Annual "Humies" Awards
For Human-Competitive Results

Produced By Genetic And Evolutionary Computation

Multi-Objective Software Effort Estimation

Federica Sarro

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

f.sarro@ucl.ac.uk

@f_sarro

Multi-Objective Software Effort Estimation

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

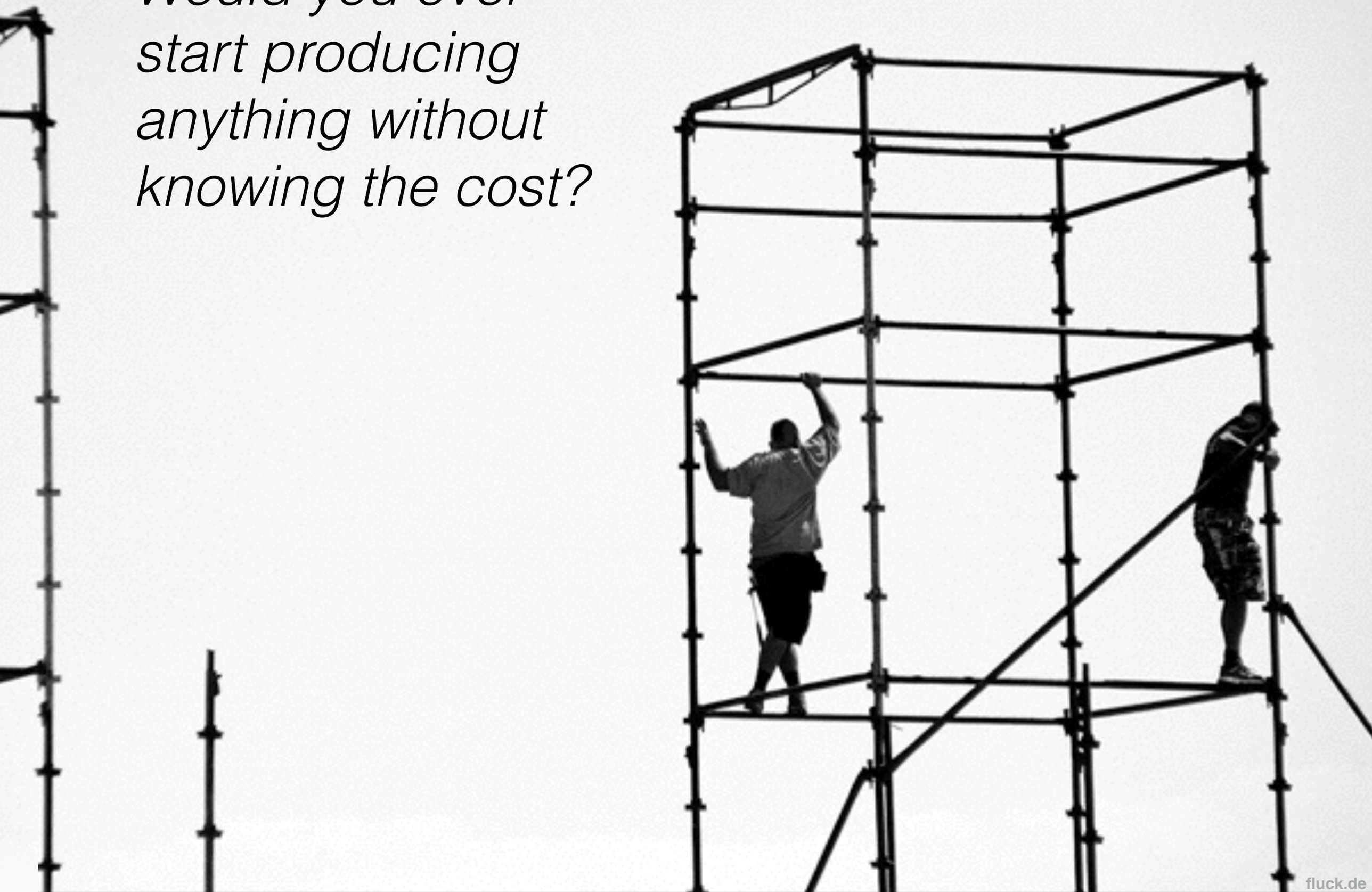
*CREST, Department of Computer Science, University College London, UK
** School of Computing, University of Portsmouth, UK

Software Development Effort Estimation

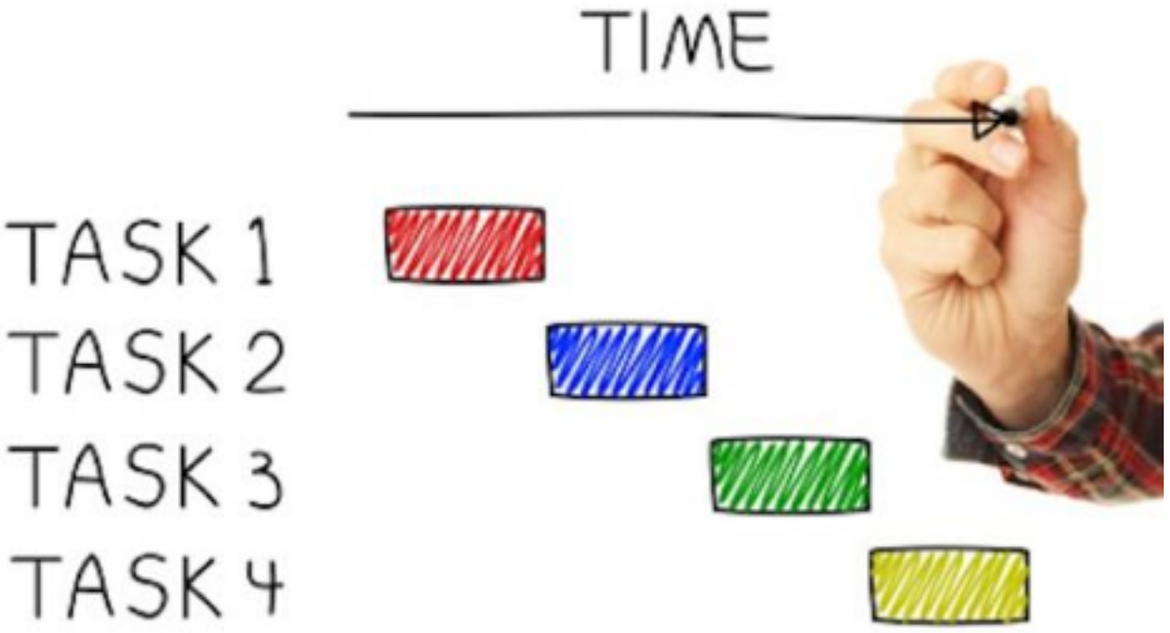
Process of predicting the most realistic amount of effort required to realise a software project

