The 47th CREST Open Workshop - CREST 10th Anniversary

Multi-objective Prediction

Federica Sarro

Senior Research Associate
Dept. of Computer Science, CREST
University College London

f.sarro@ucl.ac.uk

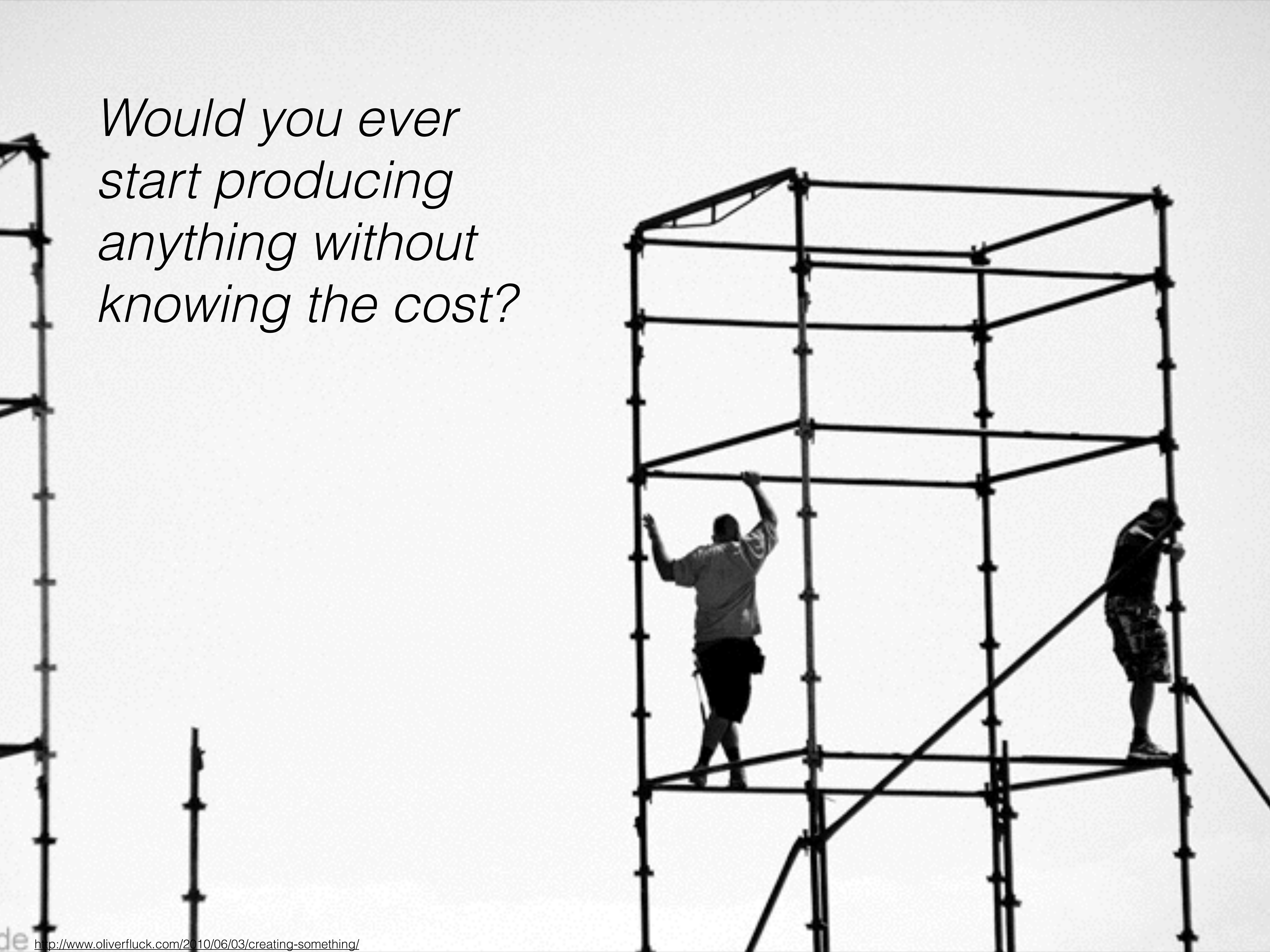
7th Sept 2016

Annual "Humies" Awards
For Human-Competitive Results
Produced By Genetic And Evolutionary Computation

Multi-objective Software Effort Estimation

F. Sarro*, A. Petrozziello**, M. Harman*

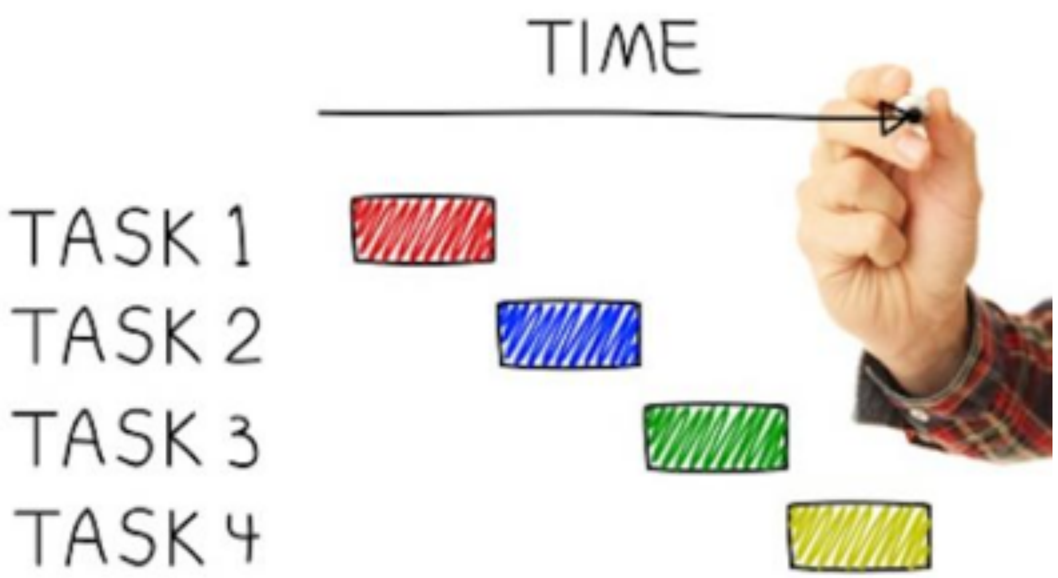
*CREST, Department of Computer Science, University College London, UK
** School of Computing, University of Portsmouth, UK


Would you ever start producing anything without knowing the cost?
http://www.oliverfluck.com/2010/06/03/creating-something/

