Quicker and Better Quality Improvement Business Cases with Bayesian Belief networks and Six Sigma

Software Engineering Institute
Carnegie Mellon University
Pittsburgh, PA 15213

Ben Linders, SEI Affiliate (Ericsson)
Bob Stoddard, Senior Member Tech Staff, SEI

E-SEPG: Monday June 11, 2007

Contents / Agenda

Introduction

Six Sigma Methods

Exercise 1

Defect Modeling

Exercise 2

Conclusions

Introduction

Introduction

Quality improvement needed in many organizations

Business case required

- Identification of problem areas
- Selected improvement
- Quantified costs & benefits

Problem: No data available

- Measurement programs are costly
- Long lead time

Solution

Requirements

- Value/result driven
- Comprehensible, easy to use
- Objective & reliable
- Industry Standard Compatible (Benchmarking)
- Re-use best practices

Technologies

- Six Sigma
- GQIM, Balanced Scorecard
- Bayesian Belief Networks
- Cost of Quality, Root Cause Analysis

Two step approach

Quality Factor Model

- Expert opinion, extended with data
- Quick Quality Scan
- Rough Prediction Fault Slip Through
- Improvement Areas

Selected Improvement Model

- Data, tuned with expert opinion
- Detailed Prediction Fault Slip Through
- Improvement Business Case

Collaboration

NL: Market Unit Northern Europe & Main R&D Center

R&D: Value Added Services

- Strategic Product Management
- Product marketing & technical sales support
- Provisioning & total project management
- Development & maintenance
- Customization
- Supply & support

+/- 1300 employees, +/- 350 in R&D

Software Engineering Measurement & Analysis

Modern Measurement Methods

- Goal Driven Measurement
- Managing Projects with Metrics
- Measuring for Performance-Driven Improvement -I, -II
- Understanding CMMI High Maturity Practices
- Client Support & Research
- Training Development & Delivery

Affiliate Assignment

Joint effort: Ericsson (Ben Linders) and SEI (Bob Stoddard)

- Time, money, materials
- Knowledge & experience

Deliverables Ericsson

- Defect data & benchmarks
- Improved decisions skills
- Business case & Strategy 2007:
  - Early phases: **Improvements**
  - Late test phases: **Reduction**

Research contribution

- Apply Six Sigma business cases
- Verify technology (CoQ, RBT, FST, etc)

Six Sigma Methods

DMAIC Roadmap

ERICSSON

Define
- Define project scope
- Establish formal project

Measure
- Identify needed data
- Obtain data set
- Evaluate data quality
- Summarize & baseline data

Analyze
- Explore data
- Characterize process & problem
- Update improvement project scope & scale

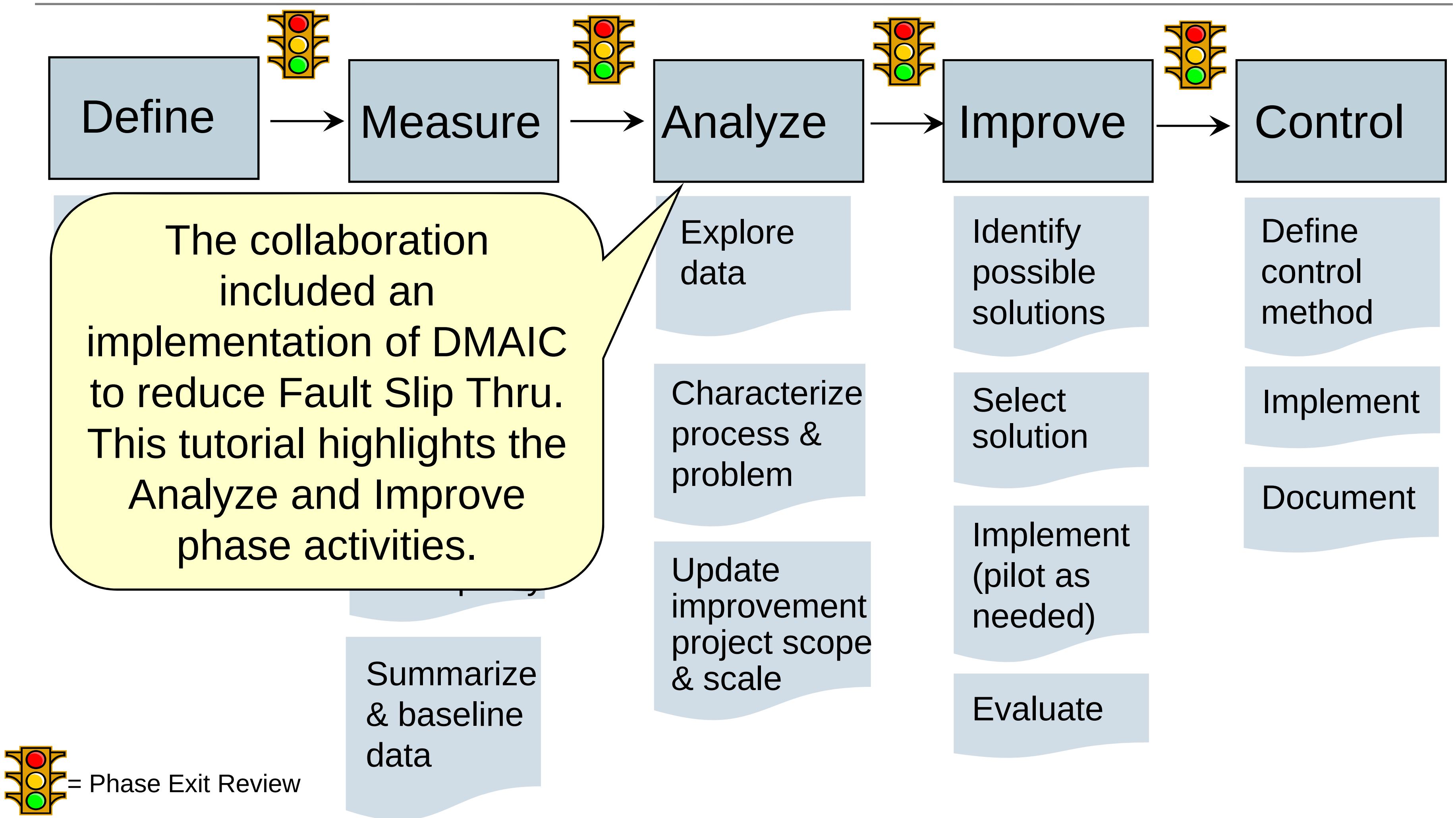
Improve
- Identify possible solutions
- Select solution
- Implement (pilot as needed)
- Evaluate