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

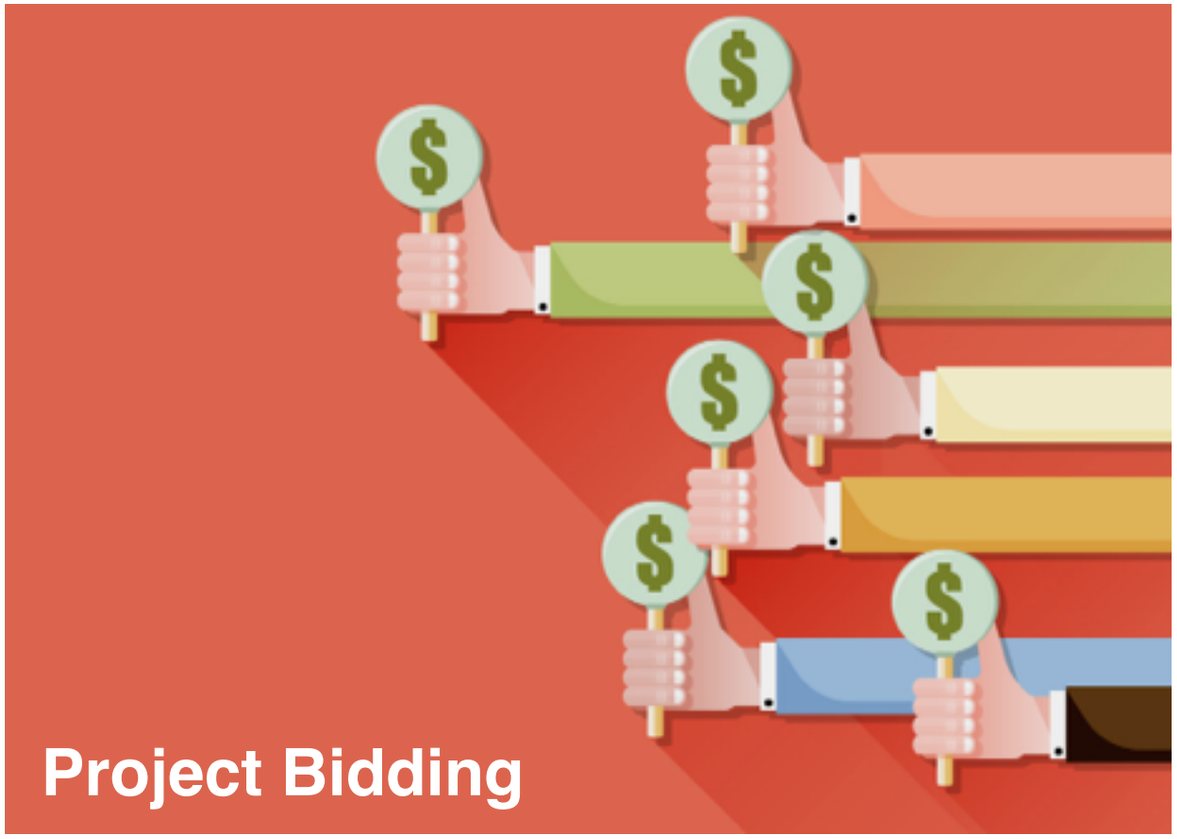
Would you ever start producing anything without knowing the cost?



Why is it Important?



Project Scheduling /Staffing



Project Bidding

Why is it Difficult?

Options for Estimation

Experts tend to under-estimate

What is the margin of error?