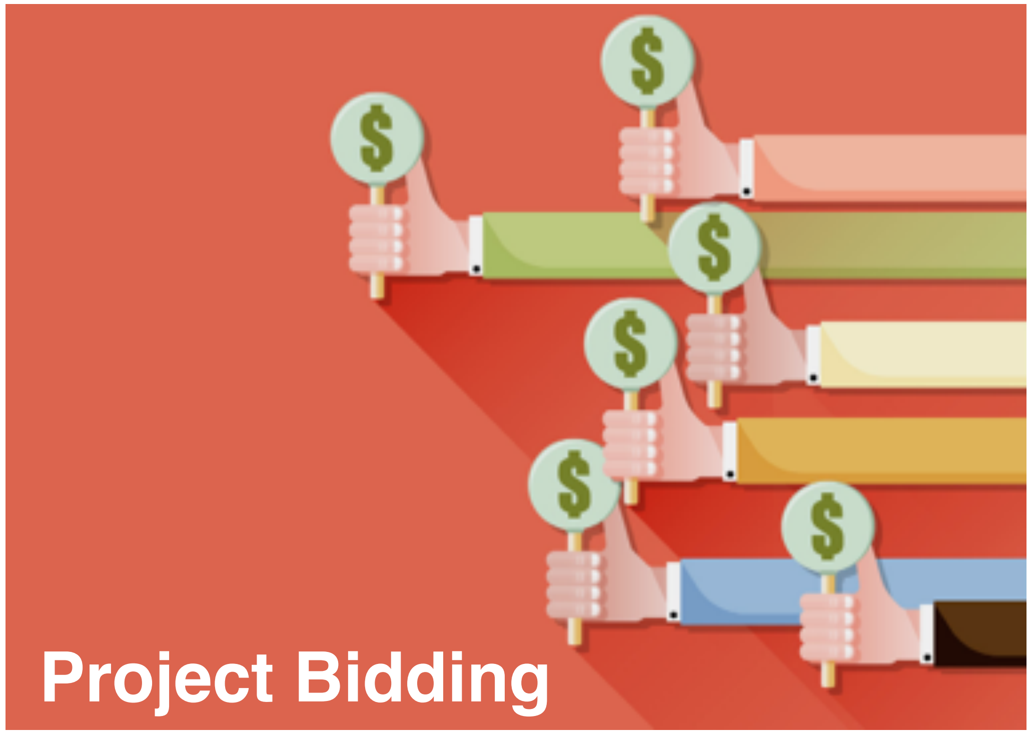
Software Effort Estimation (SEE)

Process of predicting the most realistic amount of effort required to realise a software project

(effort usually quantified in person-hours/person-months)



Project Scheduling /Staffing



Options for Estimation

Experts tend to under-estimate

What is the margin of error?

Predictions of project effort lie within **30**%-**40**% of its true value

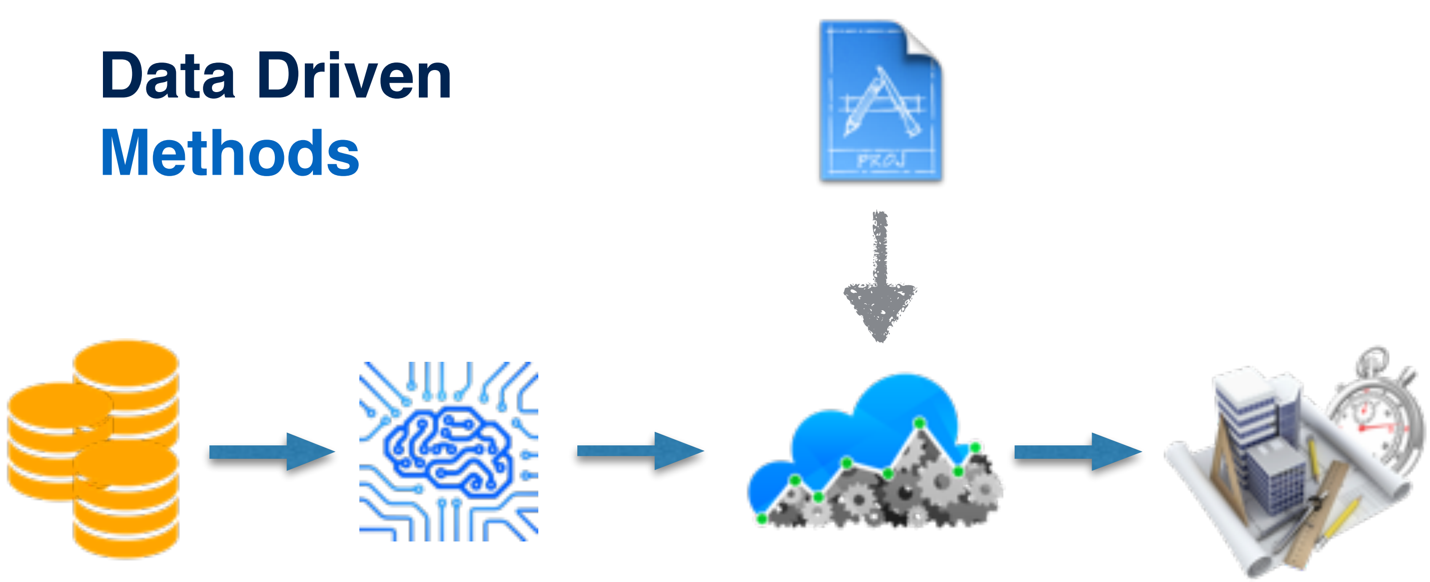
K. Molkken and M. Jorgensen. A review of surveys on software effort estimation. ISESE'03.
S. McConnell. Software Estimation: Demystifying the Black Art. Microsoft Press, 2006

Options for Estimation

Expert Judgment

Experts tend to under-estimate within 30%-40% of the true value

Data Driven Methods

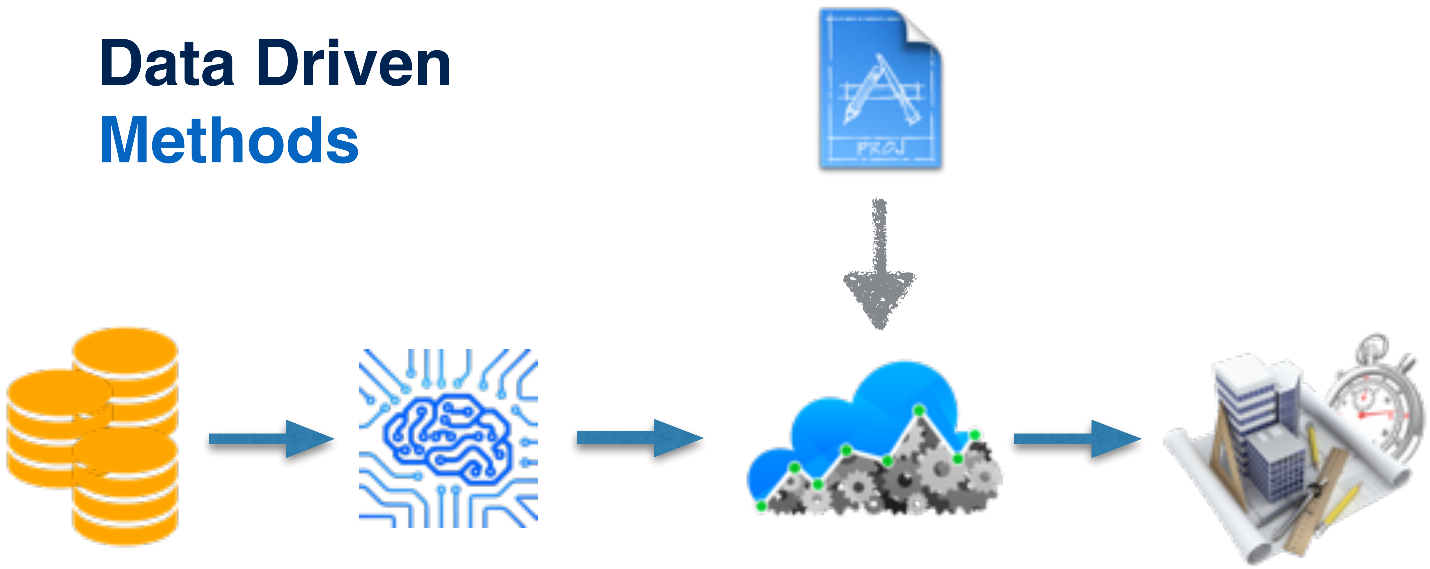


- Regression-based
- Analogy-based
- Search-based

Options for Estimation

After ~30 years of research…

Data Driven
Methods



- Linear Regression
- Stepwise Regression
- Support Vector Regression
- Classification and Regression Trees
- Case-based Reasoning
- K-Nearest Neighbours
- Genetic Algorithms
- Genetic Programming
- Tabu Search
- Simulated Annealing
- …

… data-driven methods are still unable to beat human-estimates!

