



Software Development Cost: How Much? You Sure? Technical Summary

Jairus Hihn

Karen Lum

Jet Propulsion Laboratory

California Institute of Technology

jhihn@jpl.nasa.gov

Tim Menzies

Oussama Elrawas

Lane Department of Computer Science

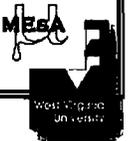
West Virginia University

tim@menzies.us

Dan Baker

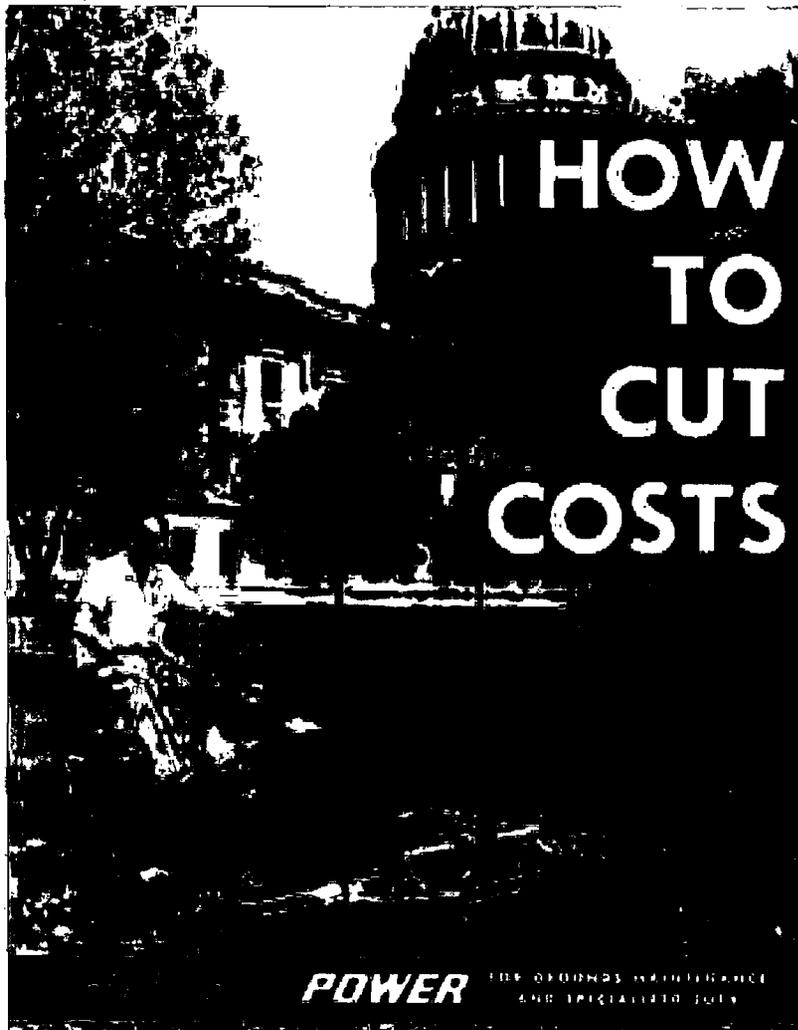
Omid Jalali

**NASA Software Assurance Symposium,
Morgantown, WV
September 25-27, 2007**

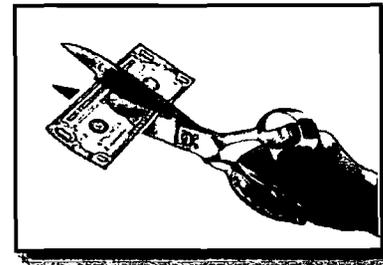




TASK Problem



- ✦ “Cost” is a quality issue
 - ✦ If development costs is underestimated...
 - ✦ ... developers will be forced into many quality-threatening cost-cutting measures.



- ✦ Can you think of anything you can do ...
 - ✦ ... that harms a project more...
 - ✦ ... than allocating insufficient resources?



Problem: Are we using the right models?