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

30%

40%

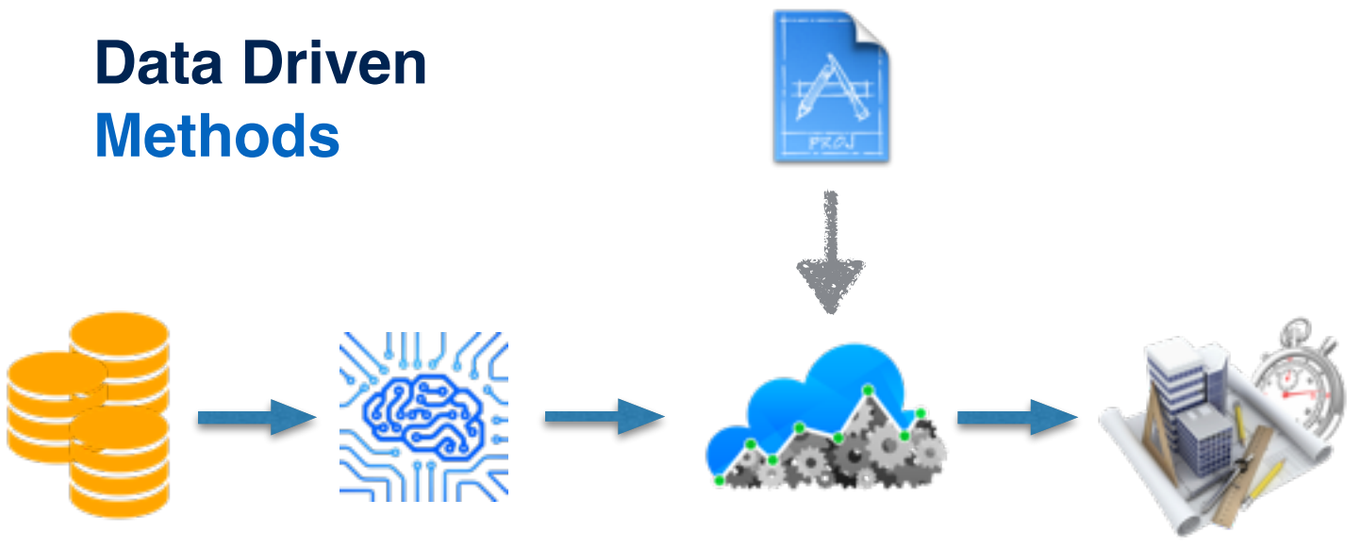
K. Molkken and M. Jorgensen. A review of surveys on software effort estimation. ISESE'03.
S. McConnell. Software Estimation: Demystifying the Black Art. Microsoft Press, 2006