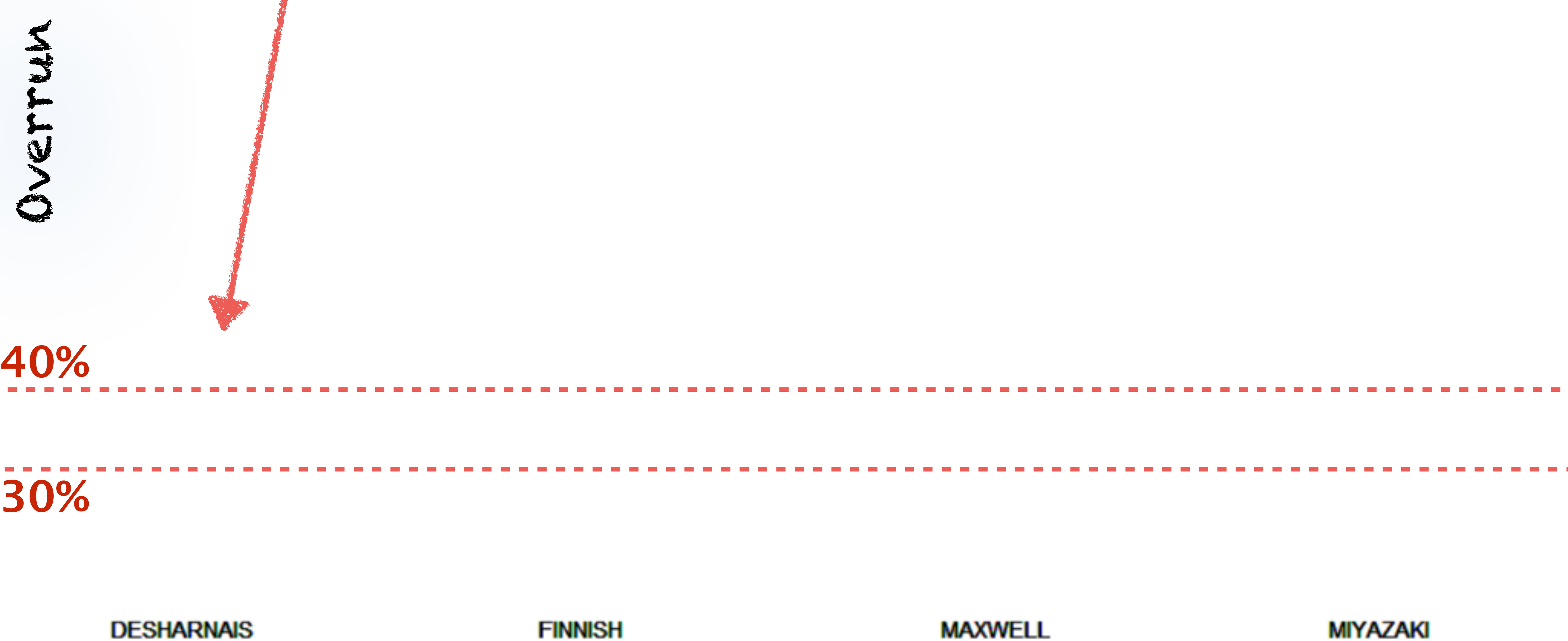
RQ4. Comparison to Industrial Practices

How does our approach, CoGEE, compare to human-expert-based estimates?

RQ4. Comparison to Industrial Practices

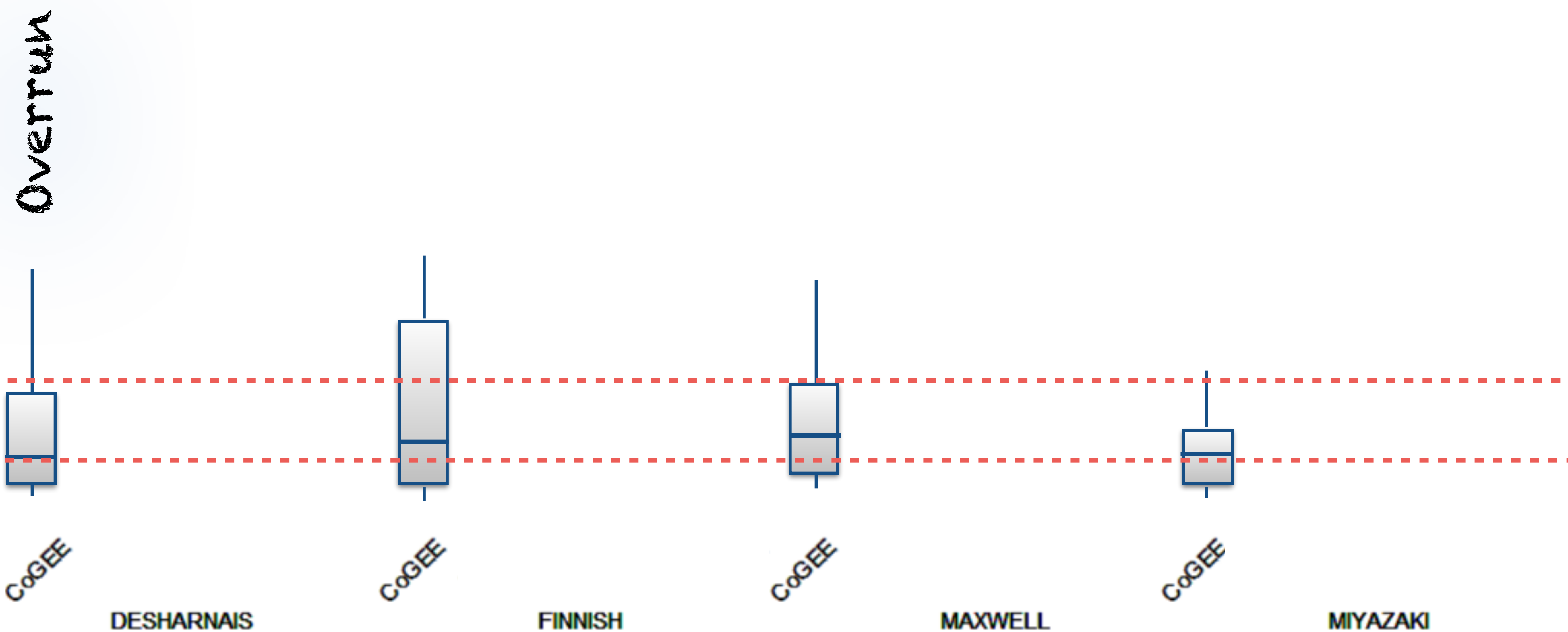
Human-expert-based predictions of project effort lie within 30% and 40% of the true value (overrun)

The evidence for these **thresholds** comes from a **survey of industry practices** by Molkken and Jørgensen



