



Software Management Research Group, METU
<http://smrg.ii.metu.edu.tr/smrgp/>

COMPARISON OF SOFTWARE BENCHMARKING REPOSITORIES FROM EFFORT PREDICTION PERSPECTIVE

UKSMA/COSMIC International Conference on Software Metrics & Estimating
London, October 27th/28th 2011

Özden Özcan Top, Mina Nabi, Onur Demirörs

Presentation Plan

- .. Introduction
- .. Selection and Evaluation Processes
- .. Features of Benchmarking repositories
 - ✧ General Features
 - ✧ Properties of Data Collection Processes
 - ✧ Properties of Data Quality/Data Validation Processes
 - ✧ Project Characteristics
- .. Findings and Discussion
- .. Conclusions



Software Management Research Group

Introduction

- .. Success of estimation models
 - Accuracy
 - Reliability
 - And validityof the data in repositories
- .. Benchmarking
 - Continuous and repeatable process
- .. Benchmark Data Repository
 - Product of benchmarking process
 - A cluster of project data collected for a variety of goals such as comparison, evaluation and construction of estimation model



Software Management Research Group

Introduction (Cont'd)

- .. Benchmarking
 - ⌘ Internal or External
 - ⌘ Public (e.g. ISBSG), semi-public (e.g. IPA/SEC) and private (e.g. SPR)
- .. Purpose of this comparative study
 - ⌘ Evaluation of the capabilities of external software benchmarking repositories using their suitability for construction, or validation of effort estimation models



Software Management Research Group

Selection and Evaluation Processes

Selection of the Repositories

1. ISBSG	International Software Benchmarking Research Group
2. CSBSG	Chinese Software Benchmarking Standards Group
3. IPA/SEC	Information-Technology Promotion Agency/ Software Engineering Center
4. Albrecht Repository	By Yanfu and Keung
5. China Repository	By Fang
6. Desharnais Repository	By Desharnais
7. Finnish Repository	By Keung
8. Maxwell Repository	By Yanfu
9. Kemerer Repository	By Keung
10. COCOMO81	by Barry Boehm
11. NASA	By Menzies
12. COCOMO SDR	In the form of COCOMO by SoftLab
13. Miyazaki94	By Amasaki
14. COCOMO_NASA93	In the form of COCOMO by NASA



Software Management Research Group

Features of Benchmarking repositories

- .. General Features
- .. Properties of Data Collection Processes
- .. Properties of Data Quality/Data Validation Processes
- .. Project Characteristics



Software Management Research Group

General Features

.. ISBSG

- Initiated in: 1994
- Non-profit organization
- Collects software project data in global range
- 5052 projects from the organizations in 24 different countries

.. CSBSG

- Initiated in: 2006
- Collects project data in national level
- 1012 projects from 16 cities of china

.. IPA/SEC

- Initiated in: 2004
- 1774 data collected from the companies operating in 20 cities of Japan, by 2008



Software Management Research Group

General Features (Cont'd)

- .. Albrecht Repository (24 projects with size range of 199FP to 1902FP)
- .. China Repository (499 projects with size range of 9FP to 17518 FP)
- .. Desharnais Repository (81 projects with size range of 62 FP to 1116 FP)
- .. Finnish Repository (38 projects with size range of 65 FP to 1814FP)
- .. Maxwell Repository (62 projects with size range of 48 to 3643 FP)
- .. Kemerer Repository (15 projects with size range of 99 to 2306FP)
- .. COCOMO81 Repository (63 in Lines of Code level)
- .. NASA93 Repository (93 projects in Lines of Code level)
- .. COCOMO SDR Repository (12 projects in Lines of Code level)
- .. Miyazaki Repository (48 projects in Lines of Code level)
- .. COCOMO_NASA Repository (60 projects in Lines of Code level)



Software Management Research Group

Properties of Data Collection Processes

- .. ISBSG, IPA/SEC and CSBSG
 - ✧ fully or partially defined data collection processes
- .. Other 11 data repositories
 - ✧ do not include information about data collection processes



Properties of Data Collection Processes (Cont'd)

.. ISBSG

- Utilizes Data Collection Questionnaires (DCQs) specialized according to the size measurement methods
- Glossary of terms provides brief description about the terms
- Tool support is not provided in data collection process
- Confidentiality of the data is ensured by the administrator

.. CSBSG

- Data collection process is not defined
- Confidentiality of data was ensured by encoding the company information