Options for Estimation

Expert Judgment

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

Data Driven Methods

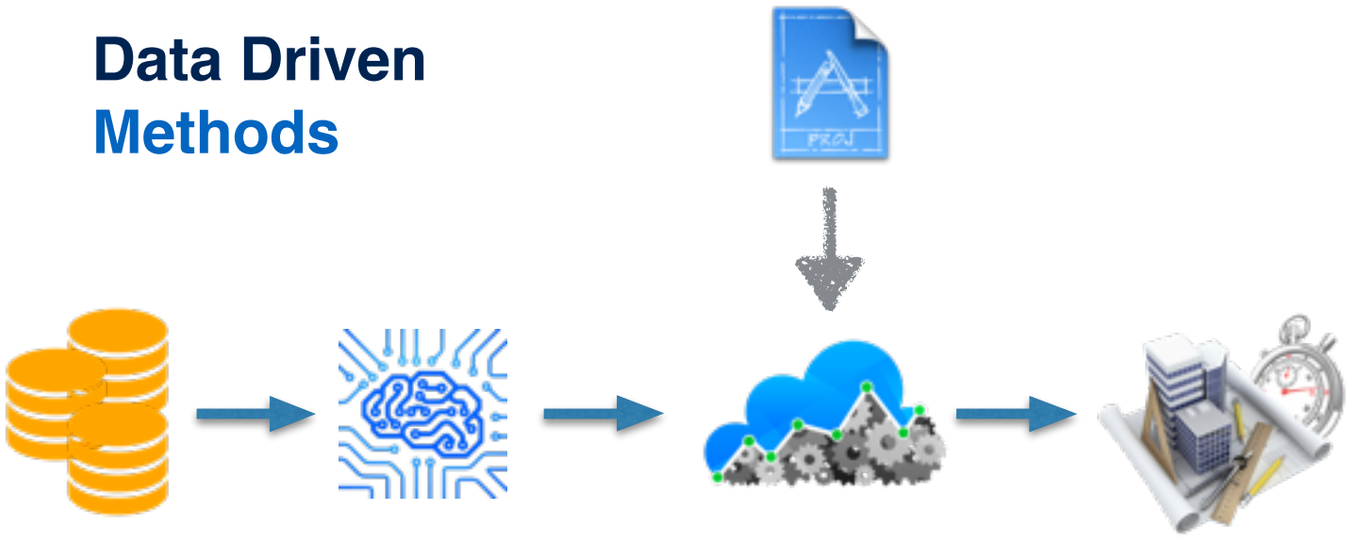


- Regression-based
- Analogy-based
- Search-based

Options for Estimation

After ~30 years of research…

Data Driven Methods



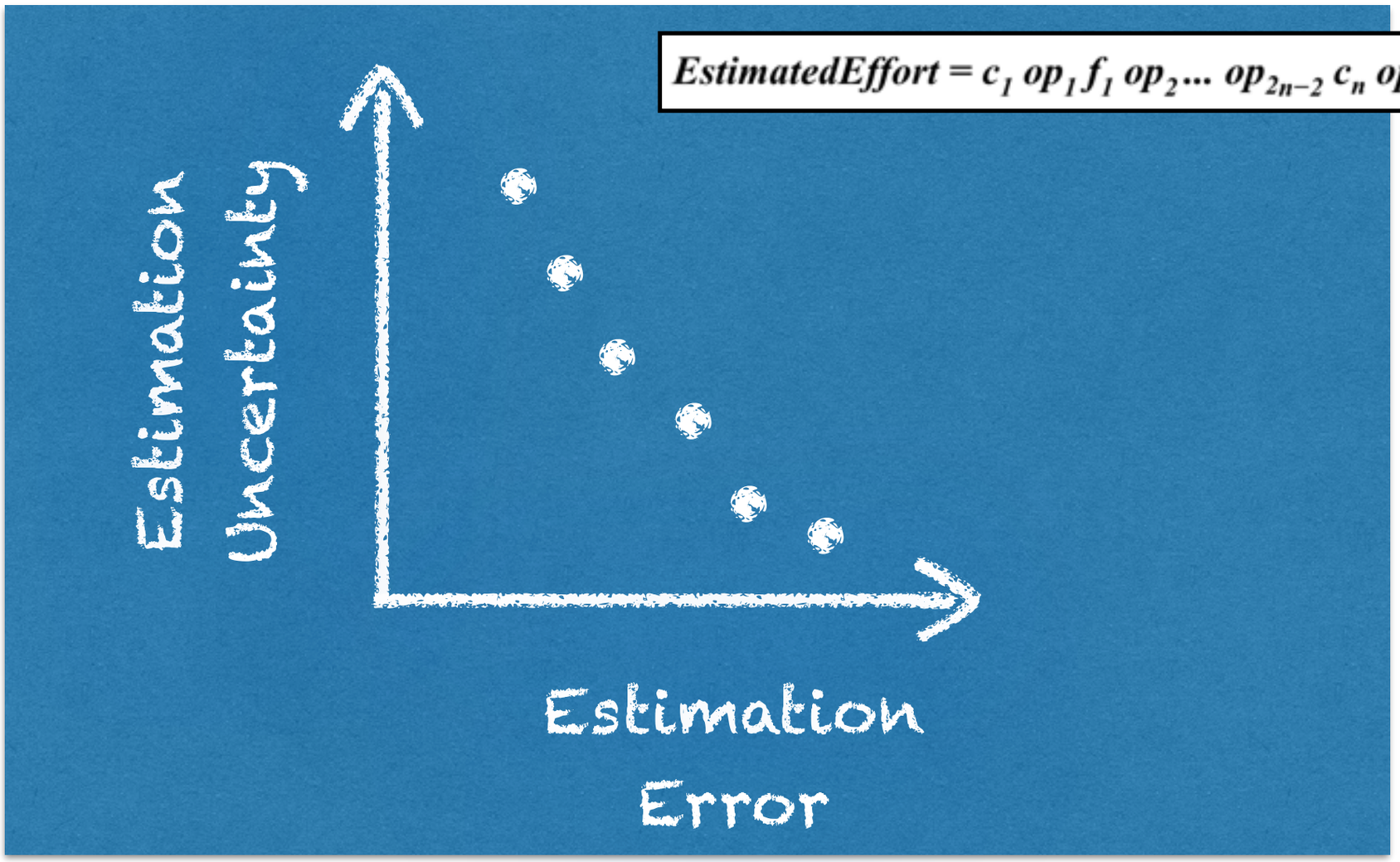
- Linear Regression
- Stepwise Regression
- Manual Stepwise Regression
- Support Vector Regression
- Classification and Regression Trees
- Case-based Reasoning
- K-Nearest Neighbours
- Genetic Algorithms
- Genetic Programming
- Tabu Search
- Simulated Annealing
- …

… still unable to par human-estimates!

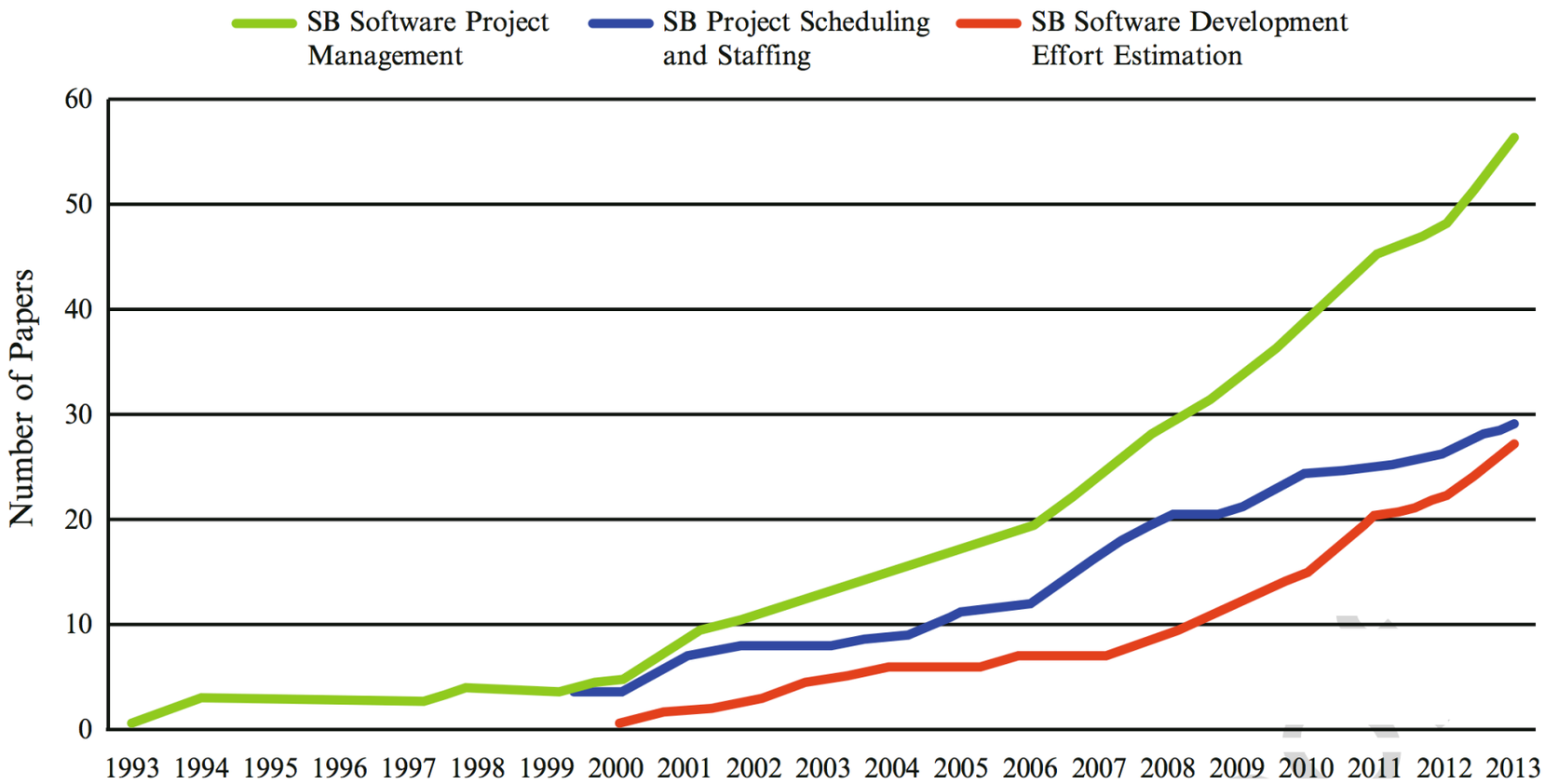
Confidence Guided Effort Estimation (CoGEE)