- A major reason for poor software cost estimation:
 - ... NASA's managers don't have information they need
 - Not enough relevant data
 - Current costing models are brittle and improperly tuned

- e.g. “officially”, COCOMO's tuning parameters vary

$$-2.5 \leq a \leq 2.94$$

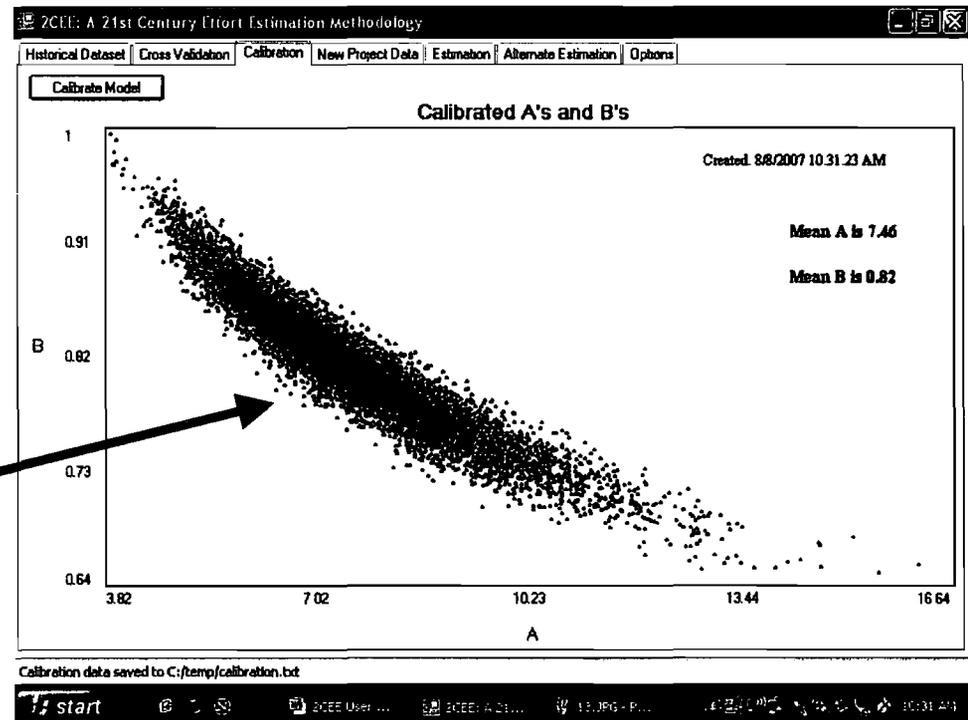
$$-0.91 \leq b < 1.01$$

- Which is nothing like what we see with real NASA data,

$$-3.5 \leq a \leq 14$$

$$-0.65 \leq b \leq 1$$

- Conclusion: we have to take far more care when building our cost models





Our Journey

- Well known throughout the software industry that software projects chronically overrun
- It became quickly apparent that there was a complete disconnect between the techniques used by estimation practitioners and the numerous ideas being addressed in the research community
- It also became clear that many fundamental estimation questions were not being addressed
 - What is a model's real estimation uncertainty?
 - How many records required to calibrate?
 - Answers have varied from 10-20 just for intercept and slope
 - If we do not have enough data what is the impact on model uncertainty
 - Data is expensive to collect and maintain so want to keep cost drivers and effort multipliers as few as possible
 - But what are the right ones?
 - When should we build domain specific models?
 - What are the best functional forms?
 - What are the best ways to tune/calibrate a model?



Our Journey Continued

- **Data mining techniques provided us with the rigorous tool set we needed to explore the many dimension of the problem we were addressing in a repeatable manner**
 - **Different Calibration and Validation Datasets**
 - **Analyze standard and non-standard models**
 - **Perform exhaustive searches over all parameters and records in order to guide data pruning**
 - **Rows (Stratification)**
 - **Columns (variable reduction)**
 - **Measure model performance by multiple measures**
 - **We have even been able to determine what performance measures are best.**



APPROACH: validated, relevant models that handle uncertainty

- Change the rules of the game
- Old way: reuse someone else's cost model
- New way: build relevant and validated models,
 - Used with uncertainty “what-if” queries
- Relevant: NASA historical data = table
 - Rows are projects (some of which aren't relevant)
 - Columns are project features (some which don't matter to you)
 - Row pruners: automatically find relevant projects
 - Column pruners: automatically dump uninformative features
- Validated:
 - In studies with real NASA data
 - This automatic method out-performed existing state-of-the-art methods
- Uncertainty & “what-ifs”
 - Planned projects aren't “one thing”
 - But a range of possible “things”
 - Explore the “things” looking for the range of possibilities

