University of Southern California
Center for Systems and Software Engineering

Building Cost Estimating Relationships for Acquisition Decision Support

Brad Clark, Ray Madachy, Barry Boehm, Wilson Rosa

14th Annual Practical Software and Systems Measurement User Group Conference

July 26 to 30, 2010

University of Southern California
Center for Systems and Software Engineering

Topics

- Research problem and objectives
- Data challenges and resolution
- Results
- Future work

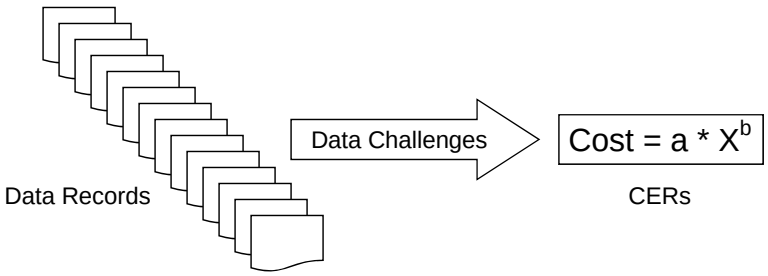
Project led by the Air Force Cost Analysis Agency (AFCAA) working with service cost agencies, and assisted by University of Southern California and Naval Postgraduate School

Problem

- For many years, there have been efforts to collect data from multiple projects and organizations
  - Data Analysis Center for Software (DACS)
  - Software Engineering Information Repository (SEIR)
  - International Software Benchmarking Standards Group (ISBSG)
  - Large Aerospace Mergers (Attempts to create company-wide databases)
  - USAF Mosemann Initiative (Lloyd Mosemann Asst. Sec. USAF)
  - USC CSSE COCOMO II repository
  - DoD Software Resources Data Report (SRDR)
- Purpose: to derive estimating relationships and benchmarks for size, cost, productivity and quality
- All have faced common challenges such as data definitions, completeness and integrity

Research Objectives

- Using SRDR data, improve the quality and consistency of estimating methods across cost agencies and program offices through guidance, standardization, and knowledge sharing.
  - Characterize different Application Domains and Operating Environments within DoD
  - Analyze collected data for simple Cost Estimating Relationships (**CER**) within each domain
  - Develop rules-of-thumb for missing data
- Make collected data useful to oversight and management entities



Background

- DoD has collected software resource data for a number of years
  - Product and development description
  - Product size
  - Resources and schedule
  - Product quality
- Analysis of ~140 records out of ~300
  - Additional data is coming in
- Data is segregated into Application Domains and Operating Environments
  - 21 Application Domains
  - 8 Operating Environments

Software Resources Data Report

SECURITY CLASSIFICATION

Software Resources Data Report: Final Developer Report - Sample

Due 60 Days After Final Software Delivery and 60 Days After Delivery of Any Release or Build

Page 1: Report Context, Project Description and Size

1. Report Context			
1. System/Element Name (version/release)			2. Report As Of:
3. Authorizing Vehicle (MOU, contract/amendment, etc.):			4. Reporting Event: Submission # 1 (Supersedes # _____, if applicable)
Description of Actual Development Organization			
5. Development Organization:	6. Certified CMM I Level (or equivalent):		8. Lead Evaluator:
	7. Certification Date:		9. Affiliation:
10. Precedents (list up to five similar systems by the same organization or team):			
Comments on Part 1 responses:			

2. Product and Development Description	Percent of Product Size	Actual Development Process	Upgrade or New?
1. Primary Application Type:	%	3.	4.
5. Secondary Application Type:	%	7.	8.

