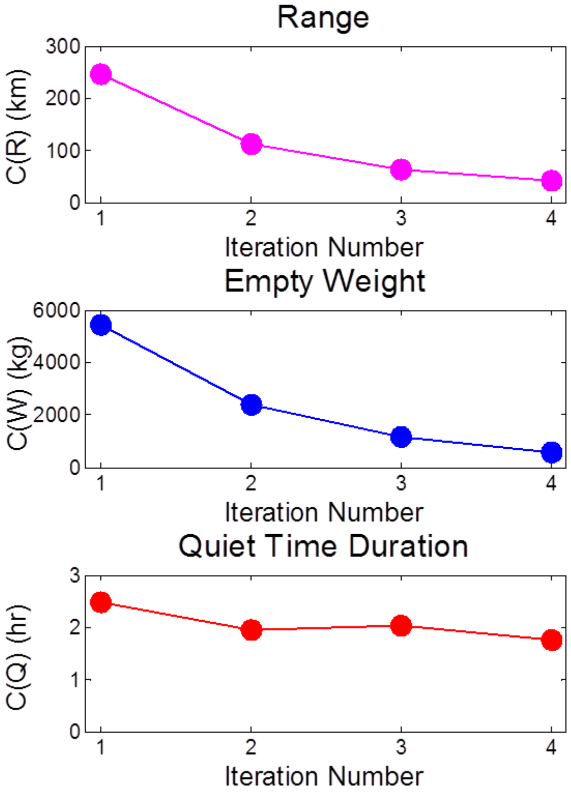
Sensitivity Indices

Battery-related state variables (**η_b**, **E_b**) contribute greatly to complexity in QOI



Complexity Estimates

Complexity in QOI estimates is slower to converge



Relevance of the Example

Source: www.boeing.com.

Source: blog.seattlepi.com.

Conclusions

- Summary
  - Proposed an information-theoretic metric of complexity
  - Developed a set of sensitivity indices as indicators of key sources of complexity
  - Calculated the metric of complexity and apportioned the complexity to key sources for an IFV application
  - Demonstrated the importance of feedback in design
- Conclusions
  - For simulation-based design and analysis, all sources of uncertainty must be included
  - Data regarding model discrepancy is critical
  - Sensitivity analysis can be used to allocate resources aimed at reducing large uncertainties in quantities of interest
  - The quantification and evolution of information in system design is essential.
  - System design and analysis is a problem of information management / uncertainty control
- Future work
  - Information fusion (models, sensors, experts…)
  - Compositional UQ

Bringing high fidelity forward

"The most important development in aviation in 2011" -Time

- Aerospace vehicle design typically involves custom parts for nearly every aspect of the system
- Design options include:
  - Use high fidelity tools to analyze
  - Start with low fidelity tools and identify where fidelity increases are required
  - Deal with emergent behavior as it emerges