Control
- Define control method
- Implement
- Document

= Phase Exit Review

DMAIC Roadmap

ERICSSON



Basic Statistical Prediction Models

X \ Y	Continuous	Discrete
Discrete	ANOVA & MANOVA	Chi-Square & Logit
Continuous	Correlation & Regression	Logistic Regression

Example ANOVA Output

ERICSSON

One-way ANOVA: Escaped Defect Density versus Quality Check

```
Source          DF       SS      MS      F      P
Quality Check    4  139.519  34.880  48.30  0.000
Error          330  238.306   0.722
Total          334  377.825

S = 0.8498   R-Sq = 36.93%   R-Sq(adj) = 36.16%

                                     Individual 95% CIs For Mean Based on
                                     Pooled StDev
Level             N    Mean   StDev  ----+---------+---------+---------+-----
System Test      94  3.7298  0.6768  (*-)
Inspection      110  4.5164  0.8615          (*-)
Walkthrough      88  4.8568  0.8417             (-*)
Informal w/Peer  37  5.6081  1.1186                   (--*--)
Email Comments    6  6.6500  1.2755                          (------*------)
                                     ----+---------+---------+---------+-----
                                       4.0       5.0       6.0       7.0
```

We predict a range of escaped defect density for each type of quality check.

Example Regression Output

ERICSSON

Regression Analysis: Defect Densi versus ReqtsVolatil, YearsDomainE

```
The regression equation is
Defect Density = 0.484 + 0.480 ReqtsVolatility - 0.0242 YearsDomainExperience
```

Predictor	Coef	SE Coef	T	P
Constant	0.48367	0.03957	12.22	0.000
ReqtsVolatility	0.47963	0.09511	5.04	0.000
YearsDomainExperience	-0.024215	0.001941	-12.48	0.000

S = 0.00893207   R-Sq = 85.9%   R-Sq(adj) = 84.8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	2	0.0126076	0.0063038	79.01	0.000
Residual Error	26	0.0020743	0.0000798		
Total	28	0.0146819			

Use of Design of Experiments

Essentially a sophisticated method of sampling data to conclude relationships

Provides more confidence in possible cause-effect relationships

Enables us to define a small, efficient set of scenarios which we can then include in surveys of experts

Results help to populate relationships in the Bayesian Belief Network (BBN) model

Example of Design of Experiments

Welcome to Minitab, press F1 for help.

Fractional Factorial Design

Factors: 5 Base Design: 5, 8 Resolution: III
Runs: 8 Replicates: 1 Fraction: 1/4
Blocks: 1 Center pts (total): 0

* NOTE * Some main effects are confounded with two-way interactions.

A	B	C	D	E		Response
1	-1	-1	-1	-1		
-1	-1	1	1	-1		
-1	-1	-1	1	1		
1	1	-1	1	-1		
1	1	1	1	1		
-1	1	1	-1	-1		
-1	1	-1	-1	1		
1	-1	1	-1	1		

Why Use Monte Carlo Simulation?

