Reflections on Statistical Engineering and Its Application

Geoff Vining

VirginiaTech
Invent the Future

Outline

- Overview of Statistical Engineering
- Initial Thoughts on “Building Blocks”
- Initial Thoughts on “Statistical Engineering Principles”
- NASA Example
- Impact on the DoD, NASA, and the National Labs

VirginiaTech
Invent the Future

Acknowledgement: My Co-Conspirators

- Roger Hoerl
- Ron Snee
- Pete Parker

Virginia Tech
Invent the Future

Overview of Statistical Engineering

- Future focus: Large, unstructured, complex problems!
- Solutions require collaboration among high profile interdisciplinary teams!
- Problems cut across the organization

Overview of Statistical Engineering

- Building upon Six Sigma
  - Good strategic structure
  - Need for something tactical in between
  - How do we deploy our tools?
- Success requires new tools and mindset
- Need to ask how we can generalize solution tactics to solve future problems

Overview of Statistical Engineering

- One pathway: Statistical Engineering
- Goal: Develop appropriate theory
  - to apply known statistical principles and tools
  - to solve high impact problems
  - for the benefit of humanity.
- Minimize “one-off” solutions

Overview of Statistical Engineering

- The heart of Statistical Engineering is the scientific method.
- Most theories underlying statistical engineering involve strategic application of the scientific method.
  - Deming-Shewhart PDCA (Plan, Do, Check, Act)
  - DMAIC (Define, Measure, Analyze, Improve, Control)

VirginiaTech
Invent the Future

Initial Thoughts on the Building Blocks

- The Scientific Method Is a Fundamental Approach for Discovery and Problem Solving
- Statistical Thinking Is Essential Developing Solutions
- Success Requires Teams that Function Well
  - Subject Matter Expertise
  - Statistical/Analytical Expertise
- “All Models Are Wrong; Some Are Useful”

Initial Thoughts on the Building Blocks

- Probability Is the Basic Language for Quantifying Uncertainty
- All Probability Statements Are Subjective, Depending on Critical Assumptions (Beliefs!)
- Statistical Methods Must Be as Robust as Possible to Assumptions and Models
- Other People Must Be Able to Duplicate Results

VirginiaTech
Invent the Future

SE Principle 1

- Proper Data Collection, Analysis, and Interpretation Are Essential for the Scientific Method
  - Dependence on the proper question of interest
  - Impact of restrictions on data collection
  - Proper consideration of constraints on factors/regressors
  - Must avoid error of the third kind!
  - In early phases, data include expert opinion.

VirginiaTech
Invent the Future

SE Principle 2

- All Data Collection, Especially Experimentation, Must Be Sequential
  - Iterative procedure
    - Adaptive
    - Able to mitigate problems
  - Each phase targets different questions
  - Final Phase: Data must dominate opinion to extent possible.

VirginiaTech
Invent the Future

SE Principle 3

- All Data Collection Must Recognize Sources of Variability
  - Local control of error (blocking, co-variates)
  - Basis to minimize biases, understand true precision
  - Understanding sources necessary for variation reduction
  - More complicated the problem, the more sources of variability!

SE Principle 4

- Approximate Models that Include Uncertainty Are Fundamental to Analysis
  - At least two sources of error:
    - Model: over- or under-specified; linear or non-linear
    - Background noise – Often combination of several sources!
  - Important to understand error propagation, especially as the system becomes more complex

VirginiaTech
Invent the Future

SE Principle 5

- Analyses Require Clear Statements about All Modeling Assumptions
  - Essential for other researchers to duplicate
  - States and justifies the beliefs of the research team
    - Subject matter experts
    - Analysis
  - Essential for both Bayesian and Frequentist Analysis!

SE Principle 6

- All Analyses Require the Proper Use of Data to Assess Assumptions
  - Residual analysis typically essential
    - Raw residuals never appropriate!
    - Must standardize as closely as possible to appropriate distribution
    - Translate residual to subject matter language/understanding
  - Reserve data for model validation/confirmation

VirginiaTech Invent the Future

SE Principle 7

- Difference between Data Cleansing and Data Manipulation
  - Data cleansing: Identifying and correcting bad data
  - Data manipulation: throwing away data not consistent with assumed model (original beliefs of the research team)
  - Outliers often are the most interesting data points!
  - Cannot discard data without proper assignable cause!