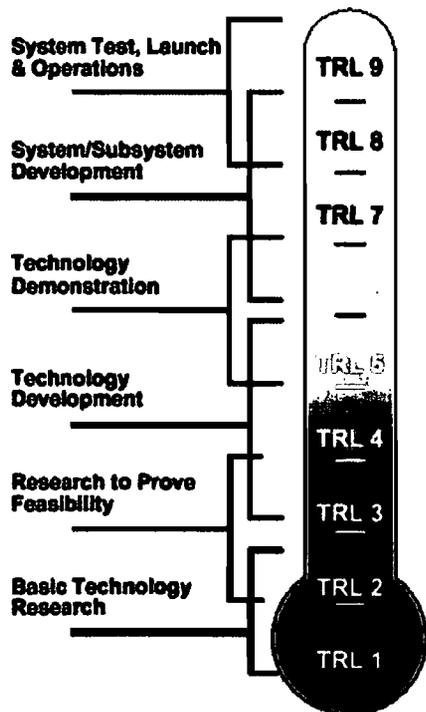
2011: A 21st Century Effort Estimation Methodology

Metric	Min	Mode	Max
flex	0	2.03	5.07
resl	0	2.83	6.01
team	0	1.1	4.66
pmat	2.84	4.68	7.8
rely	0.85	1.1	1.74
cplx	0.87	1.17	1.74
data	0.9	1	1.28
ruse	0.95	1	1.11
time	1	1	1.63
stor	1	1	1.46
pvol	0.87	0.87	1.3
acap	0.71	0.85	1.19
pcap	0.71	0.88	1
pcon	0.81	1	1.21
apex	0.81	1	1.22
plex	0.85	1	1.19
lrex	0.84	0.91	1.2
tool	0.83	1.17	1.17
ecod	1	1	1.43
site	0.8	0.93	1.09
docu	0.91	1	1.23
Logical EKLOC	0.81	90	200

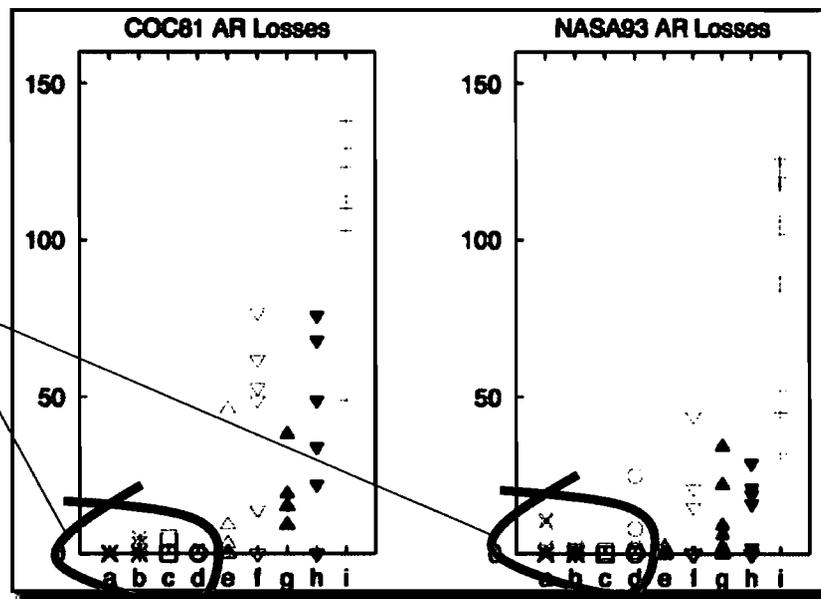
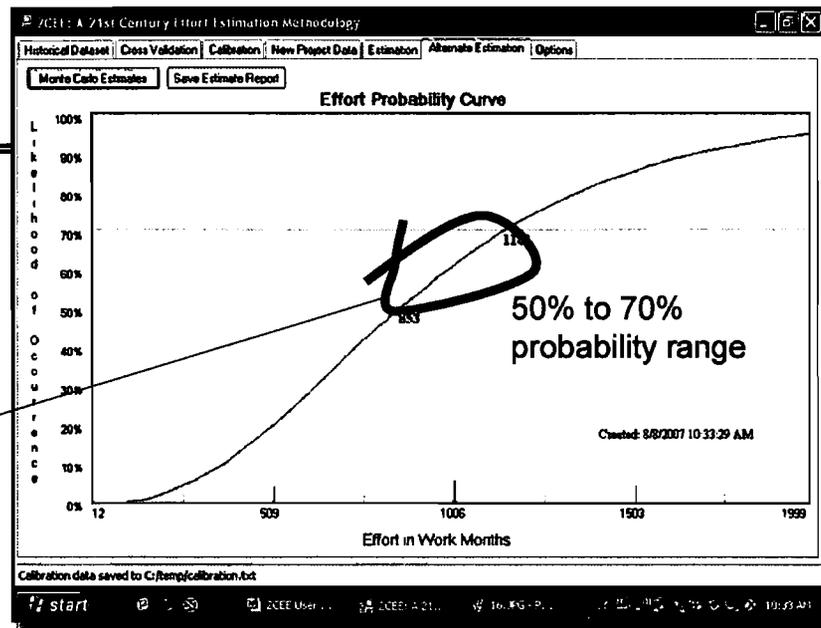