Allows modeling of variables that are uncertain (e.g. put in a range of values instead of single value)

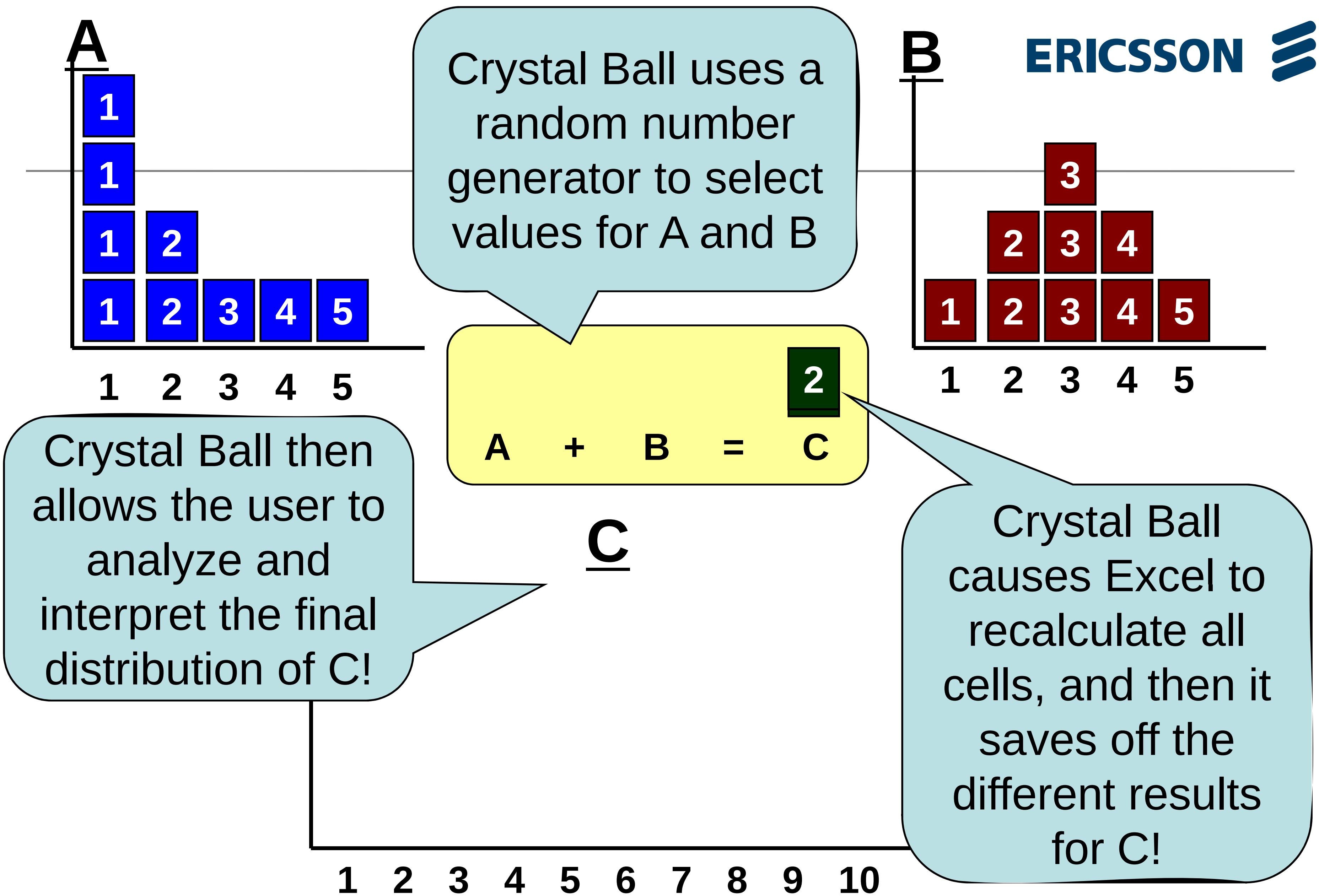
Enables more accurate sensitivity analysis

Analyzes simultaneous effects of many different uncertain variables (e.g. more realistic)

Eases audience buy-in and acceptance of modeling because their values for the uncertain variables are included in the analysis

Provides a basis for confidence in a model output (e.g. supports risk management)

"All Models are wrong, some are useful" – increases usefulness of the model in predicting outcomes


A
1
1
1 2
1 2 3 4 5
1 2 3 4 5
Crystal Ball uses a random number generator to select values for A and B
B
ERICSSON
3
2 3 4
1 2 3 4 5
1 2 3 4 5
2
A + B = C
Crystal Ball then allows the user to analyze and interpret the final distribution of C!
C
Crystal Ball causes Excel to recalculate all cells, and then it saves off the different results for C!
1 2 3 4 5 6 7 8 9 10

Why Use Optimization Modeling?