VirginiaTech
Invent the Future

SE Principle 8

- Analyses Must Take into Proper Account the Sources of Variability
  - Informal: Database records for check “interesting” cases
  - Formal:
    - Blocking
    - Variance component estimation
    - Including covariates in formal model

VirginiaTech
Invent the Future

SE Principle 9

- Complex Systems of Systems
  - Require combination of subject matter expert first principles/physics and statistical/empirical models
  - Outputs from subsystems become inputs to assemblies
  - Proper propagation of error models essential
  - Empirical confirmation of models
    - Generally easier at the simplest subsystems
    - Often, limited opportunities for complex assemblies

VirginiaTech
Invent the Future

SE Principle 9 - Continued

- Belief Networks Can Provide Basis to Combine Information from Subsystems into Assemblies
  - Combination of subject matter opinion and frequentist model outputs
  - Formal Bayesian with strong prior distributions
  - Require clearly stated and vetted assumptions
  - Empirical confirmation highly desired but impossible in certain cases
  - Common limitation: focus on probability of an event (0/1 data)

VirginiaTech
Invent the Future

SE Principle 10

- Interactions Often Are More Important than Main Effects
  - Operational-Developmental Testing
  - Insights from Robust Parameter Design
    - System robustness to environmental conditions
    - Proper mitigation strategies for operating system

NASA Example - COPVs

- Relatively Small Statistical Engineering Project
- Overarching Question of Interest: Reliability of COPVs at Use Conditions for Expected Life of Mission
- Issues:
  - Many different types of COPVs used in spacecraft
  - Vessel tests are very expensive: money and time
- NASA Engineering Safety Center (NESC) Project

VirginiaTech
Invent the Future

COPVs

- The Core NESC Analytics Team:
  - Reliability Engineers:
    - JPL
    - Langley Research Center
    - Glenn Research Center
  - Statisticians:
    - Marshall Space Flight Center
    - Virginia Tech

VirginiaTech
Invent the Future

COPVs

- NASA Team’s Approach: Focus on Strands Used to Wrap Vessels
  - Less expensive
  - Can have many more experimental units than for vessels
- Still Issue with Time to Test
- Problem: How Do Strands Predict Vessel Behavior?

VirginiaTech
Invent the Future

COPVs

- Initial Study: Previous Strand and Vessel Tests
  - Relevant strand study conducted at a national lab:
    - 57 strands at high loads for 10 years
    - Net information learned: Strands either fail very early or last more than 10 years
  - Vessel studies:
    - Also 10 years
    - Weibull model parameters seem similar to strand studies

VirginiaTech
Invent the Future

COPVs

- Team's Initial Concept
  - Much larger study
  - Censor very early
    - Reduces time
    - Allows the larger study in a practical amount of time
  - Proceed in phases
  - Have detailed data records to track any problems

VirginiaTech
Invent the Future

COPVs

- Phase A: Conducted During Shake-Out of Equipment
  - Small study (although bigger than the national lab study!)
  - Statistical goal: Determine if the parameters from the national lab study are valid as the basis for planning the larger study!
  - Note: Phase A gave the team an opportunity to re-plan the larger experiment, if necessary!

COPVs

- Phase B: "Gold Standard" Experiment
  - Planned time required: 1 year
  - Used 4 "blocks" of equal numbers of strands
    - Allowed the team to correct for time effects
    - Allowed the team to mitigate problems, especially early
  - Study assumed the "classic" Weibull model
  - Size of the experiment assured ability to assess model

COPVs

- Total Size of the Database: Huge
  - Kept data from start of specific strand test to failure on the second
  - Kept the last 2 minutes at the .01 second from buffer
  - Buffer allowed team to investigate unusual phenomena at failure
  - Essential for proper data cleansing

VirginiaTech
Invent the Future

COPVs

- Parallel Vessel Study
  - Reasonably large ISS study targeted to end early (< 10 yrs)
  - Opportunity to step up loads to mimic strands
  - Censored but longer censor time than strands

VirginiaTech
Invent the Future

COPVs

- Results to Date:
  - Phase A: Surprisingly similar to national lab study
  - Phase B:
    - Serious problem occurred with the gripping in the first block
    - Serious conversations with possibility of replacing!
    - Other three blocks well behaved and by themselves produced better than the planned precision for the estimates
    - Residual analysis confirmed the Weibull model

Why is COPVs Statistical Engineering?

- Application of Scientific Method to a Complex Problem
- Sequential Data Collection/Experimentation
- Each Phase Targeted Different Questions
- Clearly Documented Assumptions, Assessed via Data
- Took Proper Steps to Cleanse Data
- Real Research Question Involves System of Systems

VirginiaTech
Invent the Future

SE Impact for DoD, NASA, Labs

- Large, Unstructured, Complex Problems Everywhere!
- Can No Longer Afford "One-Off" Solutions
- Time/Resource Restraints Demand Effective Tactical Approaches for Problem Solutions
  - Issue has not been a lack of tools
  - Issue has been how to deploy these tools!
  - Major overlap with standards of practice for data analysis

SE Impact for DoD, NASA, Labs

HELLO STATISTICAL ENGINEERING!