CoGEE is a multi-objective evolutionary approach

that seeks to build robust estimation models



Novelty of Our Approach



F. Ferrucci, M. Harman, F. Sarro, "Search-Based Software Project Management" in Software Project Management in a Changing World, G.Ruhe and C. Wholin (Editors), Springer, 2014

All previous evolutionary approaches sought to improve only point estimates

none of them was clearly better than the state-of-the-art

none of them parred human-expert estimates

Empirical Evaluation

CoGEE realised as a Non-dominated Sorting Genetic Algorithm-II (NSGAII)

Compared VS.

3 baselines

3 state-of-the-art effort estimators

3 alternative single/multi-objective formulations

5 industrial datasets from the PROMISE repository (724 projects)

RQ1.
Sanity Check

RQ2. State of the Art Benchmark

RQ3. Benefits from Multi-objective Formulation

RQ4. Comparison to Industrial Practices

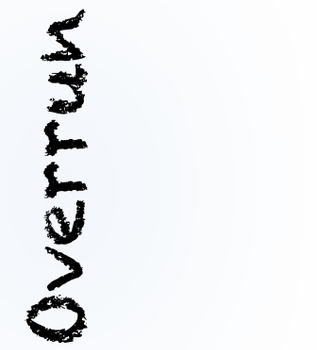
RQ4. Comparison to Industrial Practices

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

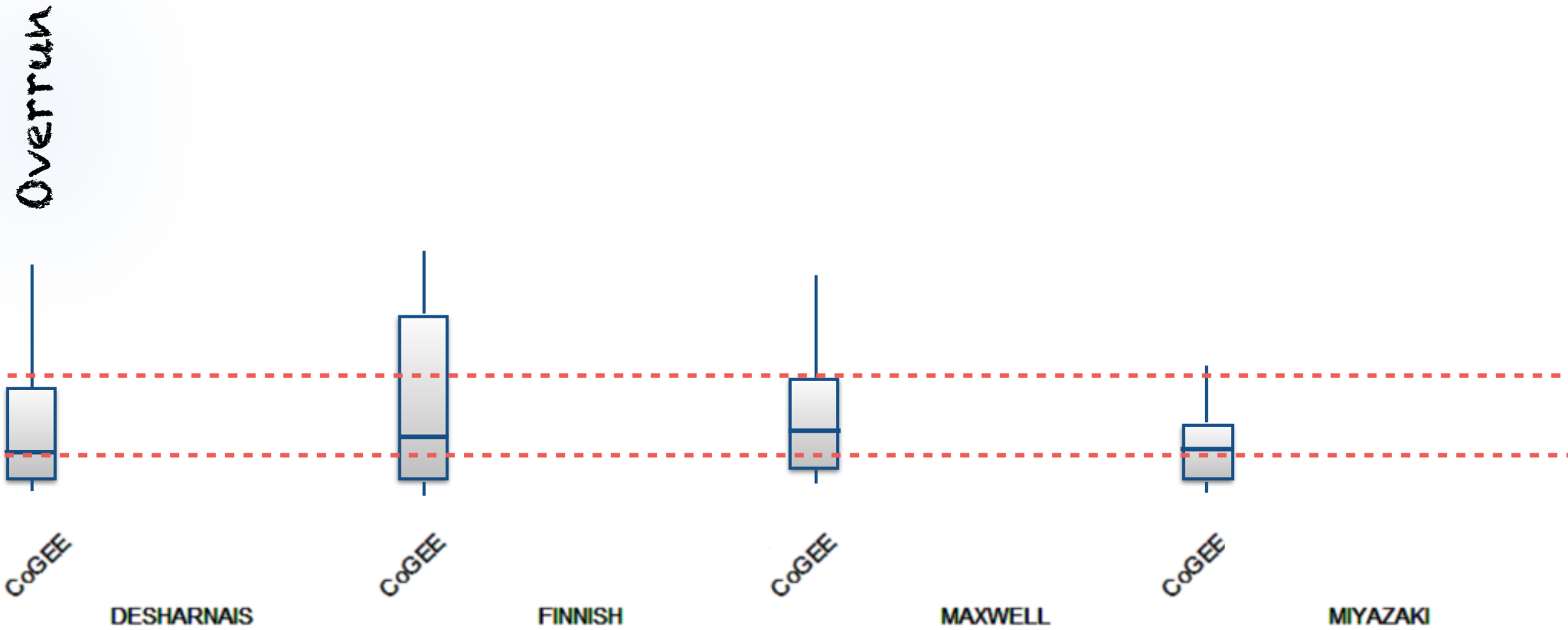
The evidence for these **thresholds** comes from a **survey of current industry practices** by Molkken and Jørgensen