Partners with Monte Carlo simulation to automate tens of thousands of “what-ifs” to determine the best or optimal solution

Best solution determined via model guidance on what decisions to make

Easy to use by practitioners without tedious hours using analytical methods

Uses state-of-the-art algorithms for confidently finding optimal solutions

Supports decision making in situations in which significant resources, costs, or revenues are at stake

Several Example Tools

http://www.palisade.com/trials.asp

@RISK

The world's most powerful risk analysis tool. Take into account all possible scenarios using Monte Carlo simulation. Work directly in Excel, create presentation-quality graphs, use distribution fitting, and more!

@RISK for Project

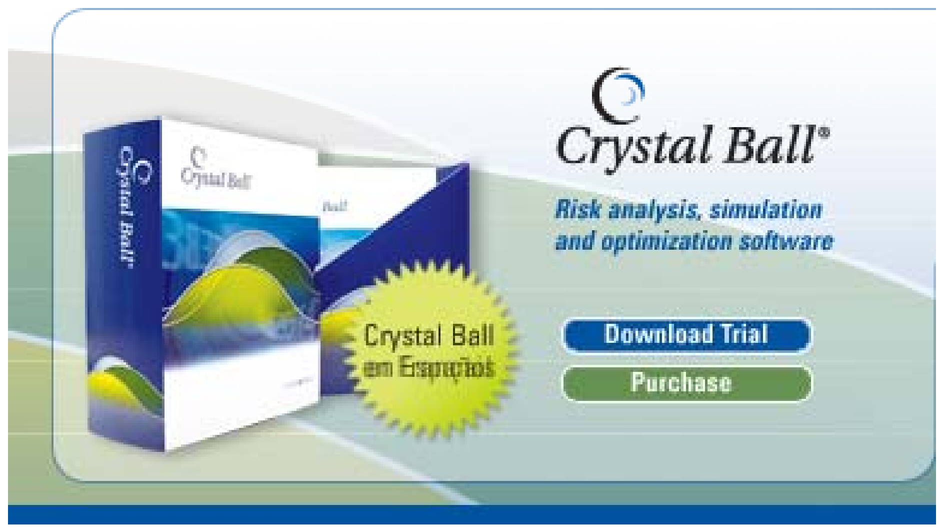
Analyze cost and schedule risks in Microsoft Project using Monte Carlo simulation.

- STANDARD
- PROFESSIONAL

http://www.decisioneering.com/

DECISIONEERING

Risk Resources | Products | Tra



Probabilistic Models - 1

A Bayesian network is a probabilistic graphical model, also known as a Bayesian Belief Network (BBN) or belief network.

A Bayesian network is represented by a graph, in which the nodes of the graph represent variables, and the edges represent conditional dependencies.

The joint probability distribution of the variables is specified by the network's graph structure. The graph structure of a Bayesian network leads to models that are easy to interpret, and to efficient learning and inference algorithms.

