Why Statistical Engineering?

NASA Conference on Statistical Engineering
Williamsburg, VA
May 3rd, 2011

Roger Hoerl, GE Global Research
With Considerable Input From Ronald D. Snee

Today's Agenda

- Motivation
- Today's Realities
- My "Blinding Light of the Obvious"
- Statistical Engineering Definition
  - Giving a Name to Something Too Often Done in the Shadows
- Examples
- Why Statistical Engineering Needs to be a Formal Discipline
  - Versus Done on an Ad-Hoc Basis
- Summary

My Motivation

- Internship at DuPont
- Early career at Hercules Chemical Co.
- Wake up call at Scott Paper Co.
- Career at GE

For me, this is personal

Today's Realities

The statistics discipline has much to be proud of and excited about:

- Hal Varian (chief economist at Google): "I keep saying that the sexy job in the next 10 years will be statisticians. And I'm not kidding."
- More interdisciplinary academic research being conducted
- Increased enrollments in service courses at many universities

At the same time, there is an ongoing concern that statisticians are not fully utilized within their organizations:

- Statisticians often are not sure how to take more initiative to get involved in the large, complex, unstructured "mission critical" projects in their organizations
- Management may see that statisticians can make larger contributions than passive consulting, but are not sure how to properly deploy them
- Professional colleagues may not understand how to engage statisticians in broader roles as collaborators, as opposed to going to them for support on narrow technical questions

Statisticians have much to offer: how do we unlock their potential?

My “Blinding Light of the Obvious”

Susan Hockfield – MIT President:

Around the dawn of the 20th century, physicists discovered the basic building blocks of the universe; a “parts list”, if you will. Engineers said “we can build something from this list,” and produced the electronics revolution, and subsequently the computer revolution.

More recently, biologists have discovered and mapped the basic “parts list” of life – the human genome. Engineers have said “we can build something from this list,” and are producing a revolution in personalized medicine.

Loosely quoted from January, 2010 seminar at GE Global Research

Application to Statistical Science

Two important questions we must answer:

Who is building something meaningful from the statistical science “parts list” of methods?

What are the implications of stopping at developing the parts list – the methods, and not building something of interest to society from them?

Can statisticians be thought leaders in addition to being “tools guys”?

Some Current Challenges

Some currently unsolved statistics problems:

- Ensuring that statistical projects have high impact
- How to attack big, complex, unstructured problems
  - Problems that do not "correspond to an identifiable textbook chapter" (Meng, 2009)
- The need to integrate the principles of statistical thinking with the application of statistical methods and tools
- Providing opportunities for statisticians to demonstrate true leadership to their organizations, rather than only passive consulting services

An opportunity to build something new from the parts list?

A Conjecture

Scientists, engineers, statisticians and other professionals have been building meaningful new things from the statistical science parts list of tools for some time, to society's benefit. However:

- This has typically been done on an ad-hoc basis with little or no underlying theory documented to provide guidance to others.
- Applications have generally been "one offs", requiring the wheel to be reinvented each time.
- This has significantly slowed progress, and missed opportunities to benefit society.

Not a new idea, but perhaps a new discipline

Statistical Engineering Fundamentals

Science:

- Systematic study and advancement of the facts and general laws of the physical world

Engineering:

- Study of how to best utilize scientific and mathematical principles for the benefit of humankind

Note: while scientists study and advance the fundamental laws of nature, engineers study how existing laws and principles could be put to better use, e.g., the IBM Computer "Watson"

- Engineers develop theory - How existing science can be better utilized - What works, what doesn't, and why
- The development and use of theory is the key differentiator between an "engineer" and a "practitioner"

Theory: A coherent group of general propositions used to explain a class of phenomena

- Note: there is no mention of mathematics in this or other common definitions of theory!

Terminology is important, and needs to be precise

Statistical Engineering Definition

Statistical engineering:

- The study of how to best utilize statistical concepts, methods, and tools and integrate them with information technology and other relevant sciences to generate improved results (Hoerl and Snee 2010a)
- In other words, trying to build something meaningful from the statistical science parts list

Notes

For this to be a true engineering discipline as opposed to just a sexier term for applied statistics, there must be a dynamic theoretical foundation based on rigorous research, just as there is for electrical engineering, mechanical engineering, and so on

This definition does not refer to application of statistics to engineering

- Statistical engineering can be applied to improving anything

This is a different definition than that used by Eisenhart, who we believe was the first to use this term in 1950

This definition is consistent with dictionary definitions of engineering

Elaboration of Definition

An important viewpoint to keep in mind:

The issues we raise above have nothing to do with the old distinction between applied statistics and theoretical statistics. The traditional viewpoint equates statistical theory with mathematics and thence with intellectual depth and rigor, but this misrepresents the notion of theory.