DESHARNAIS FINNISH MAXWELL MIYAZAKI

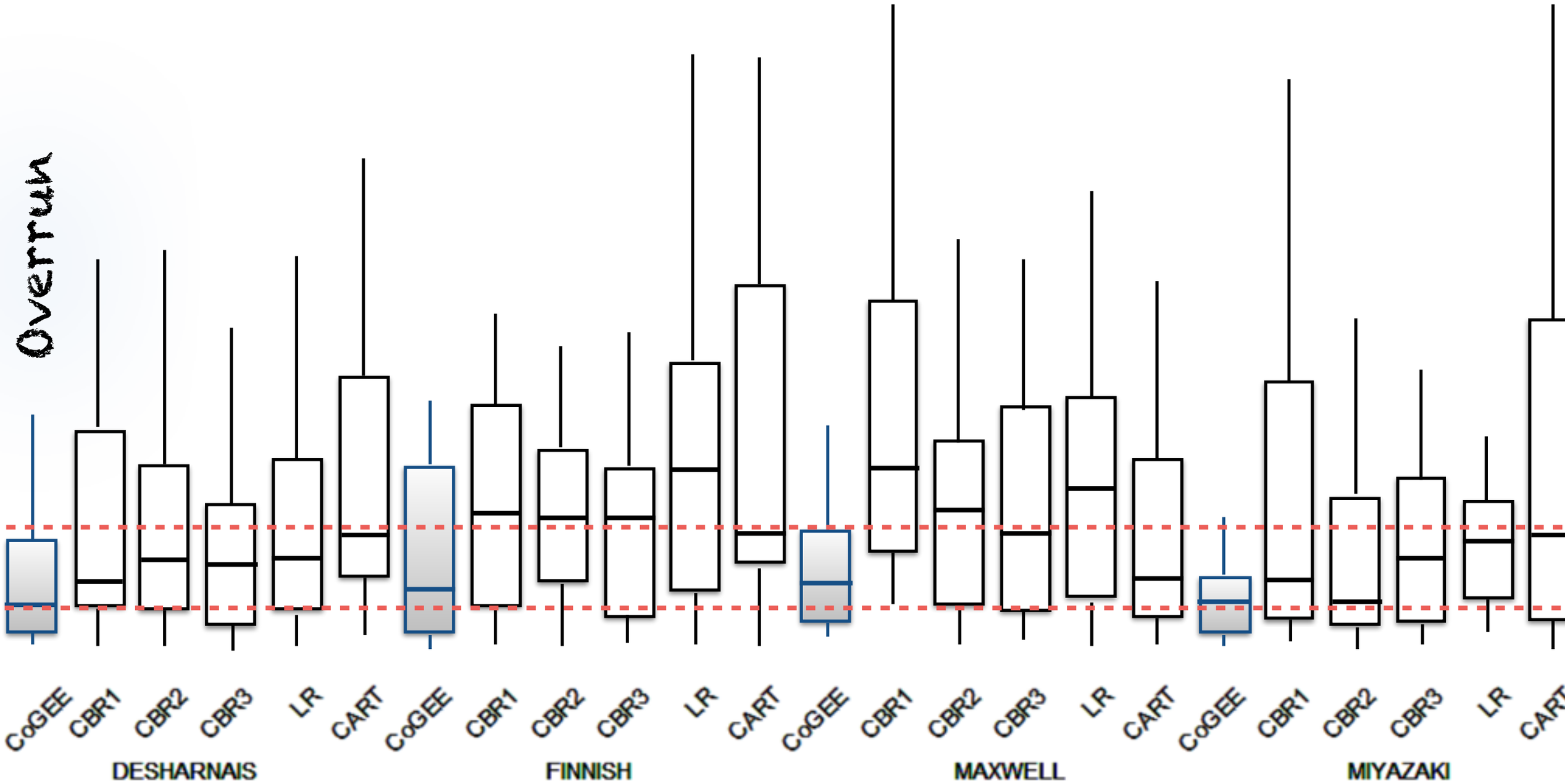
RQ4. Comparison to Industrial Practices

The median error of CoGEE lies within both thresholds for all the datasets



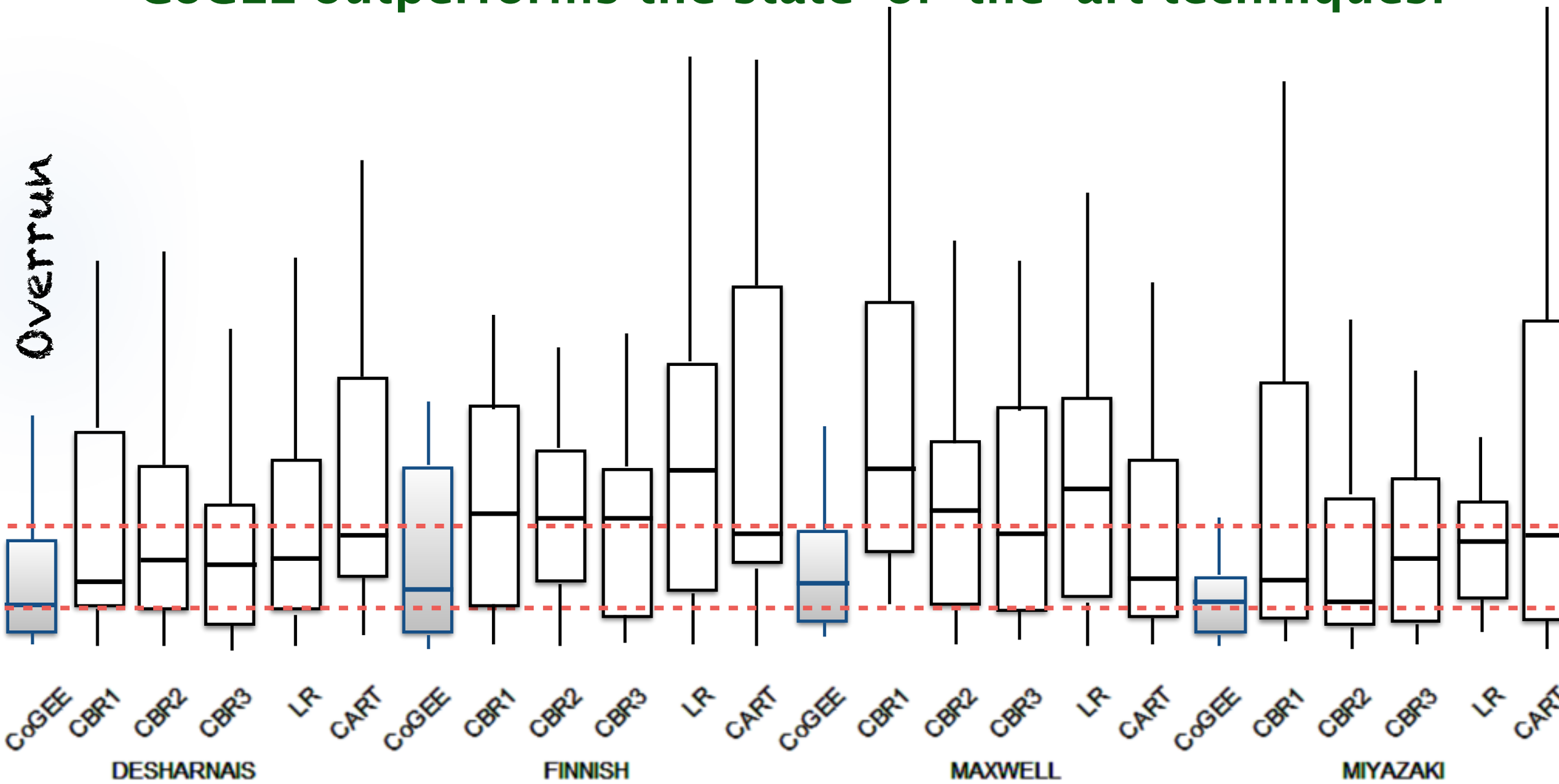
RQ4. Comparison to Industrial Practices

This is not always true for the state-of-the-art approaches



RQ4. Comparison to Industrial Practices

CoGEE provides human-competitive results!
CoGEE outperforms the state-of-the-art techniques!



Empirical Results

Our proposed bi-objective evolutionary algorithm

Creates a new state-of-the-art that pars currently claimed human-expert estimates (RQ4)

Outperforms scientific approaches previous published
(significantly and with medium and large effect size for all the datasets considered)

3 baselines (RQ1)

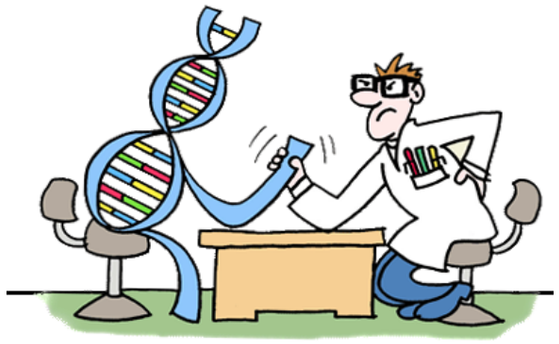
3 state-of-the-art methods (RQ2)

3 alternative single/multi-objective formulations (RQ3)

Criteria Satisfied by Our Work

- (G) The result solves a problem of indisputable difficulty in its field
- (E) The result is equal to or better than the most recent human-created solution to a long-standing problem for which there has been a succession of increasingly better human-created solutions
- (D) The result is publishable in its own right as a new scientific result independent of the fact that the result was mechanically created
- (B) The result is better than a result that was accepted as a new scientific result at the time when it was published in a peer-reviewed scientific journal