From Wikipedia, the free encyclopedia

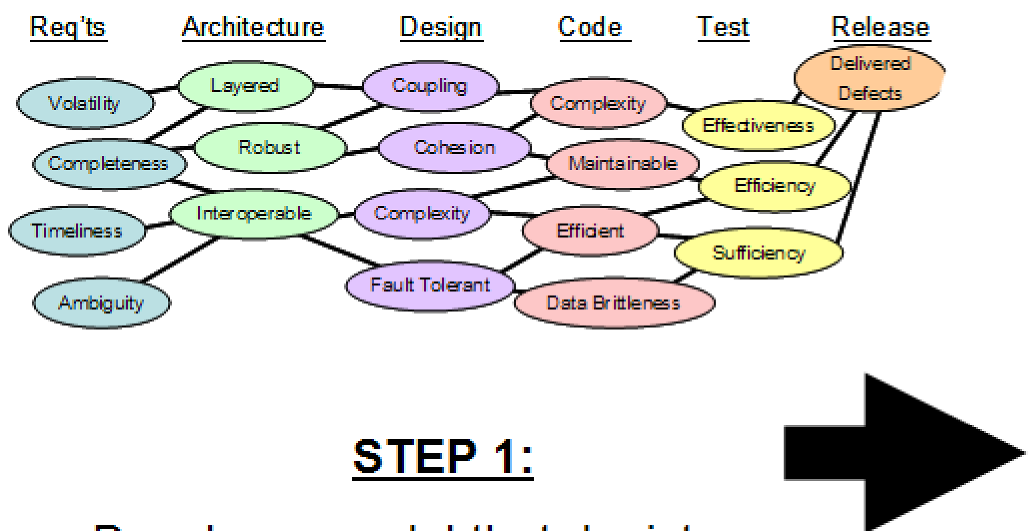
Probabilistic Models - 2

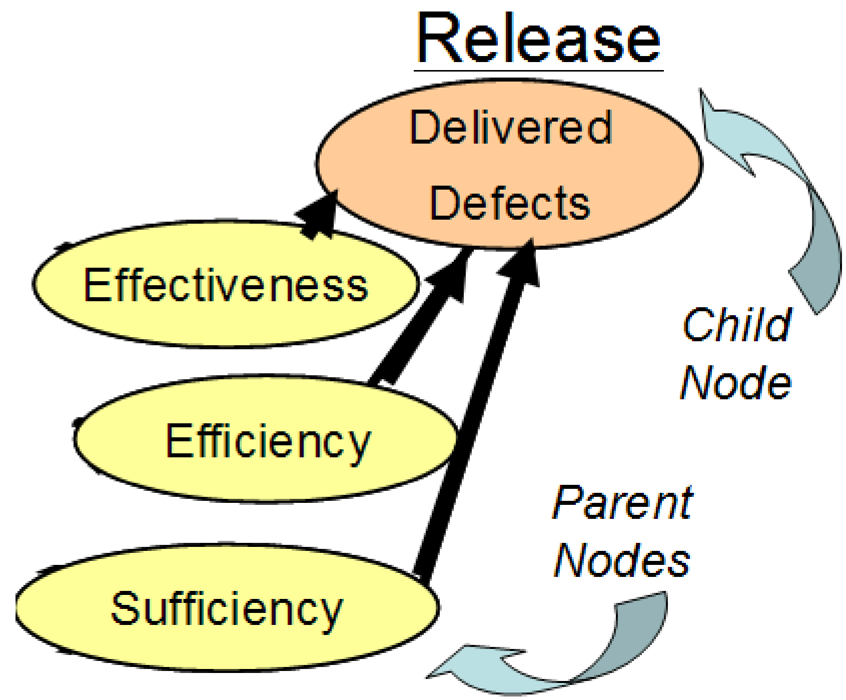
Nodes can represent any kind of variable, be it a measured parameter, a latent variable, or a hypothesis. They are not restricted to representing random variables; this is what is "Bayesian" about a Bayesian network.

Bayesian networks may be used to diagnose and explain why an outcome happened, or they may be used to predict outcomes based on insight to one or more factors.

From Wikipedia, the free encyclopedia

Example of Bayesian Belief Model





STEP 1:

Develop a model that depicts leading indicator nodes and/or root cause nodes for each node to be predicted. In this diagram, the flow is from left to right and thus, each child node (nodes with incoming lines) may be predicted with the parent nodes (nodes sending a line to a child node)

STEP 2:

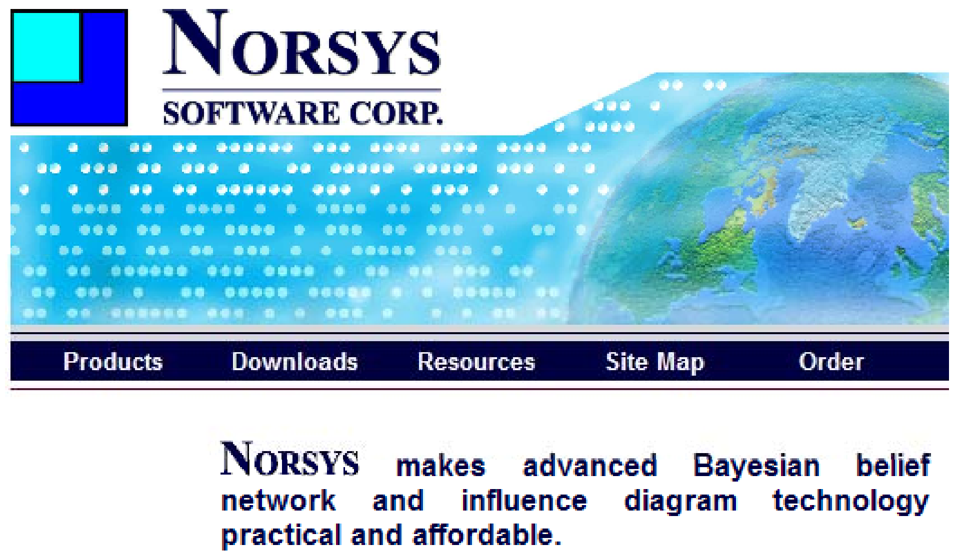
Model the relationship between each "child" node and the associated "parent" nodes for the child using:

Regression & ANOVA with Objective Data
Design of Experiments with Subjective Data

Examples of BBN Tools

“AGENARISK” http://www.agena.co.uk/

“NETICA” http://www.norsys.com/



“HUGIN” http://www.hugin.com/

Exercise 1

Defect Modeling

History Defect Modeling

ERICSSON

2001

- Defect Model defined, pilot in first project

2002/2003

- Improved based on project feedback
- First release quality predictions
- Industrialize model/tool, use in all major projects

2004/2005

- Targets: Project portfolio management
- Process Performance & Cost of Quality

2006/2007

- Process Improvement Business Cases
  SW Engineering Economics, Six Sigma
- Fault Slip Through reduction