.. IPA/SEC

- Data is collected through data definition and data entry forms
- Provides white papers annually, which include quantitative analyses results



Software Management Research Group

Properties of Data Quality/Data Validation Processes

- .. Validation
 - ✧ To identify the accuracy or inaccuracy of the collected data through cross-check calculations, inspections, comparisons, and re-measures
- .. Quality of the data
 - ✧ Not only related with data accuracy, but also related with the sparseness and the appropriateness of the data
- .. Only in five data repositories, data pass through a full or partial validation process



Properties of Data Quality/Data Validation Processes (Cont'd)

- ISBSG
 - Data quality is rated from "A" to "D" (by A repository manager)
 - In ISBSG v11, only A and B level projects were validated against correctness and completeness
- CSBSG
 - CSBSG follows the same procedures as ISBSG
- IPA/SEC
 - data is passed through an inspection process (by quality assurance division or product control divisions in SEC)
- Desharnais repository was corrected by Dan Port
- COCOMO81 repository was also fixed by Dan Port



Software Management Research Group

Project Characteristics

- Project attributes related with software effort estimation
 - q Actual effort
 - q Size of software
 - q Duration of the project
 - q Team experience
 - q Team size
 - q Domain of the software project
 - q Programming language
 - q Hardware platform
 - q And the usage of standards through the software development processes
- All of the repositories provide size and effort data
- Team experience level is on the third place
- Very few of the repositories include software domain and used programming language



Software Management Research Group

Project Characteristics (Cont'd)

ISBSG

- Size in IFPUG, COSMIC, MarkII, NESMA, FISMA, and LOC forms
- The amount of the added, changed and deleted functions are provided for a few number of projects
- Base functional components (BFC) of IFPUG and COSMIC are also available
- Total effort and phase by phase effort in person-hours has been provided

CSBSG

- Couldn't be assessed from the project characteristics category point of view, since the raw data was not available



Software Management Research Group

Project Characteristics (Cont'd)

IPA/SEC

- Size data is collected in a more detailed form than the ISBSG
- Not only the effort but also the size is collected based on the phases in software development life cycle
- Planned and actual data of size and effort are also given
- Other kinds of size indicators such as the numbers of the documents pages, numbers of the screens, numbers of DFDs, and use cases are provided

Other 11 repositories

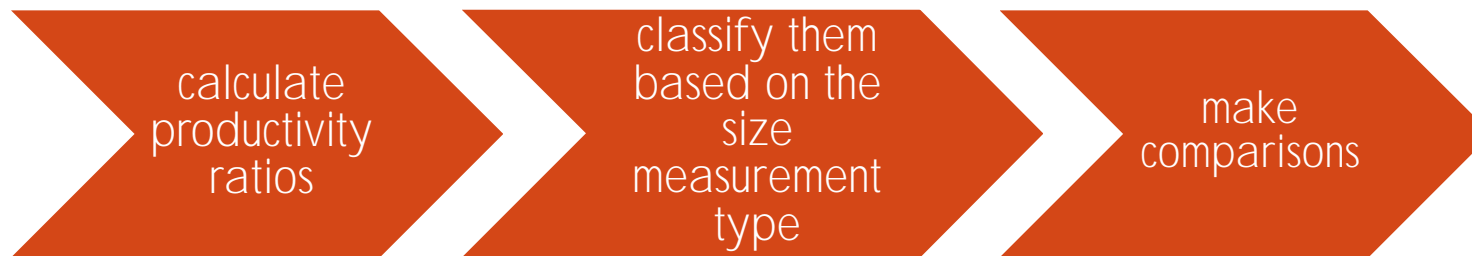
- Albrecht, China Repositories include functional size data in BFC level
- remaining repositories provide only a single size
- No effort data based on phases of software life cycle in any 11 repositories