Source: http://dcarc.pae.osd.mil/Policy/CSDR/csdrReporting.aspx

University of Southern California
Center for Systems and Software Engineering

SRDR Data

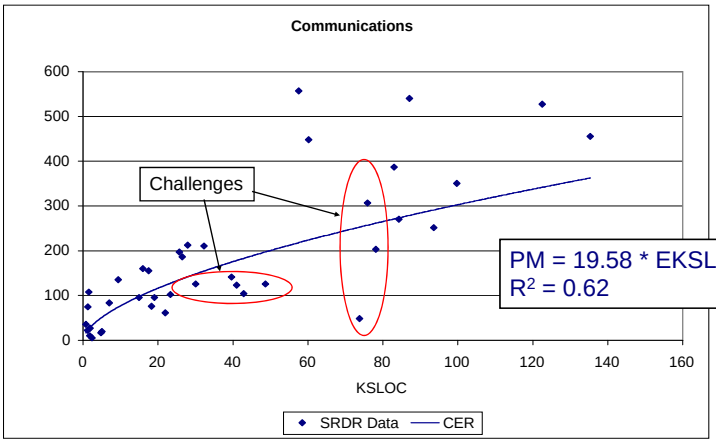
Application Domain	Operating Environment: Avionics	Fixed Ground	Missile	Missile	Mobile Ground	Ship-board	Unman-ned Airborne	Unman-ned Space	Total
Business Systems		1			4				5
Command & Control	1	14			16				31
Communications	4	42				7		1	54
Controls & Displays	2	1			2	3			8
Executive						3			3
Information Assurance		1							1
Infrastructure or Middleware		6				1			7
Maintenance & Diagnostics	1					2			3
Mission Management	20	2	3		2		1		28
Mission Planning	1	13							14
Process Control					4				4
Scientific Systems						3			3
Sensor Control and Processing	3	15				10			28
Simulation & Modeling		17				3			20
Spacecraft Payload								2	2
Test & Evaluation		2				2			4
Tool & Tool Systems		6		1					7
Training					2				2
Weapons Delivery and Control	7			9					16
Total	39	120	3	10	30	34	1	3	240

Missing Domains: Internet, Spacecraft bus

Note: SRDR is Software Resources Data Report

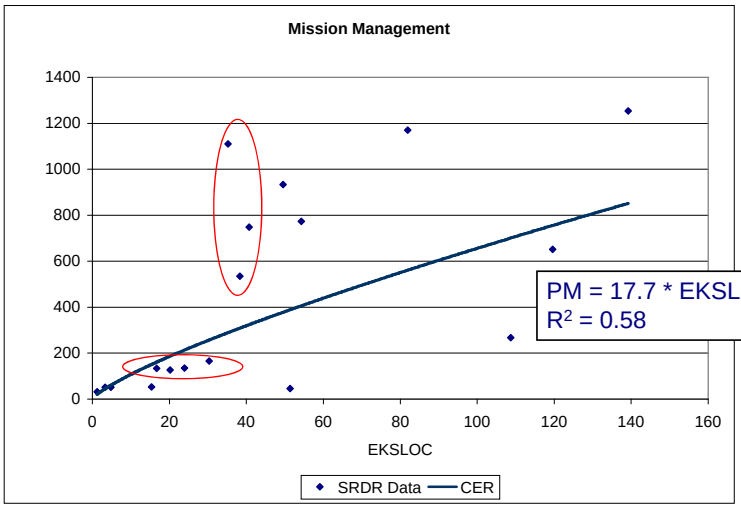
University of Southern California
Center for Systems and Software Engineering

Data Challenges-1



Notes: PM: Person Months (152 labor hours / month)
EKSLOC: Equivalent Thousands of Source Lines of Code
R^2: Proportion of variability in a data set that is accounted for by the statistical model. A measure of how well future outcomes are likely to be predicted by the model. Values vary from 0 to 1.

Data Challenges-2

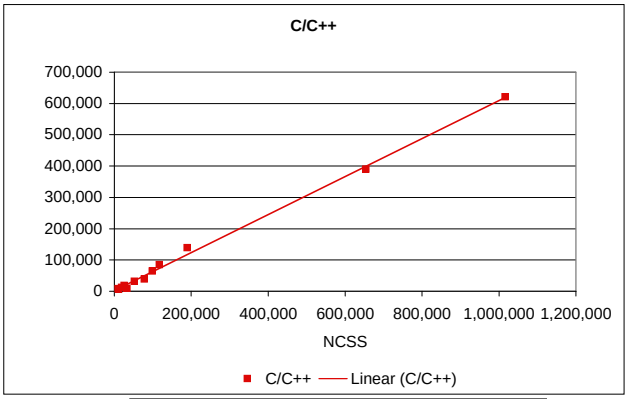


Data Issue-1 Inconsistent Counts

- Mix of SLOC counting methods
  - Total Lines: total number of lines in a file
  - Non-Commented Source Statements (NCSS): no blank or comment lines
  - Logical Lines
- No deleted code counts
- Not enough information to account for cost impact when using pre-existing software
  - Modified / Reused / Auto-generated
- SLOC conversion experiment
  - Use the results of USC's Code Count Tool to find conversion ratios
  - Total Lines to Logical Lines
  - NCSS to Logical Lines
  - Results segregated by programming language