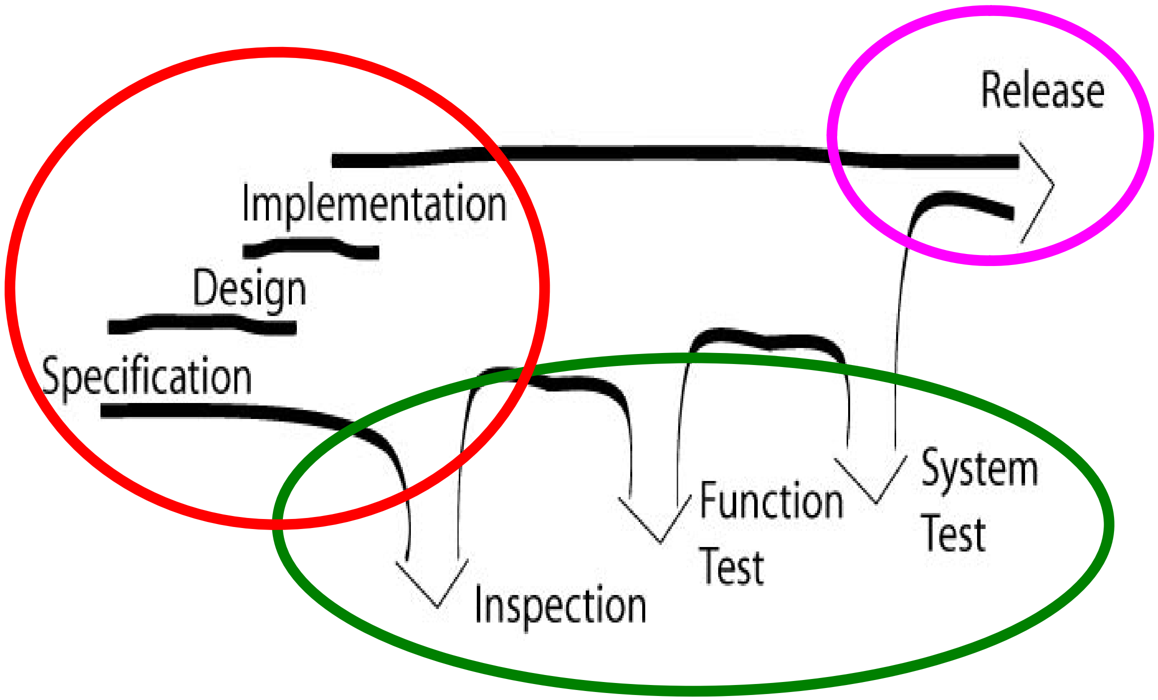
Project Defect Model

Why?

- to control quality of the product during development
- improve development/inspection/test processes

Business Value:

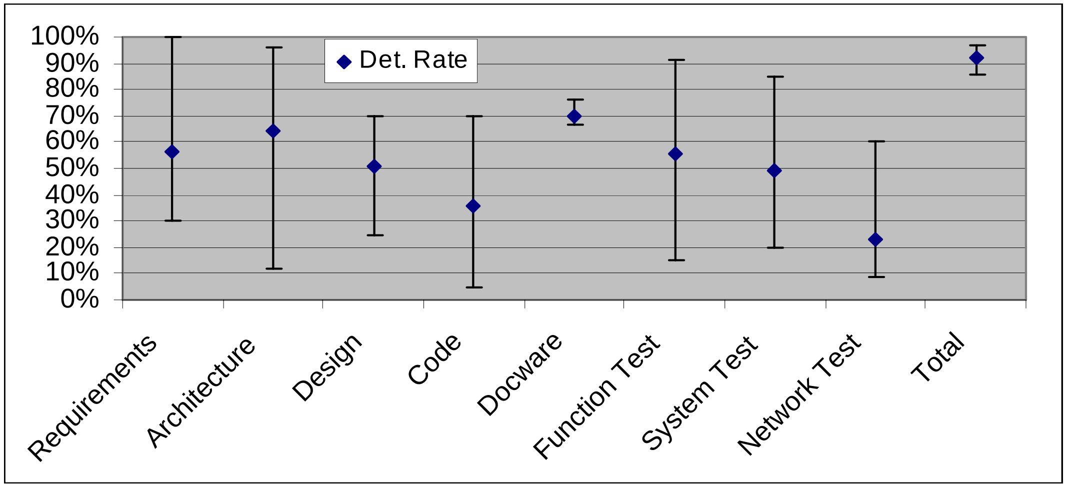
- ➔ Improved Quality
- ➔ Early risks signals
- ➔ Better plans & tracking
- ➔ Lower maintenance
- ➔ Save time and costs
- ➔ Happy customers!