Software Management Research Group

Findings and Discussion

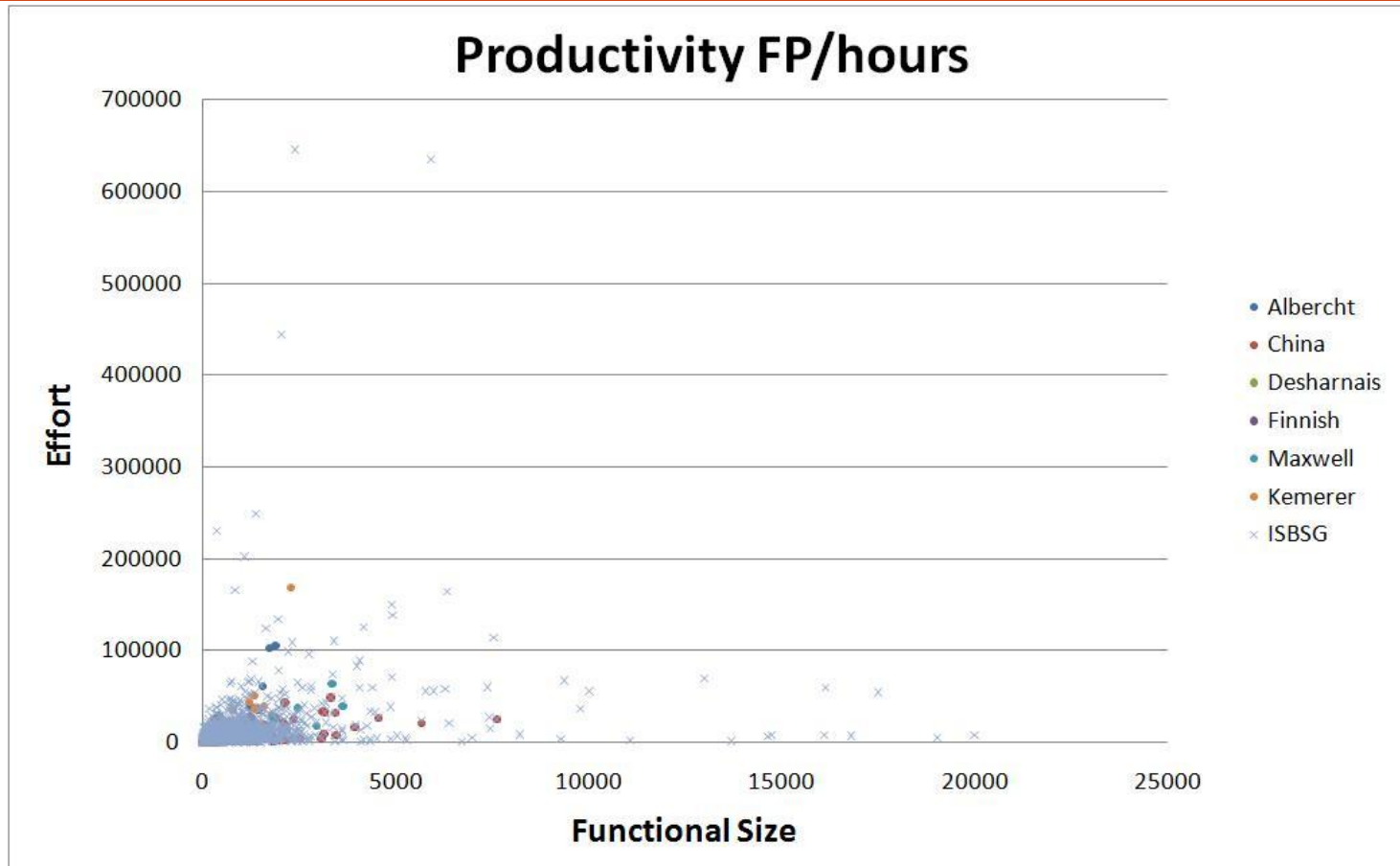
- Productivity ratios are utilized for effort estimation and performance evaluation



- Average productivities vary
 - from 0.01 to 0.21 in FP/person-hours unit
 - from 0.24 to 19.73 in SLOC/person-hours unit.



Findings and Discussion (Cont'd)