Note: USC Code Count Tool is FREE from http://sunset.usc.edu/research/CODECOUNT/

SLOC Count Conversion Experiment



Logical SLOC = 0.611 * NCSS Count
R^2 = 0.9974

SLOC Count Conversion Factors

	Data Count	Total Line to Logical	NCSS to Logical
Ada	4	0.25	0.52
C/C++	12	0.32	0.61
C#	8	0.35	0.68
Java	6	0.35	0.72
Perl	4	0.53	0.70
PHP	4	0.44	0.66
Overall	38	0.33	0.64

For example, (C++ NCSS SLOC Count) * 0.61 = (C++ Logical SLOC Count)

Data Issue-2 Equivalent Counts

- No Modified Code parameters
  - Percent Design Modified (DM)
  - Percent Code Modified (CM)
  - Percent Integration and Test Modified (IM)
  - Software Understanding (SU)
  - Programmer Unfamiliarity (UNFM)
- Program interviews provided parameters for some records (table below)

Application Domain	DM Mean	DM Count	CM Mean	CM Count	IM Mean	IM Count
Command and Control	0.01	2	0.42	4	0.51	4
Communication	0.20	15	0.26	16	0.59	16
Control and Display	0.05	2	0.05	2	0.05	2
Mission Management	0.14	5	0.51	10	0.71	10
Sensor Control and Processing	0.50	3	0.50	3	1.00	3
Simulation and Modeling	0.25	7	0.38	8	0.58	8
Weapon Delivery and Control	0.15	4	0.15	4	1.00	4

Data Issue-3 Missing Effort

- Missing effort reporting for different lifecycle phases
  - Software requirements analysis (REQ)
  - Software architectural design (ARCH)
  - Software coding and testing (CODE)
  - Software integration (INT)
  - Software qualification testing (QT)
  - Software management, CM, QA, etc. (Other – very inconsistent)

Average Effort Spent by Phase

Application Domain	REQ	ARCH	CODE	INT	QT
Command & Control	18.7%	27.8%	39.1%	13.2%	10.6%
Communications	13.4%	28.6%	33.3%	21.7%	3.9%
Control & Displays	13.4%	36.2%	34.4%	13.9%	3.7%
Mission Management	16.8%	15.7%	32.6%	26.6%	25.8%
Mission Planning	19.2%	12.5%	47.6%	17.0%	4.8%
Sensor Control & Processing	8.5%	45.5%	24.8%	15.5%	9.3%
Simulation	12.1%	29.1%	38.3%	18.8%	6.3%
Weapons Delivery & Control	14.9%	17.5%	34.2%	22.0%	12.5%

Data Issue-4 Personnel Experience

- SRDR Data Definition
  - Report the percentage of project personnel in each category
  - Highly Experienced in the domain (three or more years of experience)
  - Nominally Experienced in the project domain (one to three years of experience)
  - Entry-level Experienced (zero to one year of experience)
- Need to include Personnel Experience (PX) in CERs to estimate cost
- After analyzing the data, the following quantitative values are assigned:
  - Highly experienced: 0.60
  - Nominally experienced: 1.00
  - Entry-level experienced: 1.30

University of Southern California
Center for Systems and Software Engineering

Data Issue-5 SRDR Data Collection

Current SRDR	Proposed Modifications
Application Types (3.7.1 – 17)	• Reorganize around Operating Environments and Application Domains • Add Mission Criticality (add reliability and complexity in a single rating scale) • Revisit detailed definitions of the Application Domains
Amount of New (>25%) Modified (<25% mod) Code	• Add DM, CM, IM, SU, & UNFM factors for modified code • Incorporate DM-CM-IM questionnaire • Add IM for Reused code • Definitions for code types • Count at the level it will be maintained
Deleted Code	• Report deleted code counts
Software and External Interface Requirements	• Add anticipated requirements volatility to 2630-1, 2 • Use percentage of requirements change as volatility input (SRR baseline)
Personnel Experience & Turnover	• Add to 2630-1 • Expand years of experience rating scale to 12 years