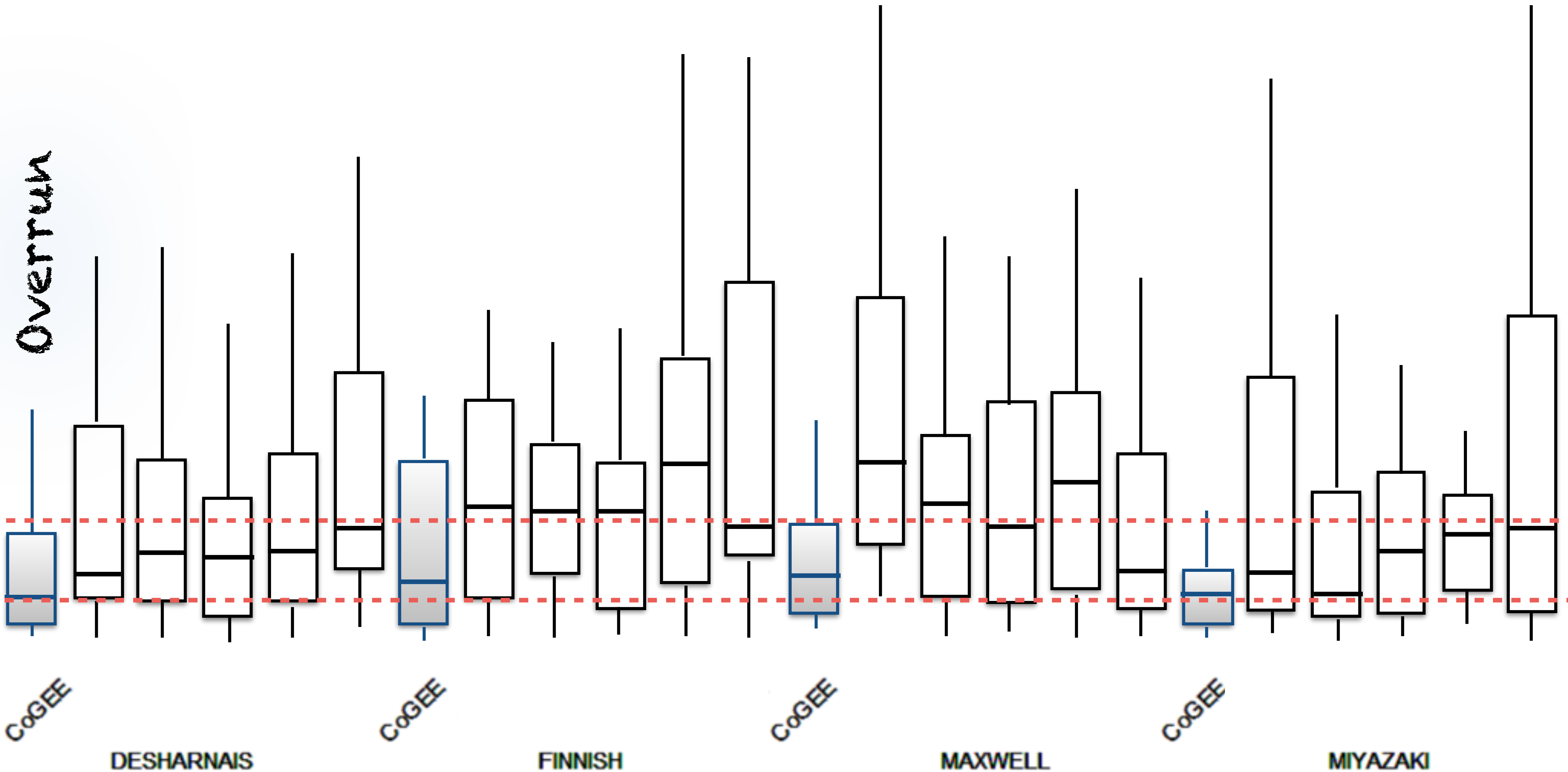
RQ4. Our Results are Human-Competitive

CoGEE provides effort estimates similar or better than those provided by human-experts



RQ2. Comparison with State-of-the-Art Benchmark

CoGEE outperforms popular automated estimation methods proposed over the last 30 years

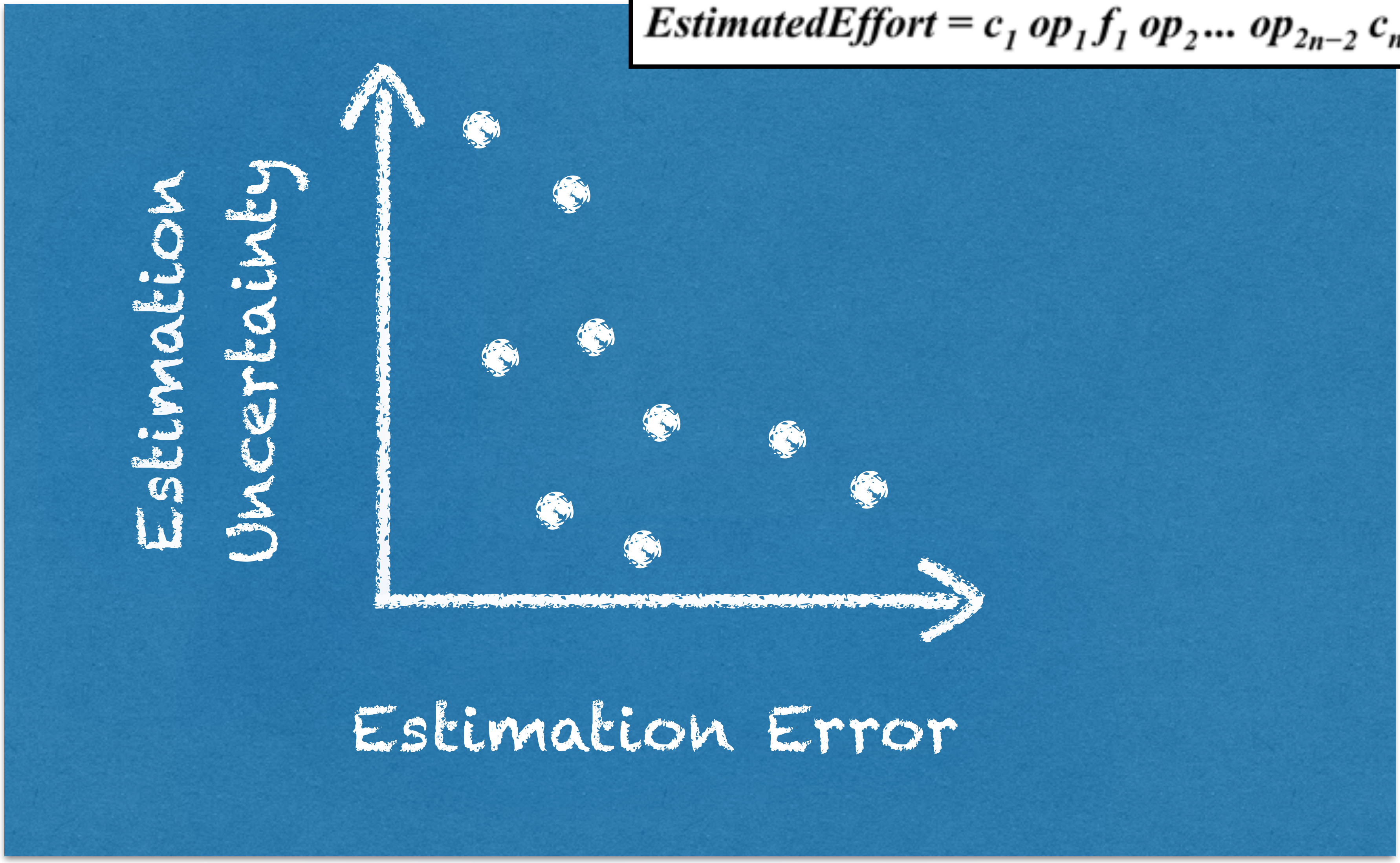


CoGEE: Confidence Guided Effort Estimator

CoGEE is a multi-objective evolutionary approach

that builds robust estimation models

$$EstimatedEffort = c_1\ op_1\, f_1\ op_2 \ldots\ op_{2n-2}\ c_n\ op_{2n-1} f_n\ op_{2n}\ C$$