- (F) The result is equal to or better than a result that was considered an achievement in its field at the time it was first discovered

Why our entry is the “best”



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

Advances the state of the art

Novelty

Thorough empirical study (724 real-word projects)

Breakthrough results published in ICSE’16 top tier SE conference

Multi-objective Software Effort Estimation

Federica Sarro*, Alessio Petrozziello* and Mark Harman*
University College London, London, United Kingdom*
University of Portsmouth, Portsmouth, UK*
f.sarro@ucl.ac.uk, alessio.petrozziello@port.ac.uk, mark.harman@ucl.ac.uk

ABSTRACT

We introduce a bi-objective effort estimation algorithm that combines Confidence Interval Analysis and assessment of Mean Absolute Error. We evaluate our proposed algorithm on three different alternative formulations, baseline comparators and current state-of-the-art effort estimators applied to five real-world datasets from the PROMISE repository, involving 724 different software projects in total. The results reveal that our algorithm outperforms the baseline, state-of-the-art and all three alternative formulations, statistically significantly ($p < 0.001$) and with large effect size ($\hat{A}_{12} \geq 0.9$) over all five datasets. We also provide evidence that our algorithm creates a new state-of-the-art, which lies within currently claimed industrial human-expert-based thresholds, thereby demonstrating that our findings have actionable conclusions for practicing software engineers.