Data Issue-5 SRDR Data Collection

Current SRDR	Proposed Modifications
Project- or CSCI-level data	• Specify the level of data reporting
All Other Direct Software Engineering Development Effort (4.7): • Project Management • IV&V • Configuration Management • Quality Control • Problem Resolution • Library Management • Process Improvement • Measurement • Training • Documentation • Data Conversion • Customer-run Acceptance Test • Software Delivery, Installation & Deployment	Break into: • Management functions • Configuration / Environment functions • Assessment functions • Organization functions (e.g. user & maintainer documentation, measurement, training, process improvement, etc.)
Product Quality (Optional)	• Are there better measures than Measured Mean Time to Serious or Critical Defect ?

Results

	Operating Environment								
Application Domain	Avionics	Fixed Ground	Missile	Missile	Mobile Ground	Ship-board	Unman-ned Airborne	Unman-ned Space	Total
Business Systems		1			4				5
Command & Control	1	14			16				31
Communications	4	42				7		1	54
Controls & Displays	2	1			2	3			8
Executive						3			3
Information Assurance		1							1
Infrastructure or Middleware		6				1			7
Maintenance & Diagnostics	1					2			3
Mission Management	20	2	3		2		1		28
Mission Planning	1	13							14
Process Control					4				4
Scientific Systems						3			3
Sensor Control and Processing	3	15				10			28
Simulation & Modeling		17				3			20
Spacecraft Payload								2	2
Test & Evaluation		2				2			4
Tool & Tool Systems		6		1					7
Training					2				2
Weapons Delivery and Control	7			9					16
Total	39	120	3	10	30	34	1	3	240

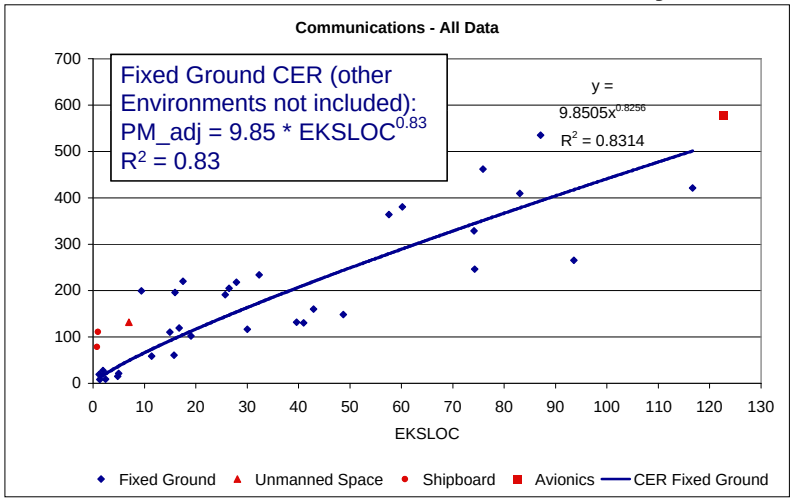
Results data set

Communication Domain Analysis-1

Domain	Examples	Brief Definition
Communications	• Radios • Microwave controller • Large telephone switching systems • Network management	Software that controls the transmission and receipt of voice, data, digital and video information. The software operates in real-time or in pseudo real-time. *Environment*: **Fixed ground**, mobile ground, manned and unmanned airborne, or unmanned space.

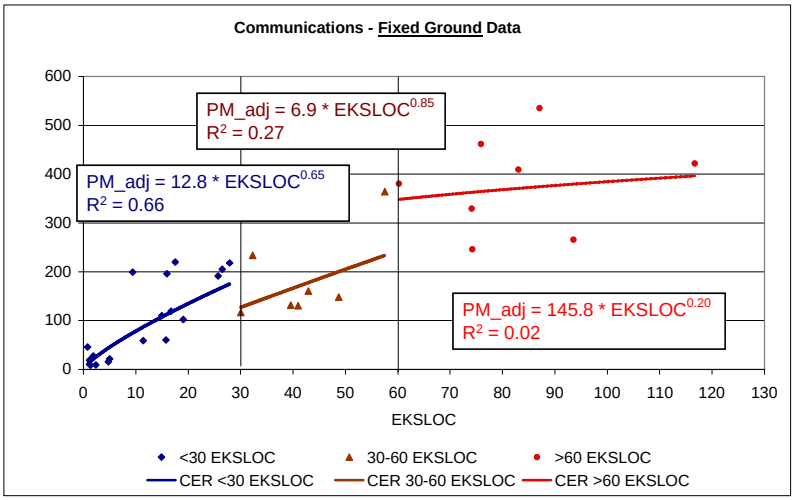
Environment	Examples	Brief Definition
Fixed Ground	• Computing facilities • Command and Control centers • Tactical Information centers • Communication centers	Manned and unmanned fixed, stationary land sites (buildings) with access to external power sources, backup power sources, physical access to systems, regular upgrades and maintenance to hardware and software, support for multiple users. Possible noisy environment.