We agree with the viewpoint that David Cox expressed at the 2002 NSF Workshop on the Future of Statistics that "theory is primarily conceptual," rather than mathematical. (Lindsay et al. 2004)

- Statistical engineering is different from traditional applied statistics
- Theory is conceptual, not necessarily mathematical

Applied Statistics or Statistical Engineering?

Three Work Environments
Experienced by Roger Hoerl Early in his Career

Work Environment	Work Description	Comments
DuPont Engineering–Summer Intern 1981, 1982	Use statistics methods to address design and analysis problems presented to him	• Applied methods learned in graduate school. • Employer, co-workers and clients were pleased
Hercules 1983-87	Identify opportunities to apply statistical methods to important design and analysis problems	• Deepened understanding of how to apply methods learned in graduate school. • Employer, co-workers and clients were very pleased
Scott Paper 1987-1995	Deploy Statistical Process Control across the company	• Not covered in graduate school • Not covered in SPC books • Other skills beyond statistics were needed • Several concepts, methods and tools needed to be integrated

Example of Statistical Engineering:

Lean Six Sigma and the DMAIC Framework

Lean Six Sigma (LSS) is an obvious example of statistical engineering – building something from the parts list of tools. I believe that this is one reason for its prolonged success. LSS has been deployed broadly not only within the private sector, but within several branches of the US armed forces.

- LSS has not invented any new tools (no new statistical science), but has achieved much greater results from existing tools
  - The tools have been more effectively integrated and linked through the DMAIC model, for example
- LSS has developed dynamically through cycles of application of the scientific method – theory and experimentation
  - Addition of the Define stage, integration of Lean approaches, etc.
- LSS development has been based on a solid theoretical foundation in continuous improvement, e.g.:
  - Pareto principle (focus on a few key process drivers)
  - Use of project-by-project improvement (Juran)
  - Utilization of small (4-6 people) project teams (Useem 2006)
  - Use of tools with sound theoretical bases; DOE, regression, SPC, etc.

A result of "the study of how to best utilize statistical concepts, methods, …"