Accomplishments

- 2CEE (Windows application)
 - 21st century estimation
 - New high-water mark in cost estimation methods
 - Generates space of possible estimates .
TRL=7

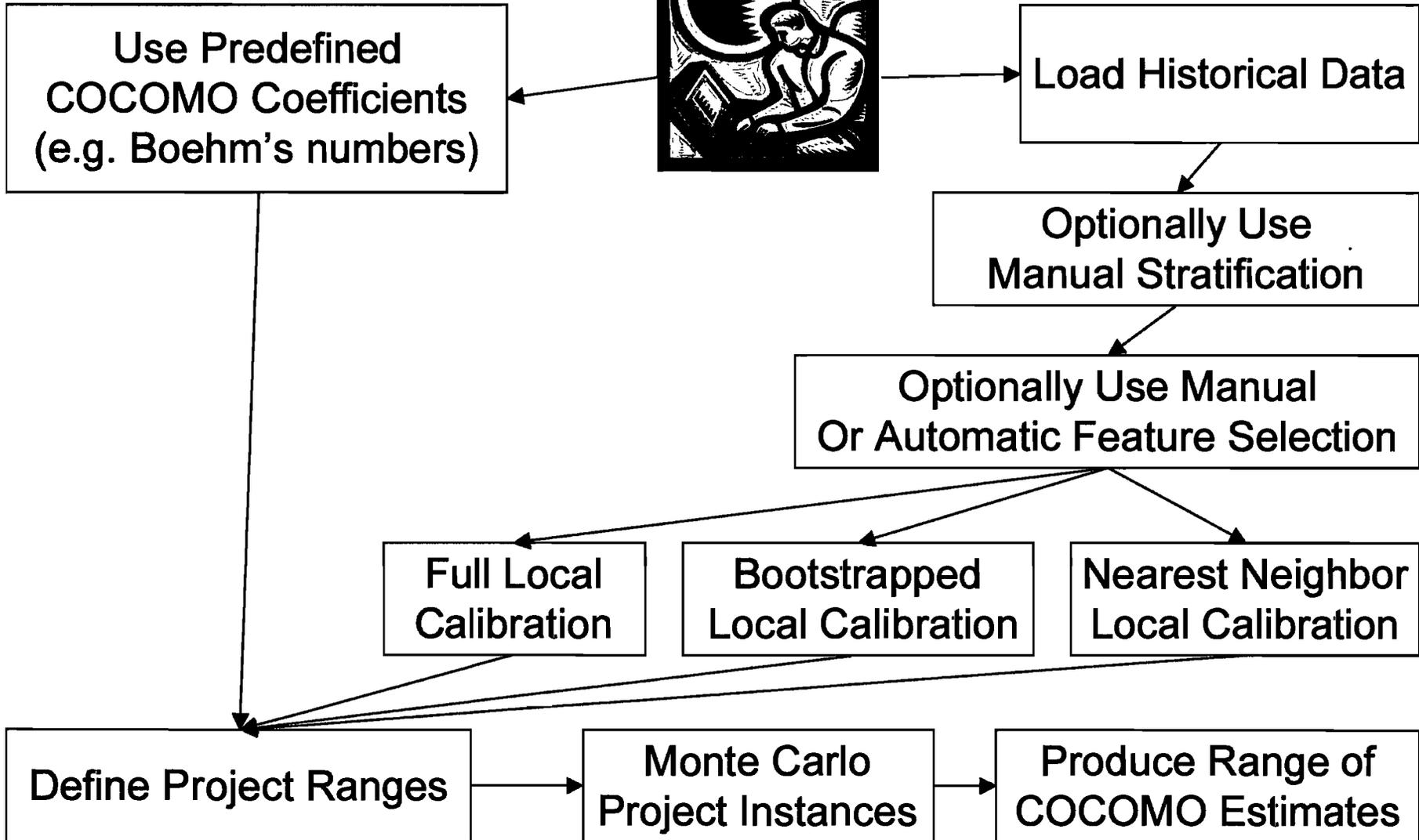


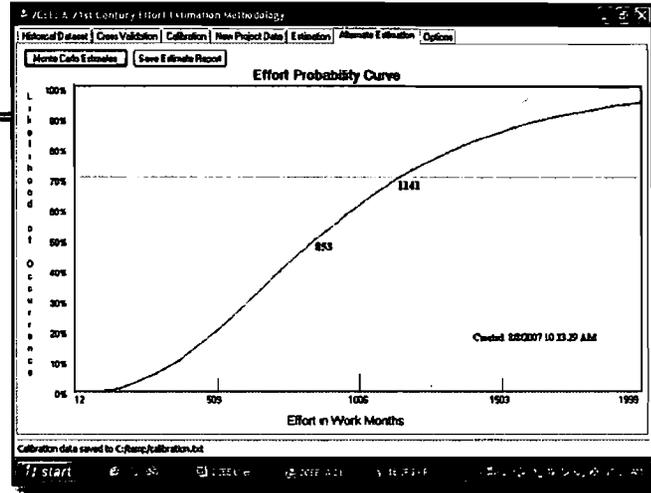
- Validated on NASA data
 - We have tried 158 ways to build cost models
 - Rejected 154 methods
 - Find the four row/column pruners that matter
 - Finds best models
- Deployed at NASA
 - Currently, undergoing a 12 month trial





2CEE





Run Demo

Following charts are
Back Up
for those who are reading the
charts on their own



2CEE Steps

Define Model Calibration



Evaluate with Cross Validation



Define Project Ranges



Monte Carlo Estimates



Historical Dataset | Cross Validation | Calibration | New Project Data | Estimation | Alternate Estimation | Options