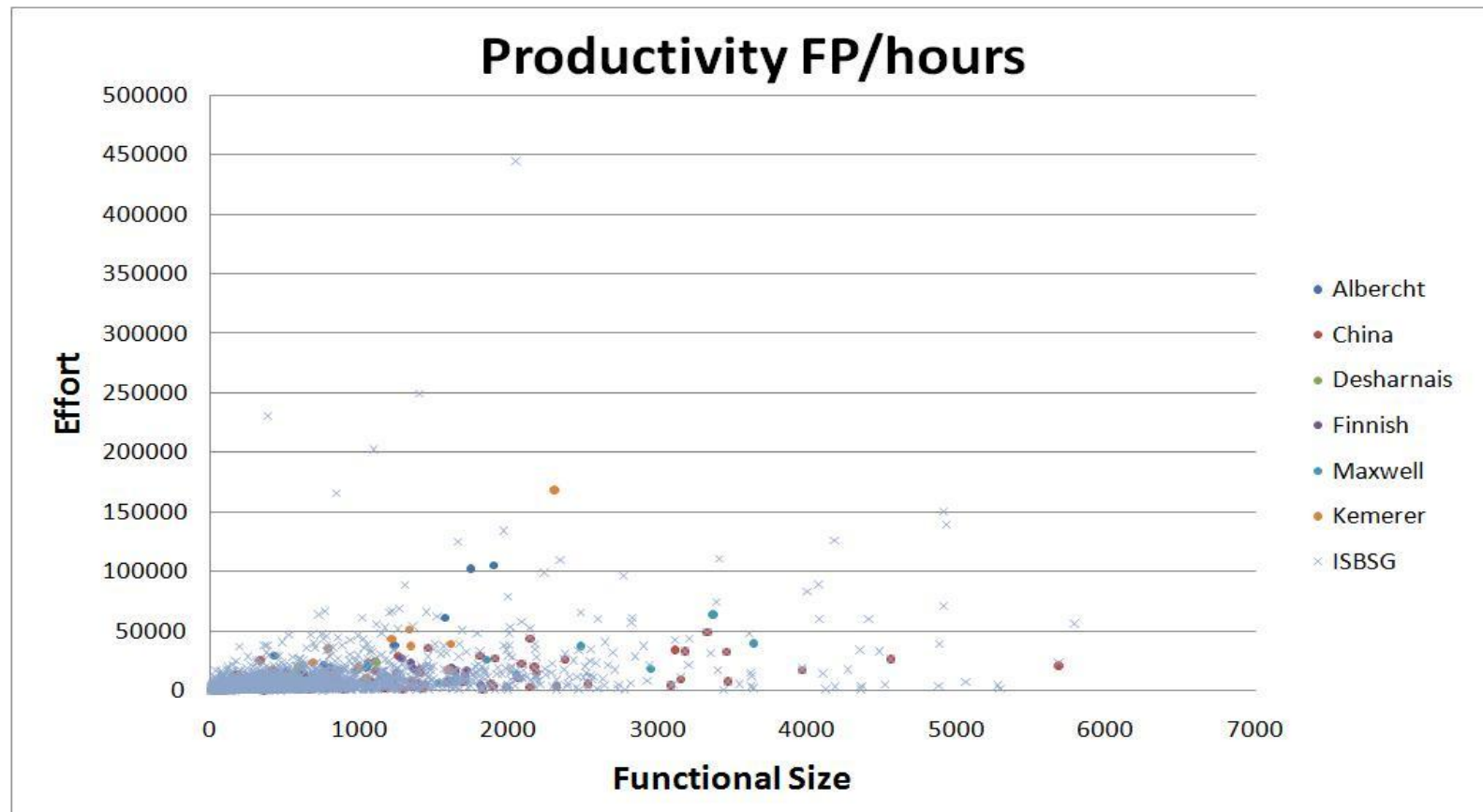


Distribution of the Data in Repositories in 0-20000 FP Interval



Software Management Research Group

Findings and Discussion (Cont'd)



Distribution of the Data in Repositories based on 0-6000 FP Size Interval



Software Management Research Group

Findings and Discussion (Cont'd)

- Validation of the collected data
 - No cross-check calculations, inspections, check lists and comparisons under a well-defined data validation
- Effort data only in total effort level
- Sparseness of the data
 - Data is presented under 118, 200 and 208 attributes in ISBSG, CSBSG and IPA/SEC respectively
 - In IPA/SEC, software size is collected both in estimated and actual levels for the phases of software development life cycle. Here the question is that, if in practice organizations keep data in such detail?!
 - %4 of ISBSG projects and %30 of CSBSG projects do not include size data, which makes these project data useless for effort estimation



Software Management Research Group

Conclusions

- .. Current data repositories usually support small and medium scale projects for effort estimation model construction
- .. Selection of the appropriate data from a mass of project data should be done based on definite data selection rules (such as pruning)
- .. Although ISBSG initiated a benchmarking standardization study, but there is not a developed benchmarking standard till now
- .. Data repository validation is the most eminent case
 - ⊠ Need for cross calculations, comparison detailed inspection check lists and question-answer sessions with data providers



Conclusions (Cont'd)

- .. Effort data is not available for
 - .. different phases of software development life cycle
 - .. or for each work task of work breakdown structure
- .. Lack of software project domain, programming language and team size in some of Repositories is observed
- .. Also, lack of data collection tools is another significant deficiency



Thanks for your attention

Questions?/
Answers



Software Management Research Group