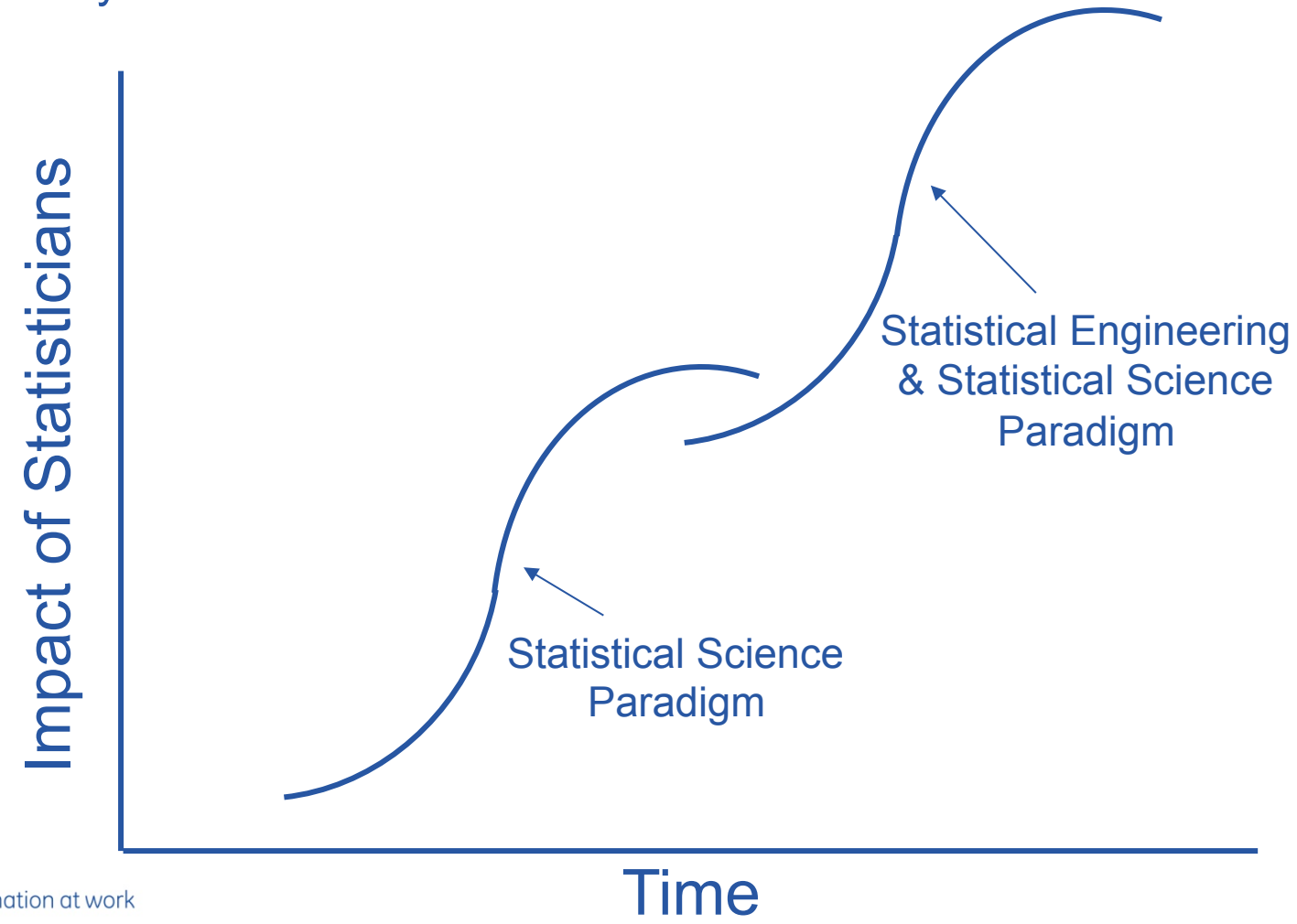
A Personal Example

- Problem: Develop a financial "default predictor" for GE
  - Big, unsolved problem!
  - Challenges of trading within a multi-billion dollar portfolio
  - No commonly accepted definition of default
  - Limited internal data – no set of "universal data" exists
  - No defined measure of success
- Approach:
  - Cross-functional team needed (statistics, OR, quantitative finance, business expertise), spread between in upstate New York, Bangalore, and Stamford, CT
  - Developed definition of default, and metrics to document success and failure
  - Data obtained externally – needed to merge disparate data sources
    - Set up direct feed from Wall Street
  - Final prediction methodology utilized:
    - Publicly available default predictor as an input
    - Smoothing algorithms, CART, simulation, Markov Chains, and censored data methods from reliability
  - Awarded US Patent for *system* – not for *algorithm*