Communication Domain Analysis-2



Note: Person Months-Adjusted are the actual Person Months with the effect of Personnel Experience removed

Communication Domain Analysis-3



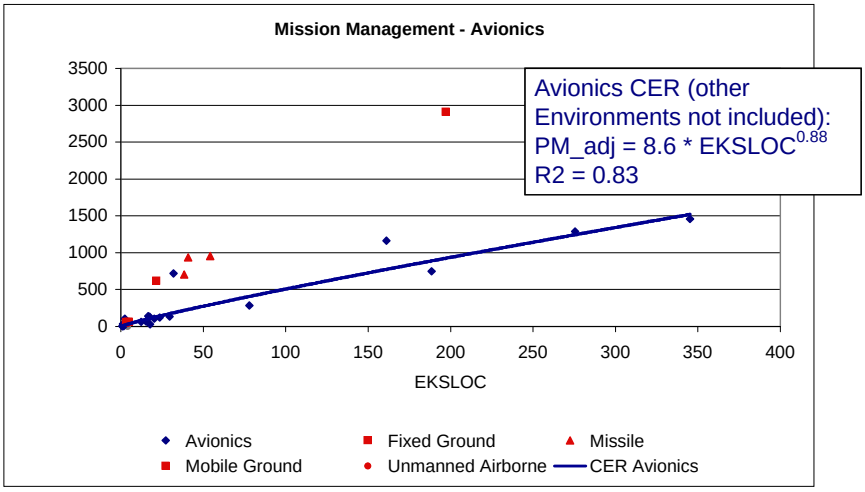
Mission Management Analysis-1

Domain	Examples	Brief Definition
Mission Management	• Operational Flight Program • Mission Computer • Flight Control Software	Software that enables and assists the operator in performing mission management activities including scheduling activities based on vehicle, operational and environmental priorities. *Environment*: Mobile ground, **avionics** or manned space.

Environment	Examples	Brief Definition
Avionics	• Fixed-wing aircraft • Helicopters	Manned airborne platforms. Software that is complex and runs in real-time in embedded computer systems. It must often operates under interrupt control to process timelines in the nanoseconds.

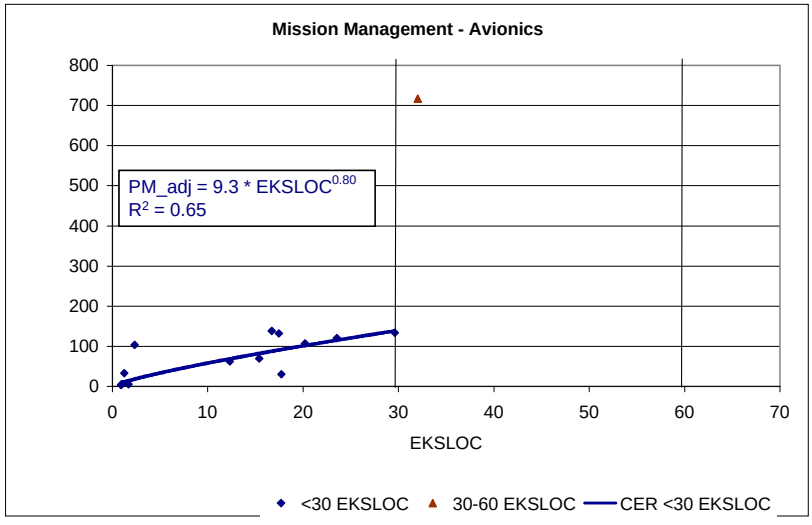
University of Southern California
Center for Systems and Software Engineering