Process Performance

Project Data

- Insertion Rates
- Detection Rates
- Defect Distribution
- Fault Slip Through
- Post Release Defects

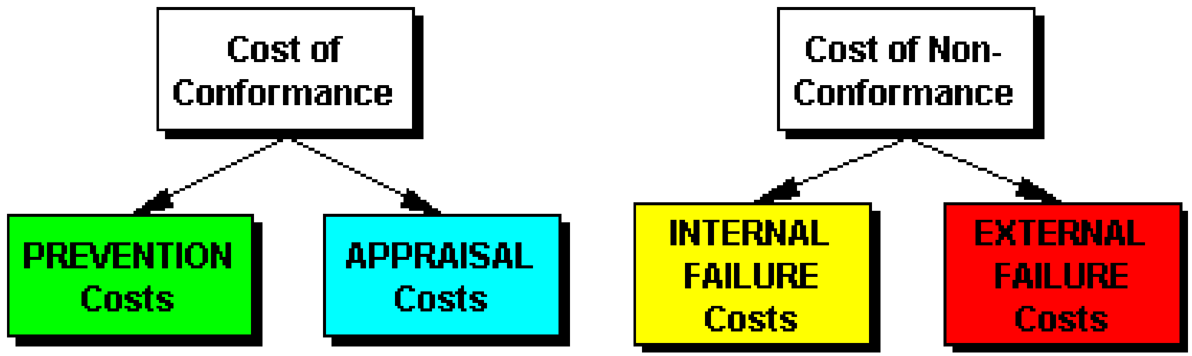


Process View

- Performance of design & test processes
- Benchmarking
- Best Practices & Improvement Areas

Cost of Quality

ERICSSON



Main value to gain:

- Increase appraisal effectiveness
- Decrease failure costs

Improve performance & Invest in Prevention

? Cost determinators, and their results

? Relationships between cost catagories (ROI)