Not at textbook problem – or solution

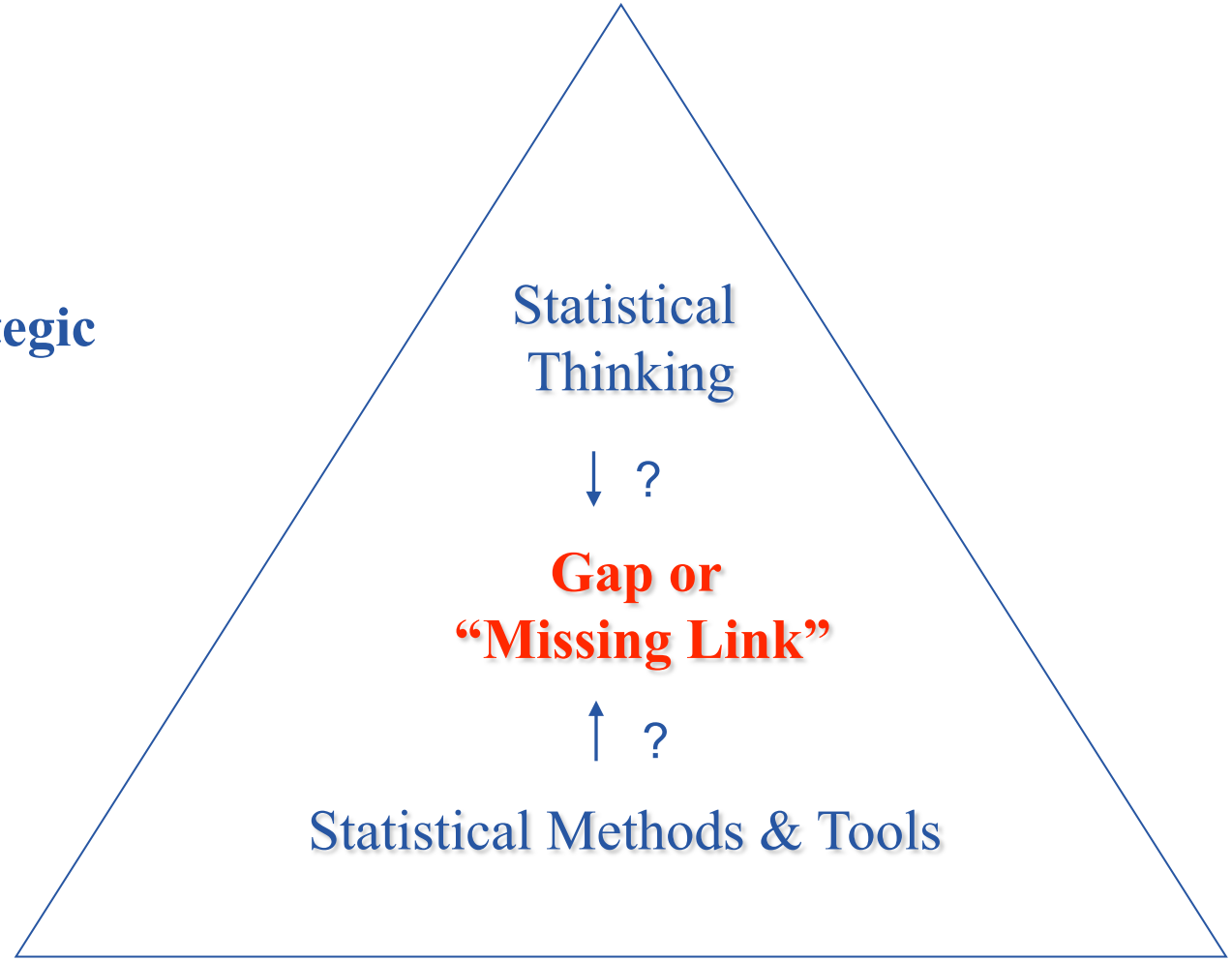
Potential Impact of Statistical Engineering

I believe that a balanced approach involving both statistical science and statistical engineering would enable us to have a much broader impact on society



Potential Impact of Statistical Engineering

Many have noted a gap between the concepts of statistical thinking and the application of statistical tools and methods



Statistical Thinking Definition

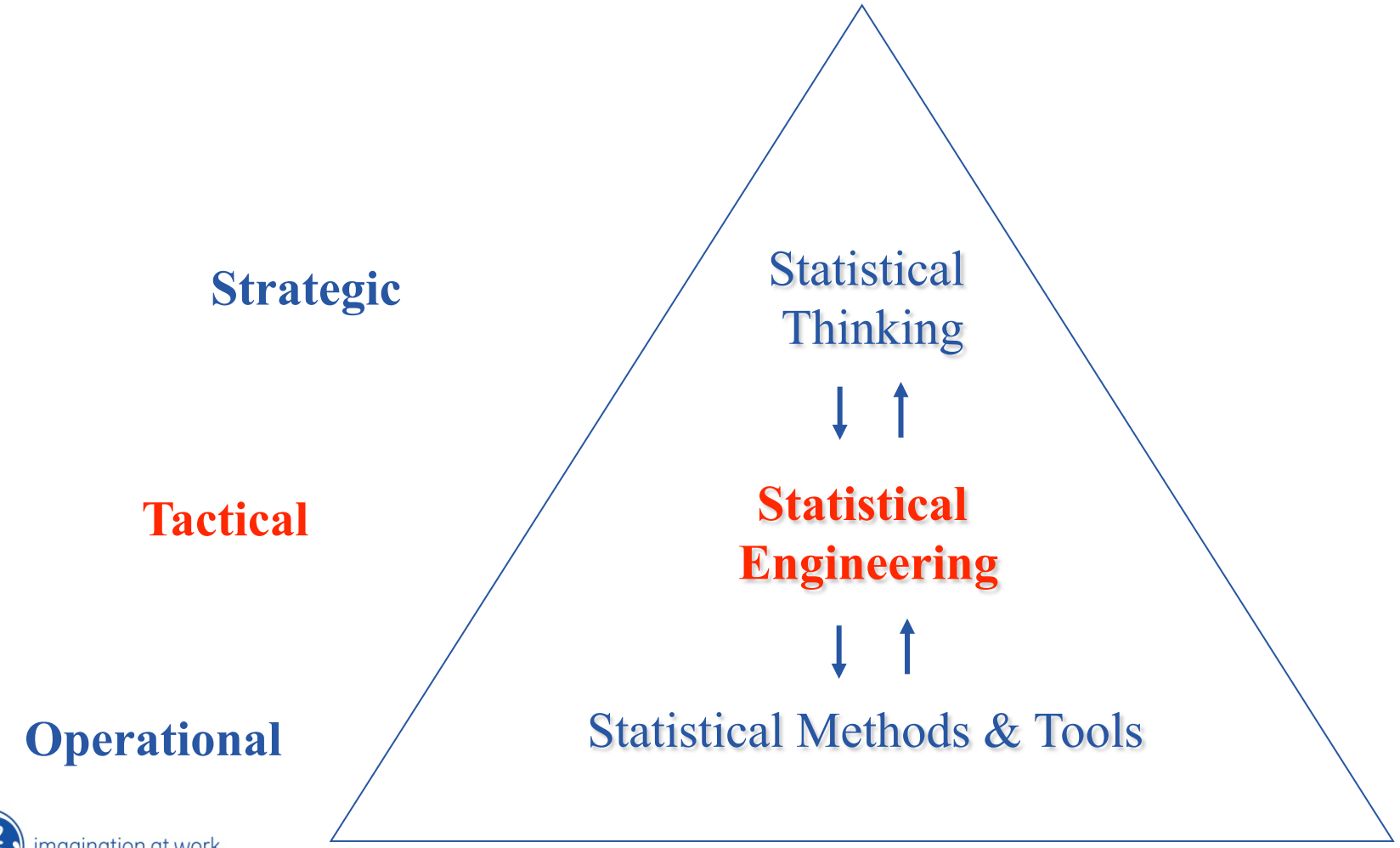
Statistical Thinking is a philosophy of learning and action based on these fundamental principles:

- All work occurs in a system of interconnected processes
- Variation exists in all processes
- Understanding and reducing variation are keys to success

ASQ Glossary and Tables for SQC, 1996

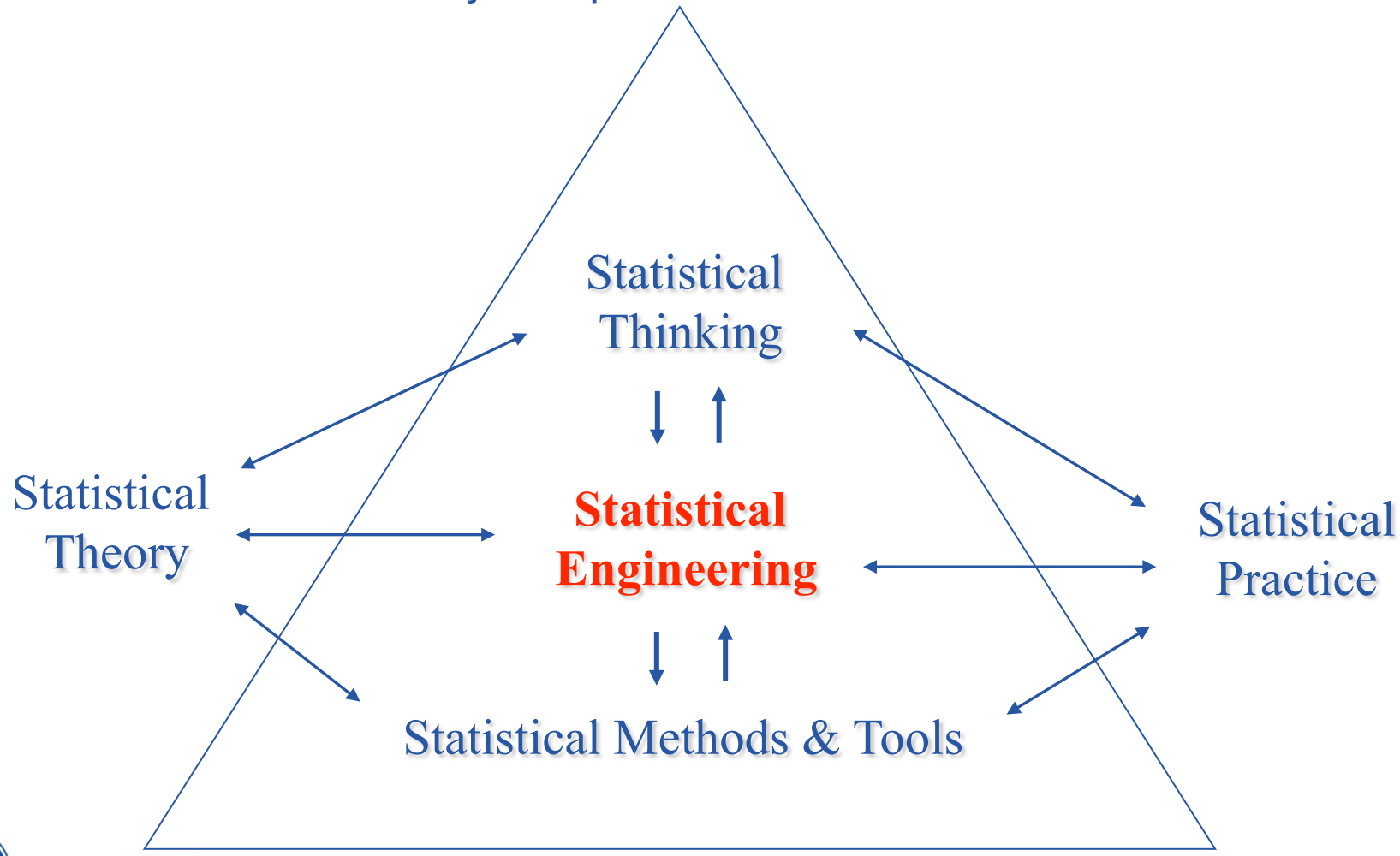
Potential Impact of Statistical Engineering