File View Column Selection Row Selection

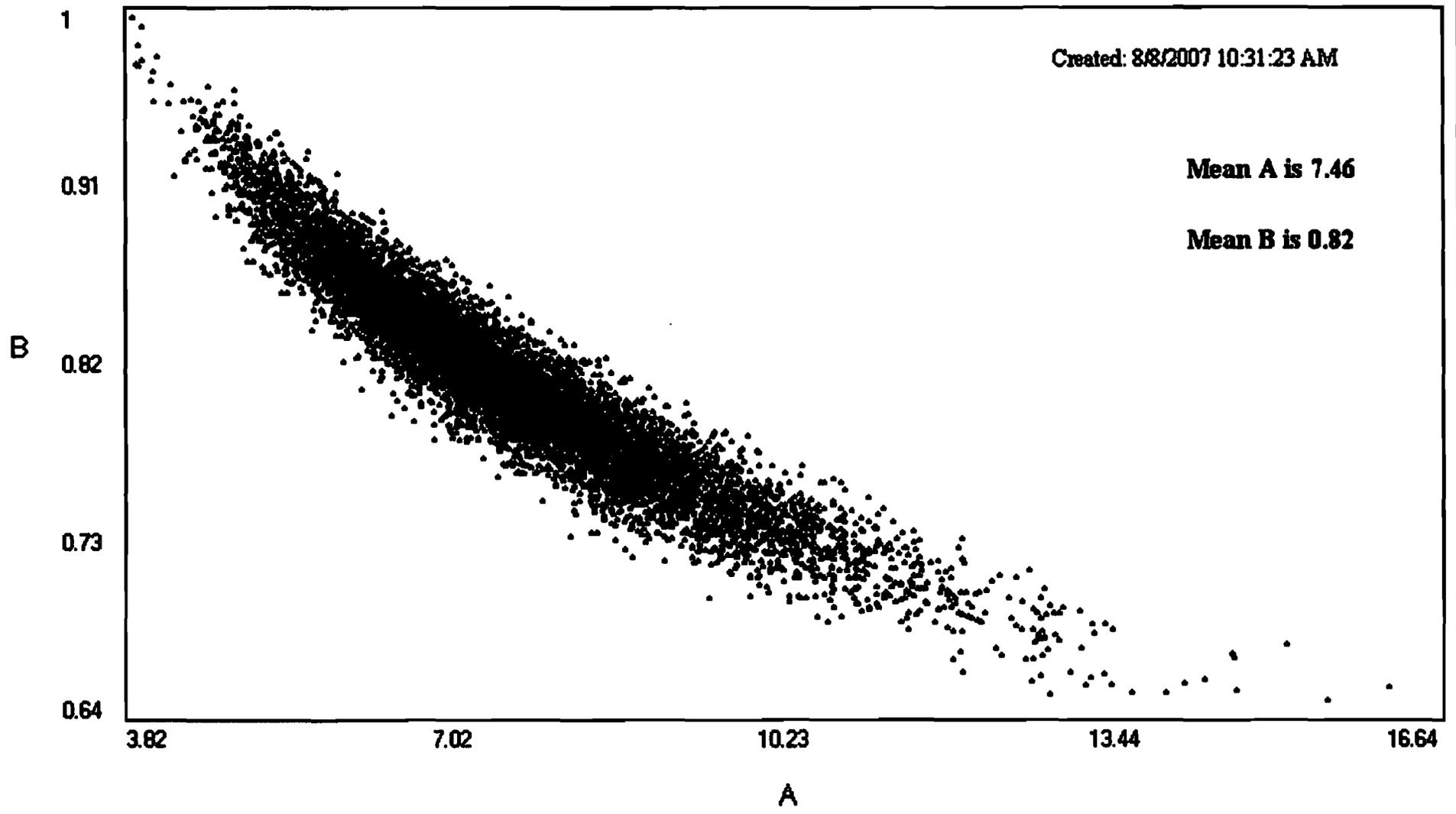
	RecordNumber	prec	flex	resl	team	pmat	rely	cpix	data	ruse	time	stor	pval
▶	1	2.48	2.03	2.83	1.1	4.68	1.1	1.17	0.9	1	1	1	0.87
	2	2.48	2.03	2.83	1.1	4.68	1.1	1.17	0.9	1	1	1	0.87
	3	2.48	2.03	2.83	1.1	4.68	1.1	1.17	0.9	1	1	1	0.87
	4	2.48	2.03	2.83	1.1	4.68	1.1	1.17	0.9	1	1	1	0.87
	5	2.48	2.03	2.83	1.1	4.68	1.1	1.17	0.9	1	1	1	0.87
	6	2.48	2.03	2.83	1.1	4.68	1.1	1.17	0.9	1	1	1	0.87
	7	2.48	2.03	2.83	1.1	4.68	1.1	1.17	0.9	1	1	1	0.87
	8	2.48	2.03	2.83	1.1	4.68	1.1	1.17	0.9	1	1	1	0.87
	9	2.48	2.03	2.83	1.1	4.68	1.1	1.17	0.9	1	1.63	1.46	0.87
	10	2.48	2.03	2.83	1.1	6.24	1	1.17	0.9	1	1	1	0.87
	11	2.48	2.03	2.83	1.1	6.24	1	1.17	0.9	1	1	1	0.87
	12	2.48	2.03	2.83	1.1	6.24	1	1.17	0.9	1	1	1	0.87
	13	2.48	2.03	2.83	1.1	6.24	1	1.17	0.9	1	1	1	0.87
	14	2.48	2.03	2.83	1.1	6.24	1	1.17	0.9	1	1	1	1.15
	15	2.48	2.03	2.83	1.1	6.24	1	1.17	0.9	1	1	1	0.87
	16	2.48	2.03	2.83	1.1	6.24	1	1.17	0.9	1	1	1	0.87
	17	2.48	2.03	2.83	1.1	6.24	1	1.17	0.9	1	1	1.46	0.87
	18	2.48	2.03	2.83	1.1	6.24	1	1.17	0.9	1	1	1	0.87
	19	2.48	2.03	2.83	1.1	6.24	1	1.17	0.9	1	1	1	0.87
	20	2.48	2.03	2.83	1.1	6.24	1	1.17	0.9	1	1	1.46	0.87

Begin by loading a dataset of prior software project's costing data.



Calibrate Model

Calibrated A's and B's



Calibration data saved to C:/temp/calibration.txt



Evaluating Performance

- **Requires Historical Project Data**
- **Uses N-Fold Cross Validation (also called Leave One Out Cross Validation)**
- **Plots an error curve**
- **Highlights theoretical model performance**