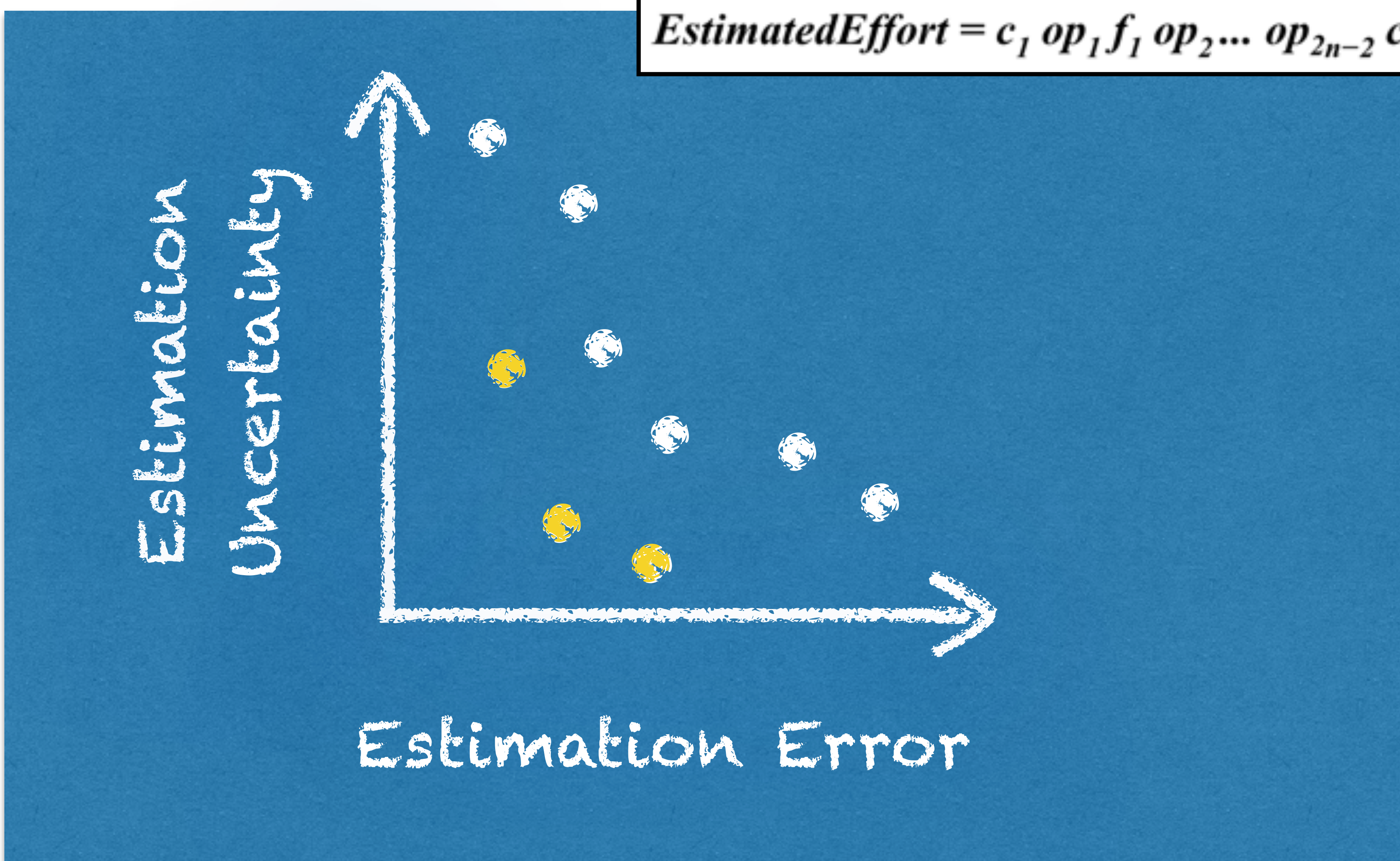


CoGEE: Confidence Guided Effort Estimator

CoGEE is a multi-objective evolutionary approach

that builds robust estimation models

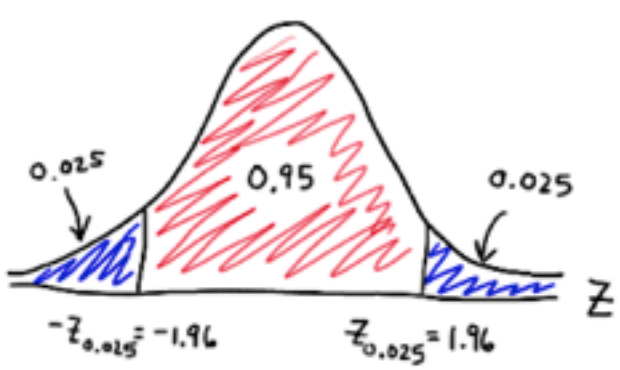
$$EstimatedEffort = c_1\ op_1\ f_1\ op_2 \ldots\ op_{2n-2}\ c_n\ op_{2n-1}\ f_n\ op_{2n}\ C$$