Categories and Subject Descriptors

D.2.9 [**Software Engineering**]: Management

Keywords

Software effort estimation; multi-objective evolutionary algorithm; confidence interval; estimates uncertainty.

1. INTRODUCTION

Effort estimation is a critical activity for planning and monitoring software project development in order to deliver the product on time and within budget [9, 51, 88]. The competitiveness (and occasionally the survival) of software organisations depends on their ability to accurately predict the effort required for developing software systems; both over- or under-estimates can negatively affect the outcome of software projects [59, 63, 84, 88]. Several algorithmic approaches have been proposed in literature to support software engineers in improving the accuracy of their estimations. These methods often produce a point estimate of the effort required to develop a new project.

Very few previous studies have accounted for the inherent uncertainty of the estimates produced [4, 6, 76, 83, 85, 86]. Some previous work has instead investigated the over-confidence and/or under-confidence of the prediction given by expert judgement [35, 36, 37, 38, 40, 41, 67]. Existing surveys on estimation practice [37, 67] suggest that human effort estimates are over-optimistic and there is a strong over-confidence in their accuracy.

We introduce a multi-objective evolutionary approach that seeks to build a robust estimation model by simultaneously maximising the estimation accuracy and minimising the uncertainty associated with the estimation model itself. We named this approach Confidence Guided Effort Estimator (CoGEE). We use the familiar sum of absolute error, abs(real effort - estimated effort), as one objective to guide our algorithm, combining this with the (less widely-known and less widely-used) confidence interval. The confidence interval is an estimated range of values that are likely (according to the chosen interval range) to include the estimated effort.

We report the results of four sets of experiments on five publicly available datasets to compare and evaluate our approach against candidate competitors: baseline estimators [82], state-of-the-art estimators [39, 50, 90], alternative single and multi-objective formulations, and currently claimed best industrial practice based on human judgment [67]. In our evaluation we follow recent best practice to assess prediction systems [82, 90] and evolutionary approaches [3, 31].

Our new bi-objective effort estimation algorithm outperforms baseline estimators (a sanity check), state-of-the-art techniques (case-based reasoning, linear regression, regression trees) and also three alternatives that we implemented in order to assess the degree to which multi-objectivity plays a role in the performance of our algorithm. These claims have been tested using a non-parametric (Wilcoxon) test for statistical significance which reveals that the results are significant ($p < 0.001$), after applying the Bonferroni correction for multiple statistical testing (the most conservatively cautious of all corrections). Furthermore, in all cases, our bi-objective algorithm outperforms these candidate competitors with a large effect size, as measured using the Vargha-Delaney non-parametric effect size measure ($\hat{A}_{12} \geq 0.9$).

We also compare both the estimation error produced by our algorithm (and the current state-of-the-art) and the budget overruns that would accrue from using them against two claimed thresholds for industrial best estimation practice. The results are very encouraging, suggesting that CoGEE moves median expected state-of-the-art performance within at least one, and sometimes both thresholds.

ICSE'16 May 14-22, 2016, Austin, TX, USA

ACM ISBN 978-1-4503-3900-1/16/05.
DOI: http://dx.doi.org/10.1145/2884781.2884830

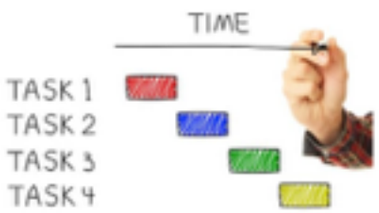
Multi-Objective Software Effort Estimation

F. Sarro, A. Petrozziello, M. Harman

Software Effort Estimation (SEE)

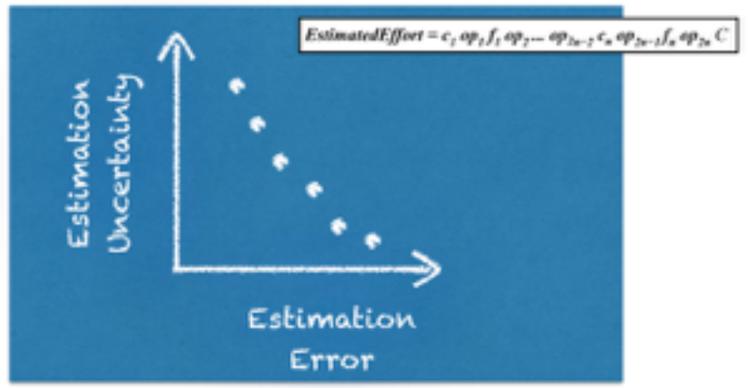
Process of predicting the most realistic amount of effort required to realise a software project

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



Confidence Guided Effort Estimation (CoGEE)

CoGEE is a multi-objective evolutionary approach that seeks to build robust estimation models



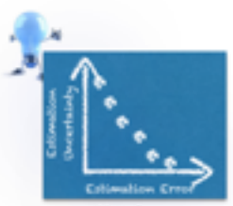
Criteria Satisfied by Our Work

- (G) The result solves a problem of indisputable difficulty in its field
- (E) The result is equal to or better than the most recent human-created solution to a long-standing problem for which there has been a succession of increasingly better human-created solutions
- (D) The result is publishable in its own right as a new scientific result independent of the fact that the result was mechanically created
- (B) The result is better than a result that was accepted as a new scientific result at the time when it was published in a peer-reviewed scientific journal
- (F) The result is equal to or better than a result that was considered an achievement in its field at the time it was first discovered

Why our entry is the "best"

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

Advances the state of the art



Novelty

Thorough empirical study (724 real-word projects)

Breakthrough results published in ICSE'16 top tier SE conference