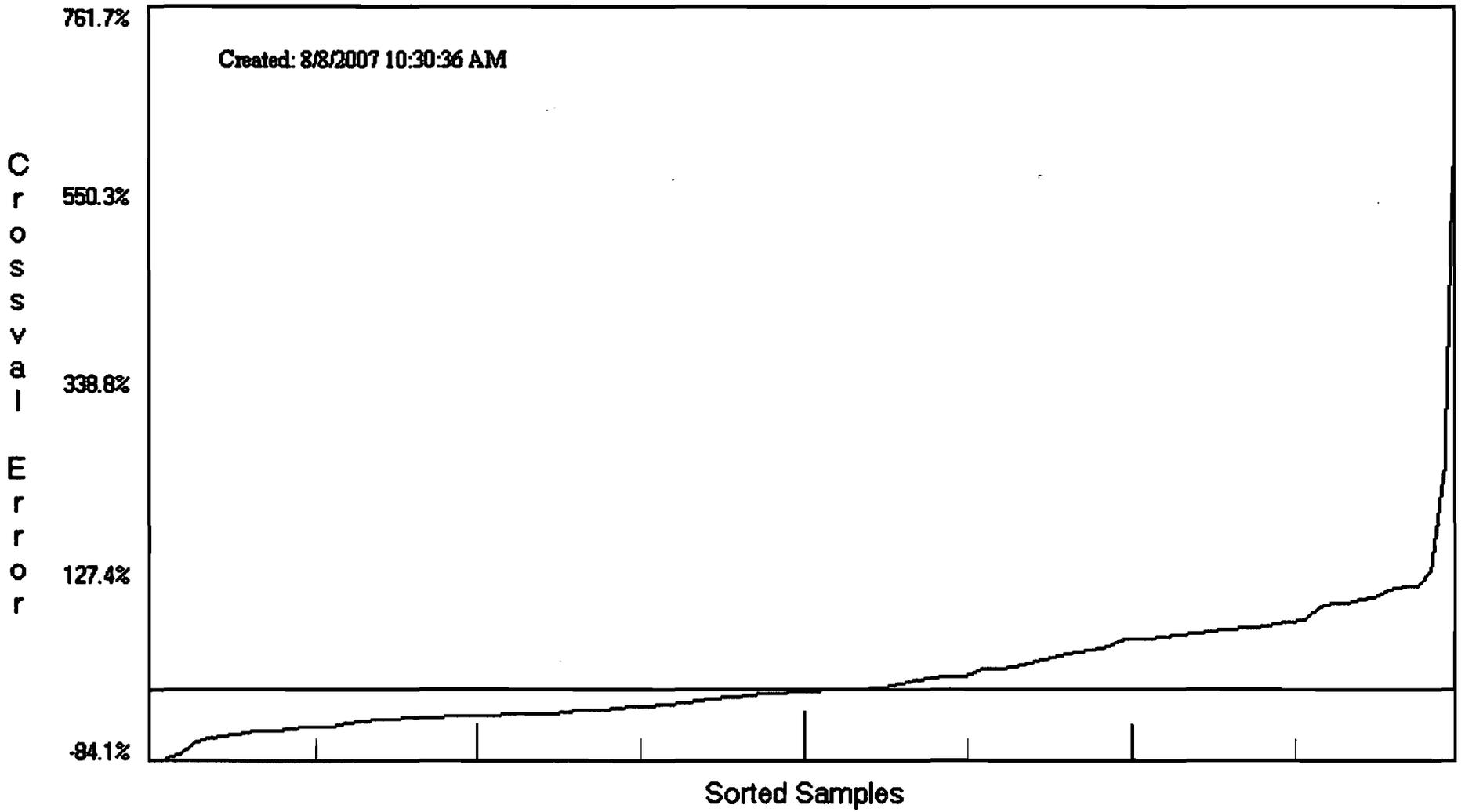


Historical Dataset | Cross Validation | Calibration | New Project Data | Estimation | Alternate Estimation | Options

Run Cross Validation

Show MRE Show RE

Sorted Relative Error



Cross Validation results saved to C:/temp/crossval.txt



Project Ranges

- Represents the knowledge of the project to estimate
- Each feature is represented by a range
- Each range has a minimum, mode, max
- Default ranges are generated from historical data
- Ranges should be modified with new project knowledge
- Allows for a range of estimates instead of a single point
- Allows estimates to be made very early in the life cycle and improved as project knowledge is gained

Historical Dataset Cross Validation Calibration New Project Data Estimation Alternate Estimation Options

- Save Project Ranges
- Load Project Ranges
- Reset to Defaults

Metric	Min	Mode	Max
flex	0	2.03	5.07
resl	0	2.83	6.01
team	0	1.1	4.66
pmat	2.84	4.68	7.8
rely	0.85	1.1	1.74
cplx	0.87	1.17	1.74
data	0.9	1	1.28
ruse	0.95	1	1.11
time	1	1	1.63
stor	1	1	1.46
pvol	0.87	0.87	1.3
acap	0.71	0.85	1.19
pcap	0.71	0.88	1
pcon	0.81	1	1.21
apex	0.81	1	1.22
plex	0.85	1	1.19
ltex	0.84	0.91	1.2
tool	0.83	1.17	1.17
sced	1	1	1.43
site	0.8	0.93	1.09
docu	0.91	1	1.23
▶ Logical EKLOC	0.81	90	200