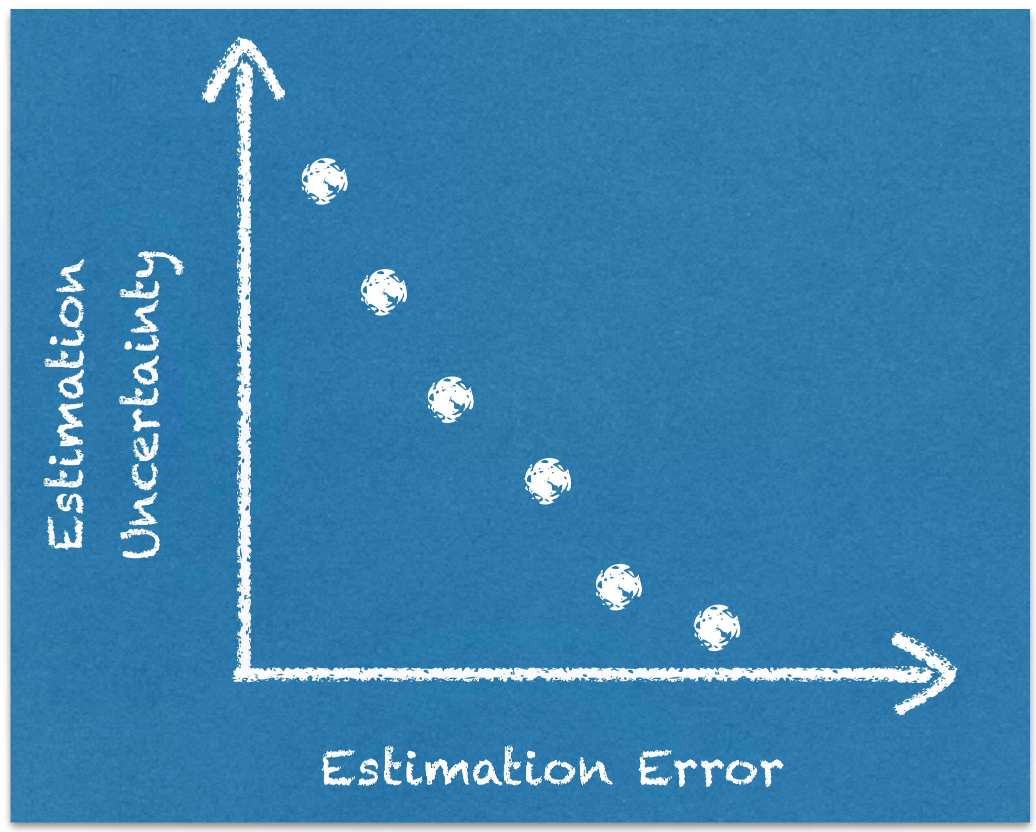


CoGEE: Confidence Guided Effort Estimator

Bi-objective estimation

Confidence Interval





Sum of Absolute Error

$$SAE = \sum_{i=1}^{N} |RealEffort_i - EstimaedEffort_i|$$

RQ1.
Sanity Check

RQ2. State of the Art Benchmark

RQ3. Benefits from Multi-objective Formulation

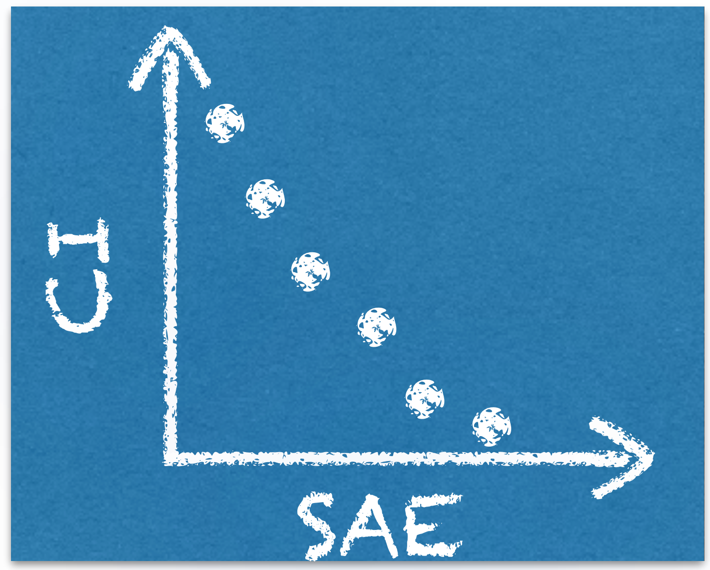
RQ4. Comparison to Industrial Practices

RQ3. Benefits from Multi-objective Formulation

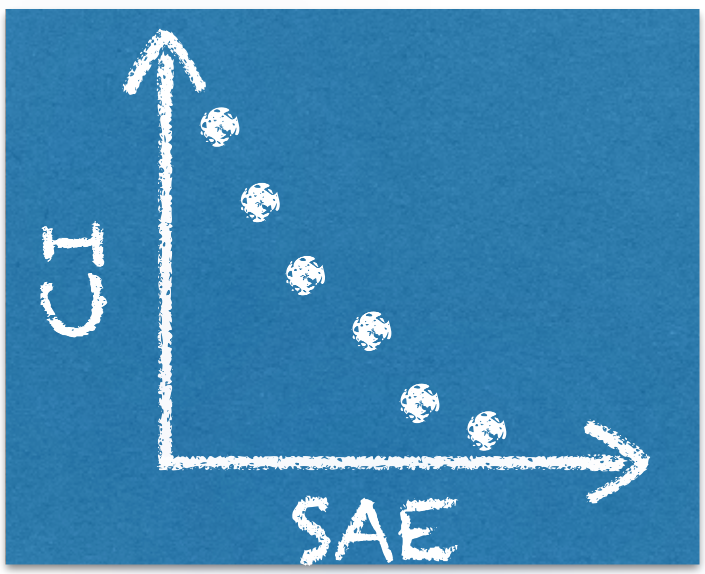
Does **CoGEE** provide more accurate estimates than *alternative* single and multi-objective *approaches*?

RQ3. Benefits from Multi-objective Formulation

GA-SAE



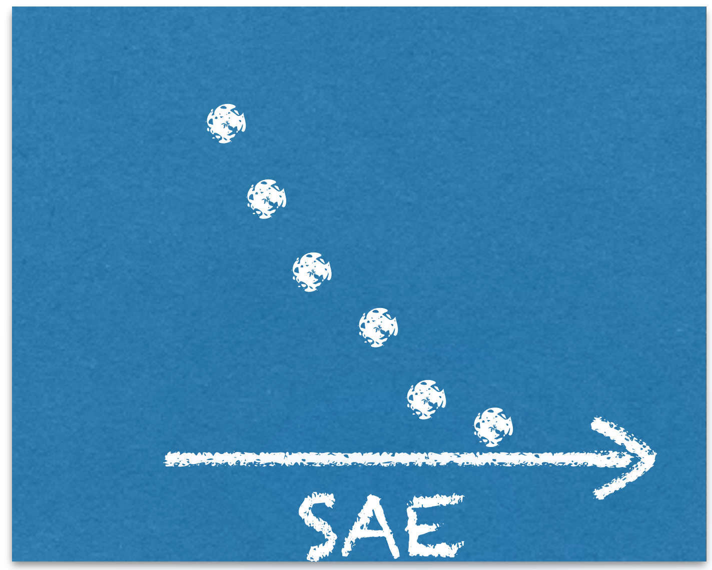
CoGEE



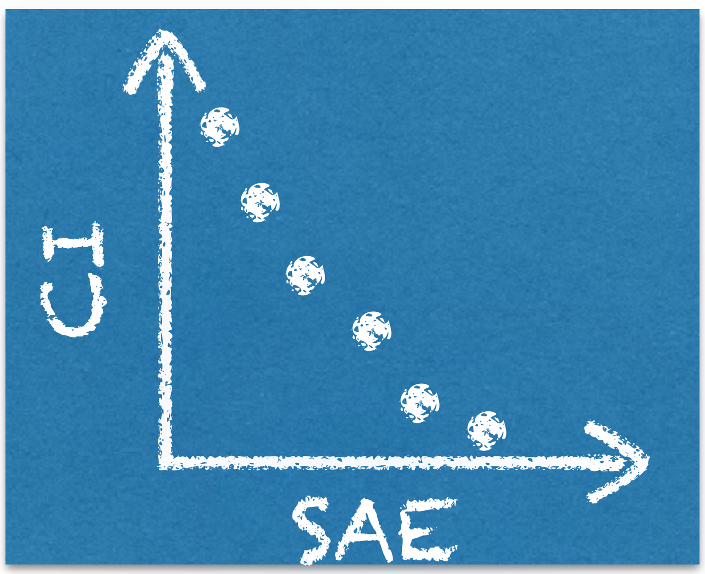
GA-CI

RQ3. Benefits from Multi-objective Formulation

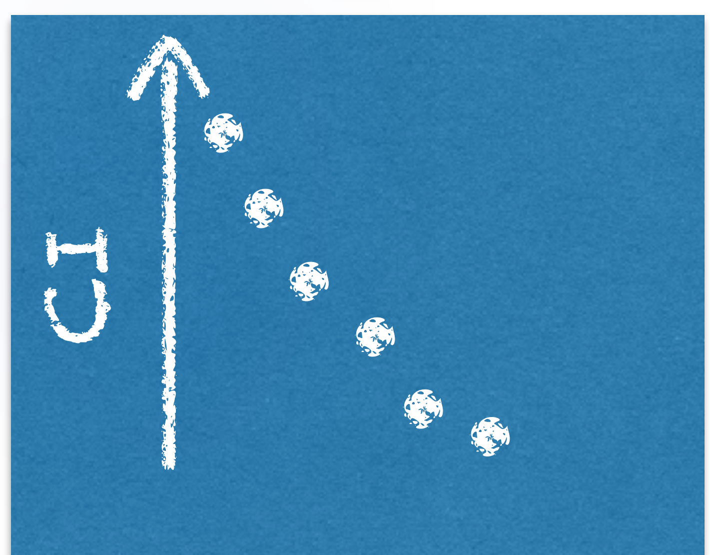
GA-SAE



CoGEE



GA-CI



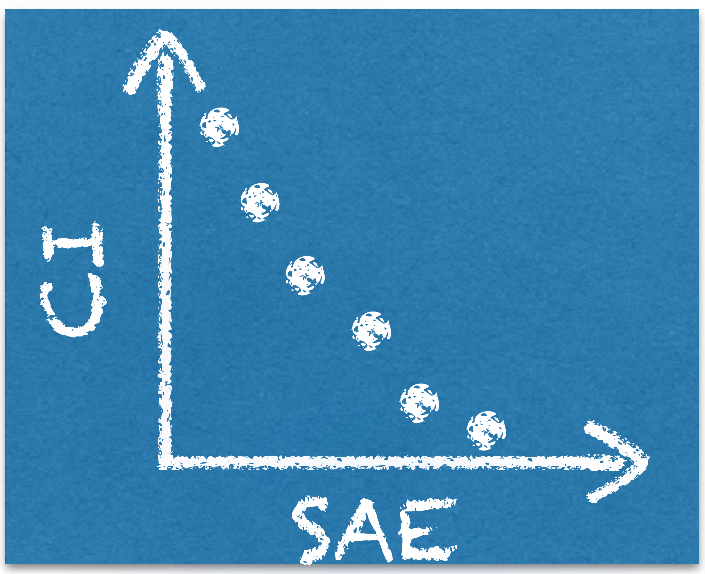
$$SAE = \sum_{i=1}^{N} |RealEffort_i - EstimaedEffort_i|$$