Statistical engineering is a potential bridge between statistical thinking and statistical tools and methods



Potential Impact of Statistical Engineering

We also believe that statistical engineering can help bridge the gap between statistical theory and practice



Large, Unstructured, Complex Problems

- The big payoff, mission critical problems that “do not correspond to an identifiable textbook chapter” (Meng 2009)
- Impact is broad – process performance, financial, customer, social and environmental
- Several departments, groups and functions are involved
- Problem has high degree of complexity involving both technical and non-technical challenges
- Multiple sources of data and information are used
- More than one statistical technique is required for solution
  - Typically non-statistical techniques are required
- Creative use of information technology is needed for success
- Long-term successes requires embedding solution into work processes typically through:
  - Use of custom software
  - Integration with other sciences and disciplines

Statistical engineering is needed for such problems – huge opportunity for the profession

My Message

To build on their successes and prosper in the 21st century, statisticians need to be more engaged in solving the big, complex, unsolved problems in their organizations

One key change needs to be a balanced focus on statistical engineering and statistical science

Statistical engineering can help:

- Ensure that statistical projects have high impact
- Provide a framework to attack large, complex, unstructured problems
- Integrate the principles of statistical thinking with the application of statistical methods and tools
- Enable statisticians to provide true leadership to their organizations,
  - Rather than focus on passive consulting services

I am anxious to hear what others have to say about statistical engineering this week.

References

Hoerl, R. W. and R. D. Snee (2002) *Statistical Thinking – Improving Business Performance*, Duxbury Press, Pacific Grove, CA.

Hoerl, R. W. and R. D. Snee (2009) "Post Financial Meltdown: What Do Services Industries Need from Us Now?", *Applied Stochastic Models in Business and Industry*, December 2009.

Hoerl, R. W. and R. D. Snee (2010a) "Moving the Statistics Profession Forward to the Next Level", *The American Statistician*, February 2010, 10-14.

Hoerl, R. W. and R.D. Snee (2010b) "Closing the Gap: Statistical Engineering can Bridge Statistical Thinking with Methods and Tools", *Quality Progress*, May 2010, 52-53.

Hoerl, R. W. and R. D. Snee (2010c) "Tried and True – Organizations put Statistical Engineering to the Test and See Real Results", *Quality Progress*, June 2010, 58-60.

Lindsay, B.G, Kettenring, J., and Siegmund, D.O. (2004) "A Report on the Future of Statistics (with Discussion)", Statistical Science, 19, 3, 387-413.

Meng, X. (2009) "Desired and Feared – What Do We Do Now and Over the Next 50 Years?", The American Statistician, 202-210.

Snee, R. D. and R. W. Hoerl (2003) Leading Six Sigma – A Step by Step Guide Based on Experience With General Electric and Other Six Sigma Companies, FT Prentice Hall, New York, NY,

Technometrics (2008) "Future of Industrial Statistics – A Panel Discussion". Technometrics Blog Link asq.org/discussionBoards/forum.ispa?forumID=77