Calibration data saved to C:/temp/calibration.txt



Estimation

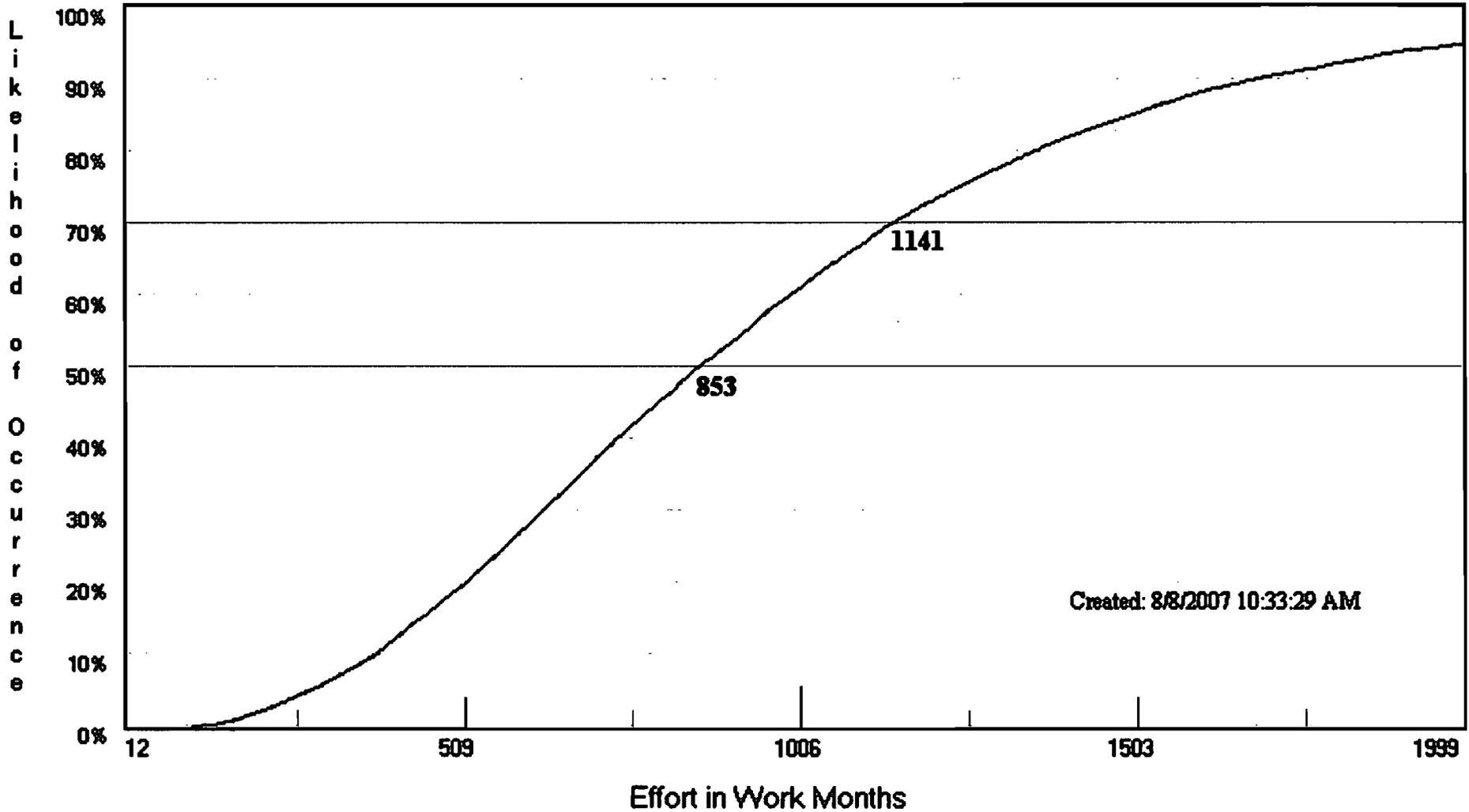
- Randomly samples instances from project ranges (10,000 times by default)
- If calibrating with bootstrapped LC, sample from the set of a,b pairs
- If calibrating with nearest neighbor, the coefficients are calibrated to be near the randomly generated sample instance
- Produces a range of estimates (S-Curve)
- Estimates are suggested from the 50th to 70th percentile

Historical Dataset | Cross Validation | Calibration | New Project Data | Estimation | Alternate Estimation | Options

Monte Carlo Estimates

Save Estimate Report

Effort Probability Curve



Calibration data saved to C:/temp/calibration.txt

End of File