Underestimates
RealEffort - EstimatedEffort > 0

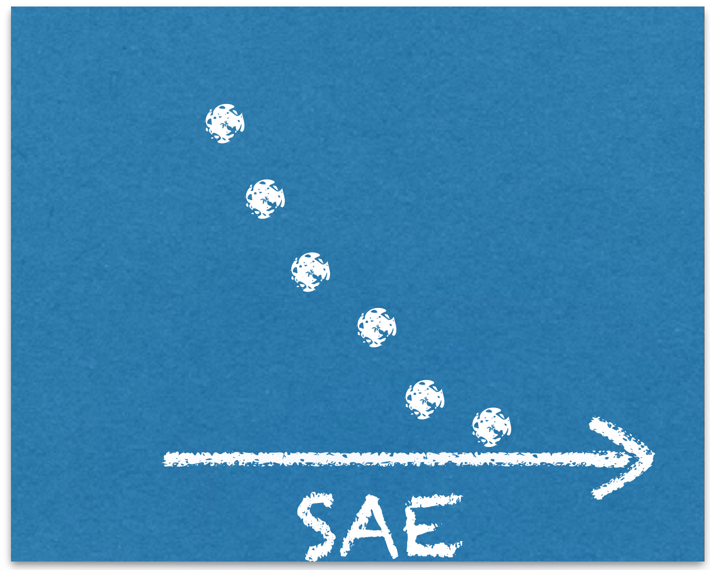
Overestimates
RealEffort - EstimatedEffort < 0

RQ3. Benefits from Multi-objective Formulation

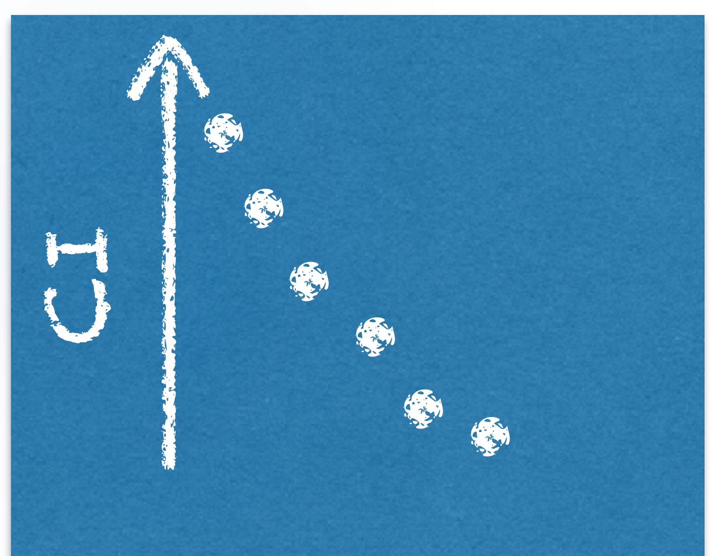
CoGEE



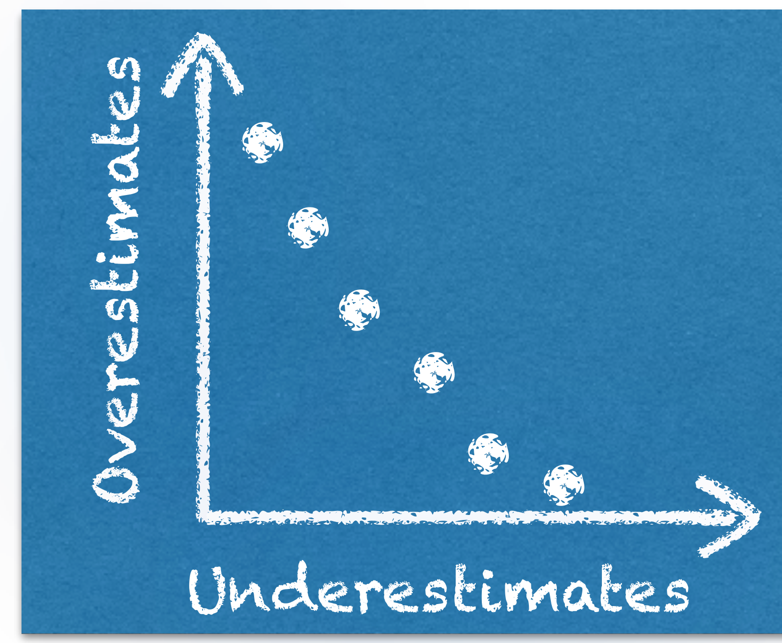
GA-SAE



GA-CI



NSGA-UO



RQ3. Benefits from Multi-objective Formulation

Dataset	Technique	I_{GD}	I_{HV}	I_C
China	CoGEE vs. GA-SAE	0.440 (0.49)	0.010 (0.60)	0.010 (0.42)
	CoGEE vs. GA-CI	0.830 (0.71)	0.010 (0.66)	0.010 (0.71)
	CoGEE vs. NSGAII-UO	<0.001 (0.26)	<0.001 (0.73)	<0.001 (0.94)
Desharnais	CoGEE vs. GA-SAE	0.010 (0.39)	<0.001 (0.33)	0.997 (0.39)
	CoGEE vs. GA-CI	0.005 (0.39)	<0.001 (0.88)	0.040 (0.66)
	CoGEE vs. NSGAII-UO	0.680 (0.56)	0.180 (0.56)	0.780 (0.44)
Finnish	CoGEE vs. GA-SAE	<0.001 (0.32)	0.130 (0.55)	0.640 (0.69)
	CoGEE vs. GA-CI	0.430 (0.51)	<0.001 (0.65)	0.920 (0.56)
	CoGEE vs. NSGAII-UO	<0.001 (0.70)	<0.001 (0.82)	<0.001 (0.71)
Maxwell	CoGEE vs. GA-SAE	<0.001 (0.08)	<0.001 (0.63)	<0.001 (0.98)
	CoGEE vs. GA-CI	<0.001 (0.25)	<0.001 (0.70)	<0.001 (0.92)
	CoGEE vs. NSGAII-UO	0.094 (0.03)	0.700 (0.72)	0.470 (0.53)
Miazaky	CoGEE vs. GA-SAE	<0.001 (0.25)	<0.001 (1.00)	<0.001 (1.00)
	CoGEE vs. GA-CI	<0.001 (0.25)	<0.001 (1.00)	<0.001 (1.00)
	CoGEE vs. NSGAII-UO	<0.001 (0.26)	<0.001 (0.95)	<0.001 (0.95)

RQ3. Results of the Wilcoxon test (Â12 effect size) which compare the quality indicators (I_{GD} , I_{HV}, I_C) of CoGEE to the ones of the other evolutionary approaches over 30 runs.