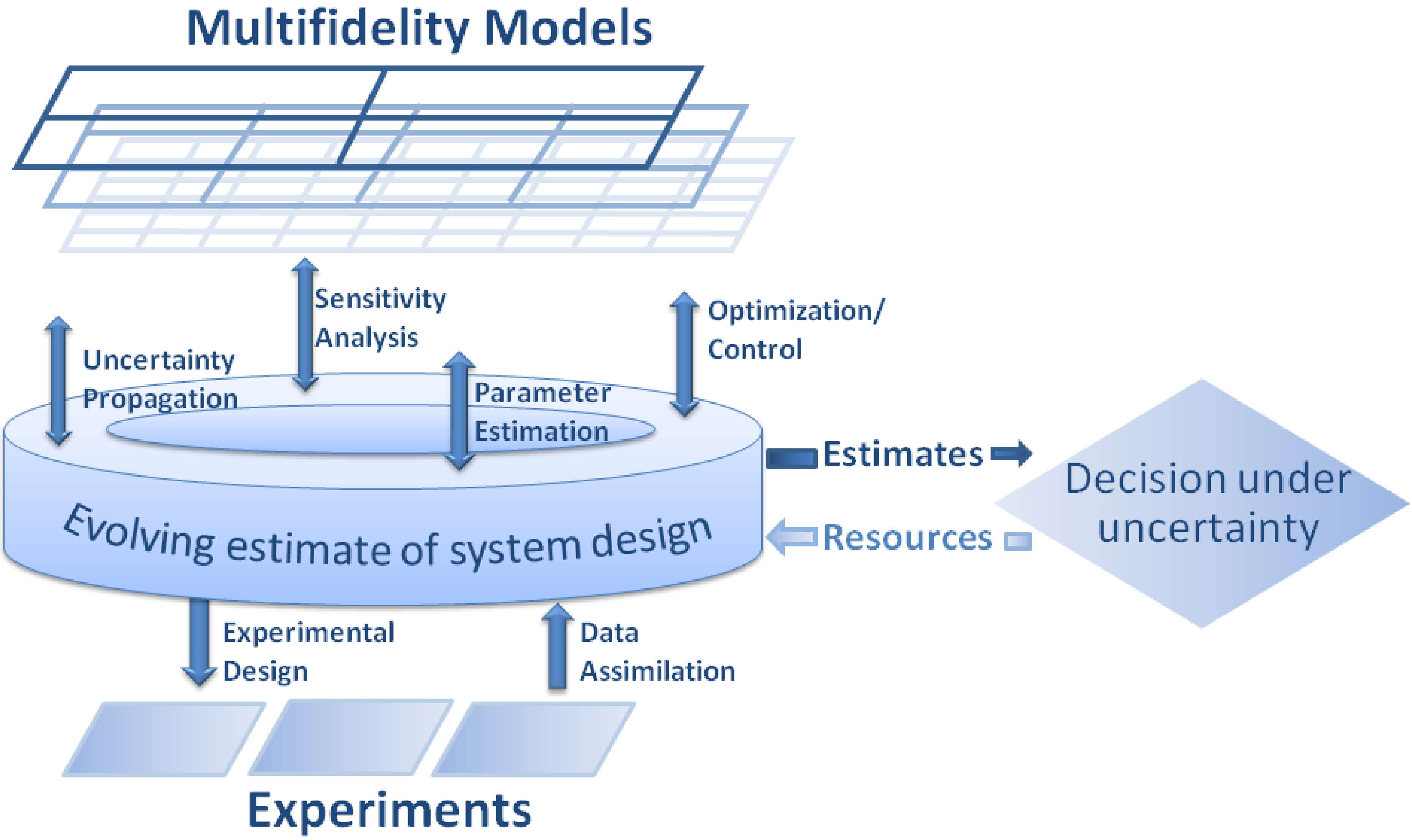
Source: www.time-az.com.

- Reuse of parts/components enables high fidelity results from "low" fidelity tools
  - Sacrifice optimality for reduced complexity designs
  - Possibly at lower cost and faster development times
- Recall visualization discussion
  - Visualizing high dimensional design parameter spaces is difficult
  - Lots of room for possibly undetected emergent behavior
  - Foundry-like approach can reduce the design space substantially

References

- (Weaver 1948) Weaver, W., “Science and Complexity,” *American Scientist*, Vol. 36, No. 536, 1948.
- (Johnson 1997) Johnson, G., “Researchers on Complexity Ponder What It’s All About,” *The New York Times: Technology*, May 6, 1997.
- (Kennedy 2001) Kennedy, M. and O’Hagan, A., “Bayesian Calibration of Computer Models,” *Journal of the Royal Statistical Society Series B*, Vol. 63, No. 3, 2001.
- (Campbell 1966) Campbell, L., “Exponential Entropy as a Measure of Extent of a Distribution,” *Z. Wahrscheinlichkeitstheorie verw. Geb.*, Vol. 5, 1966.
- (Griffin 1997) Griffin, A., “The Effect of Project and Process Characteristics on Product Development Cycle Time,” *Journal of Marketing Research*, Vol. 34, No. 1, 1997.
- (Kolmogorov 1965) Kolmogorov, A., “Three Approaches to the Quantitative Definition of Information,” *Problems of Information Transmission*, Vol. 1, No. 1, 1965.
- (Chaitin 1969) Chaitin, G., “On the Simplicity and Speed of Programs for Computing Infinite Sets of Natural Numbers,” *Journal of the ACM*, Vol. 16, No. 3, July1969.
- (Lloyd 1988) Lloyd, S. and Pagels, H., “Complexity as Thermodynamic Depth,” *Annals of Physics*, Vol. 188, 1988.
- (Homma 1996) Homma, T. and Saltelli, A., “Importance Measures in Global Sensitivity Analysis of Nonlinear Models,” Reliability Engineering and System Safety, Vol. 52, 1996.

Big Picture


Multifidelity Models
Sensitivity Analysis
Uncertainty Propagation
Parameter Estimation
Optimization/ Control
Evolving estimate of system design
Estimates
Resources
Decision under uncertainty
Experimental Design
Data Assimilation
Experiments