Ben Linders & Bob Stoddard

Software Engineering Economics

http://citeseer.ist.psu.edu/boehm00software.html

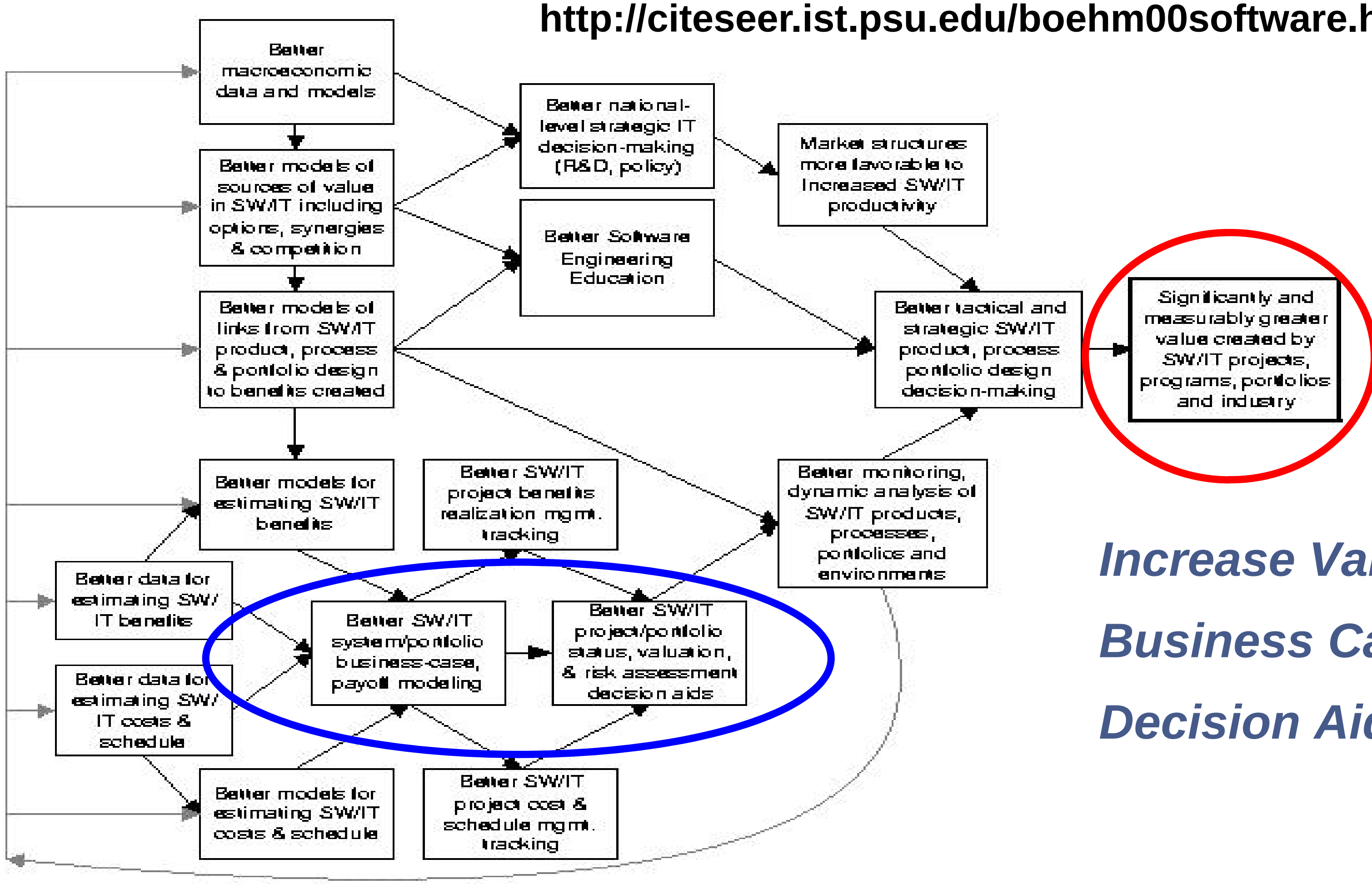


Figure 1: Roadmap for research in software engineering economics.

Increase Value

Business Cases

Decision Aids

Economic Model

Understand the costs of defects

Link process & project performance

Dialog between managers & developers

Use available operational data

Manage under uncertainty & incomplete data

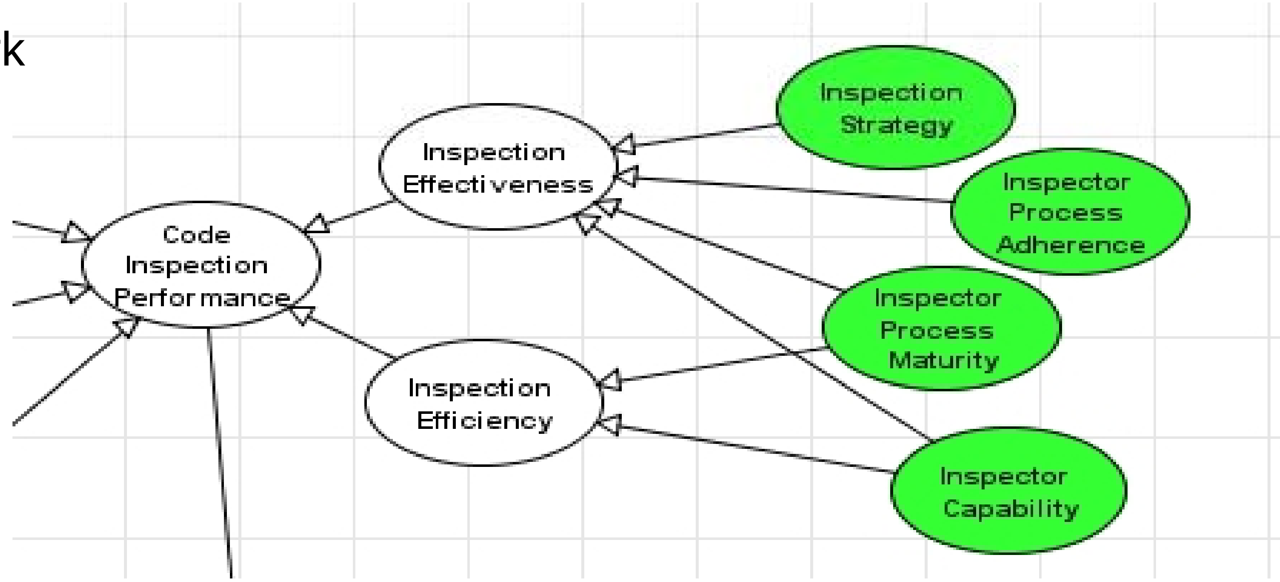
Technologies

- Cost of Quality
- Bayesian Belief Networks
- Real Options
- Lean Six Sigma

Step 1a: Quality Factor Model

Bayesian Belief Network

- Phases
- Quality Factors
- Expert opinion
- Prediction of Quality Impact



Managerial: Line, project & Process Management

Technical: Requirements, Design, Implementation, Inspection, Test

Step 1b: Prediction of Fault Slip Through

Step 2: Selected Improvement Model

ERICSSON

See ongoing discussion on modeling.

Exercise 2

Exercise: Predict Fault Slip Through

Conclusions

Conclusions

Benefits

- Quicker decisions improvement scope
- Better Business Case ????

- Our six sigma approach, which combined subjective and objective data quantified in a Bayesian Belief Network Model (BBN), along with a business benefit Monte Carlo simulation using Design of Experiment methods, is a practical and efficient approach to derive a solid business case in a short timeframe. It also helps to prioritize improvements based on the expected value for the business, which will lead to a quick return on investment.

SEI Affiliate

The Software Engineering Institute Affiliate Program provides sponsoring organizations with an opportunity to contribute their best ideas and people to a uniquely collaborative peer group who combine their technical knowledge and experience to help define superior software engineering practices.

Affiliates: http://www.sei.cmu.edu/collaborating/affiliates/affiliates.html

ERICSSON

TAKING YOU FORWARD

Software Engineering Institute | Carnegie Mellon