Multi-objective Software Effort Estimation

F. Sarro, A. Petrozziello, M. Harman

Human-competitive results to a long-standing and difficult problem

Advances the state of the art

Successful use of EC in Software Engineering

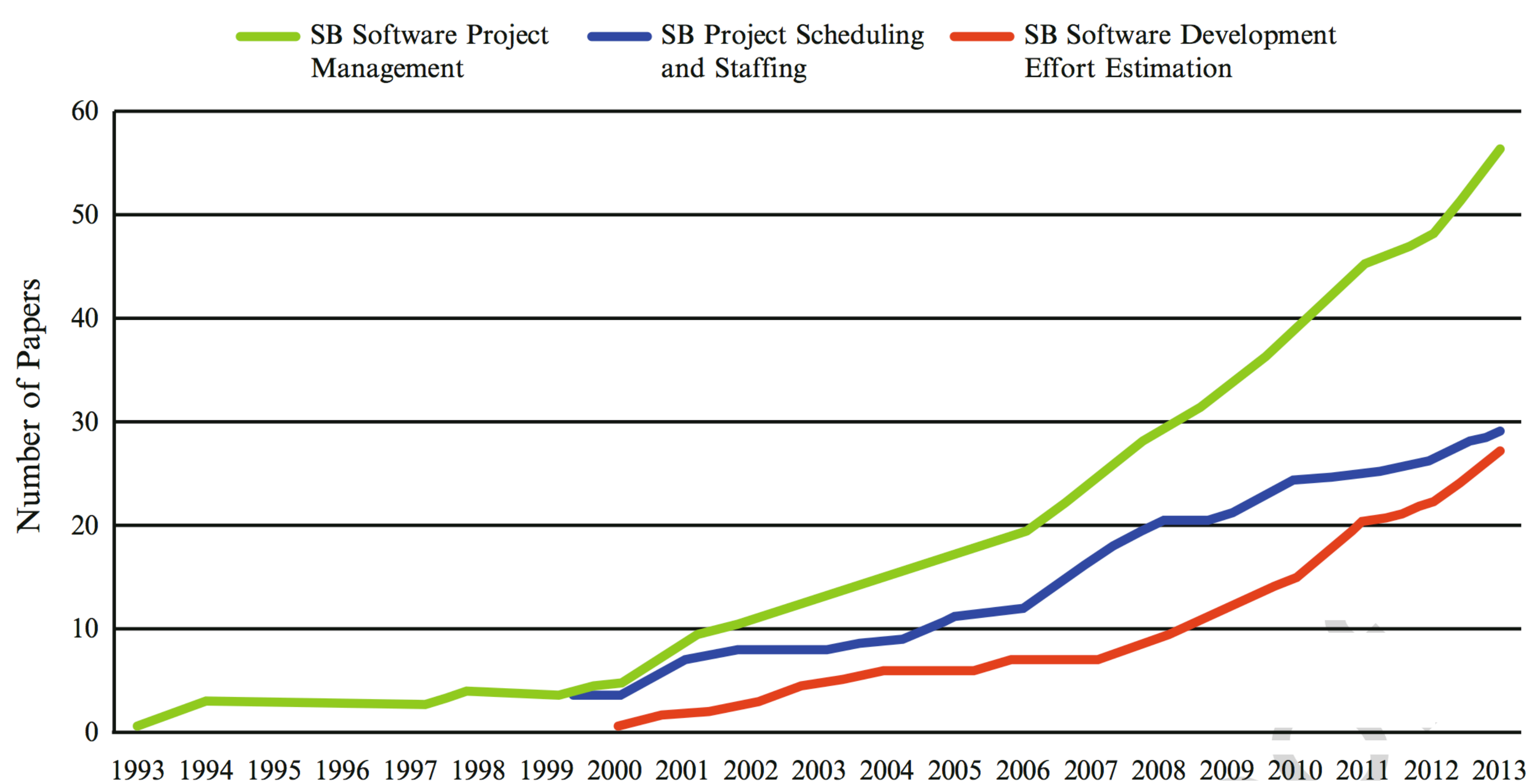
Through Empirical Study (724 real-word projects)

Breakthrough results published in ICSE'16 and awarded at the HUMIES-GECCO'16

Want to Know More about Search-Based Software Effort Estimation?

Chapter 15
Search-Based Software Project Management

Filomena Ferrucci, Mark Harman, and Federica Sarro



Mutation-Aware Fault Prediction

David Bowes[*], Tracy Hall[†], Mark Harman[‡], Yue Jia[‡], Federica Sarro[‡], and Fan Wu[‡]
[*]University of Hertfordshire, Hatfield, UK [†]Brunel University London, Uxbridge, UK
[‡]University College London, London, UK

ABSTRACT

We introduce mutation-aware fault prediction, which leverages additional guidance from metrics constructed in terms of mutants and the test cases that cover and detect them. We report the results of 12 sets of experiments, applying 4 different predictive modelling techniques to 3 large real-world systems (both open and closed source). The results show that our proposal can significantly ($p \leq 0.05$) improve fault prediction performance. Moreover, mutation-based metrics lie in the top 5% most frequently relied upon fault predictors in 10 of the 12 sets of experiments, and provide the majority of the top ten fault predictors in 9 of the 12 sets of experiments.

HIGHLIGHTS

Important gaps addressed:

- Testing information rarely used
- Industrial data rarely used

Promising findings:

- Adding mutation information improves predictive performance
- Worthwhile effect sizes occur

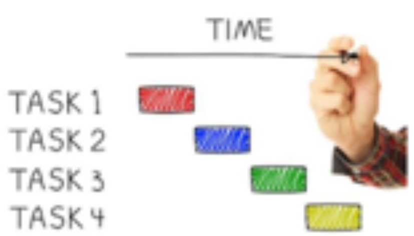
Multi-Objective Software Effort Estimation

F. Sarro, A. Petrozziello, M. Harman

Software Effort Estimation

Process of predicting the most realistic amount of effort required to realise a software project

(effort usually quantified in person-hours/person-months)



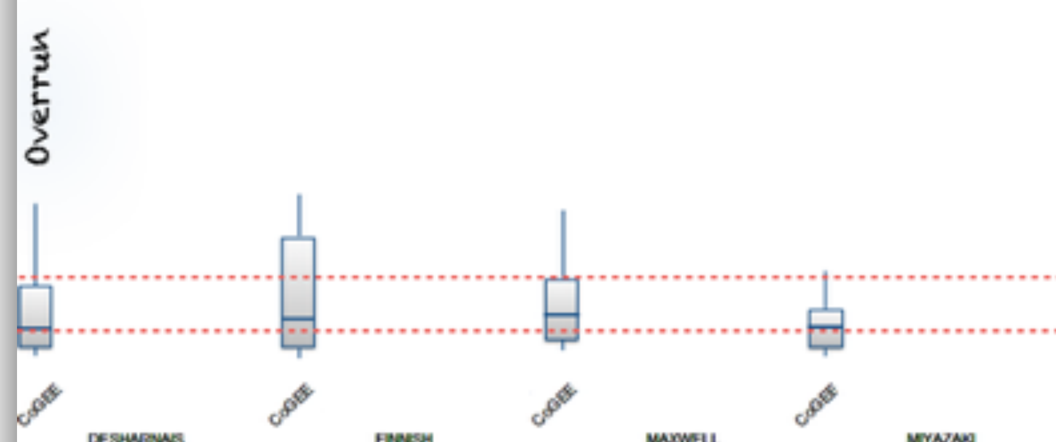
Project Scheduling /Staffing

Confidence Guided Effort Estimation (CoGEE)

CoGEE is a multi-objective evolutionary approach that builds robust estimation models

RQ4. Our Results are Human-Competitive

CoGEE provides effort estimates similar or better than those provided by human-experts



RQ2. Comparison with State-of-the-Art Benchmark

CoGEE outperforms popular automated estimation methods proposed over the last 30 years