Mission Management Analysis-2



University of Southern California
Center for Systems and Software Engineering

Mission Management Analysis-2



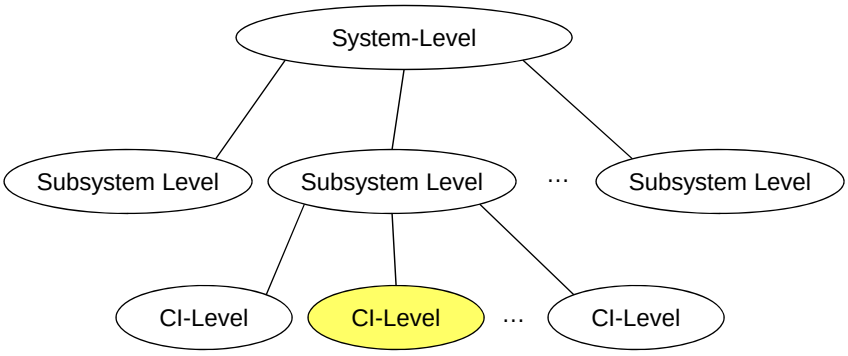
Simple Cost Estimating Relationships

	Communications - Fixed Ground	Mission Management - Avionics
CER	PM = 9.85 * EKSLOC$^{0.83}$ * PE	PM = 8.6 * EKSLOC$^{0.88}$ * PE
# Data Pts	32	19
EKSLOC Range	1.2 to 108	0.8 to 345
R^2	0.83 (was 0.62 with all env.)	0.83 (was 0.58 with all env.)
Standard Error of the Estimate	0.50	0.67
Bias	6%	-14%

Notes:
CER: Cost Estimating Relationship
PM: Person Months (152 labor hours / month)
EKSLOC: Equivalent Thousands of Source Lines of Code
R2: Correlation Coefficient
Bias: Average percentage error that estimate is above/below actual value

CERs in Cost Estimates

- Use the CERs at the level where the component size is within the data range of the CER
  - E.g. Configuration Item (CI) or lower
- Rollup the cost estimates



But what cost is missing from this methodology?

Conclusions

- Cost Estimating Relationships were improved by
  - Accounting for the impact of using pre-existing software on effort
  - Filling in missing effort data
  - Accounting for personnel experience
  - Segregating each Application Domain by the Operating Environment
    - Some Environments may have the same productivities
- Recommendations were made to improve the SRDR data collection form

Next Steps

- Work to establish a robust and cost effective software measurement collection process and a knowledge base that supports the data needs of the US DoD
  - Submitted change request to DCARC for SRDR
- We will publish the AFCAA Software Cost Estimation Metrics Manual to help analysts and decision makers develop accurate, easy and quick software cost estimates for avionics, space, ground, and shipboard platforms
  - Address data issues discussed in this presentation
  - Address integration cost issue

Manual Special Features-1

- Augment NCCA/AFCAA Software Cost Handbook:
  - Default Equivalent Size Inputs (DM, CM, IM, SU, UNFM)
  - Productivity Benchmarks by Operating Environment and Application Domain
  - Empirical Code, Effort, and Schedule Growth Measures derived from SRDRs
  - Empirically Based Cost Risk and Uncertainty Analysis Metrics
  - Calibrated SLIM-Estimate™ using most recent SRDR data
  - Mapping between COCOMO, SEER, True S cost drivers
  - Empirical Dataset for COCOMO, True S, and SEER Calibration
  - Software Maintenance Parameters

Manual Special Features-2

- Guidelines for reconciling inconsistent data
- Standard Definitions (Application Domain, SLOC, etc.)
- Address issues related to incremental development (overlaps, early-increment breakage, integration complexity growth, deleted software, impacts to maintenance) and version management (a form of product line development and evolution).
- Impact of Next Generation Paradigms
  - Model Driven Architecture
  - Net-Centricity
  - Systems of Systems

Questions?

For more information, contact:
Wilson Rosa
Wilson.Rosa@pentagon.af.mil
703-604-0395
Or
Brad Clark
bkclark@csse.usc.edu
703-754-0115
Or
Ray Madachy
rjmadach@nps.edu