Information and Software Technology xxx (2014) xxx–xxx

Contents lists available at ScienceDirect

Information and Software Technology

journal homepage: www.elsevier.com/locate/infsof

Analogy-based software development effort estimation: A systematic mapping and review

Ali Idri [a,*], Fatima azzahra Amazal [a], Alain Abran [b]

[a] *Software Projects Management Research Team, ENSIAS, Mohammed V Souissi University, Madinate Al Irfane, 10100 Rabat, Morocco*
[b] *Department of Software Engineering, Ecole de Technologie Supérieure, Montréal H3C IK3, Canada*

ARTICLE INFO

Article history:
Received 30 July 2013
Received in revised form 28 July 2014
Accepted 29 July 2014
Available online xxxx

Keywords:
Mapping study
Systematic literature review
Software development effort estimation
Analogy
Case-based reasoning

ABSTRACT

Context: Analogy-based software development effort estimation (ASEE) techniques have gained considerable attention from the software engineering community. However, to our knowledge, no systematic mapping has been created of ASEE studies and no review has been carried out to analyze the empirical evidence on the performance of ASEE techniques.

Objective: The objective of this research is twofold: (1) to classify ASEE papers according to five criteria: research approach, contribution type, techniques used in combination with ASEE methods, and ASEE steps, as well as identifying publication channels and trends; and (2) to analyze these studies from five perspectives: estimation accuracy, accuracy comparison, estimation context, impact of the techniques used in combination with ASEE methods, and ASEE tools.

Method: We performed a systematic mapping of ASEE studies published in the period 1990–2012, and reviewed them based on an automated search of four electronic databases.

Results: In total, we identified 65 studies published between 1990 and 2012, and classified them based on our predefined classification criteria. The mapping study revealed that most researchers focus on addressing problems related to the first step of an ASEE process, that is, feature and case subset selection. The results of our detailed analysis show that ASEE methods outperform the eight techniques with which they were compared, and tend to yield acceptable results especially when combining ASEE techniques with fuzzy logic (FL) or genetic algorithms (GA).

Conclusion: Based on the findings of this study, the use of other techniques such FL and GA in combination with an ASEE method is promising to generate more accurate estimates. However, the use of ASEE techniques by practitioners is still limited: developing more ASEE tools may facilitate the application of these techniques and then lead to increasing the use of ASEE techniques in industry.

Contents

[*] Corresponding author. Tel.: +212 661390943.
E-mail address: idri.ali123@gmail.com (A. Idri).

http://dx.doi.org/10.1016/j.infsof.2014.07.013
0950-5849/© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Estimating the cost of a software project in terms of effort is one of the most important activities in software project management. This is because rigorous planning, monitoring, and control of the project are not feasible if the estimates of software development cost are highly inaccurate. Unfortunately, the industry is plagued with unreliable estimates, and no effort estimation model has proven to be consistently successful at predicting software project effort in all situations [1]. Researchers in the software engineering community continue to propose new models to achieve effort prediction accuracy. Jørgensen and Shepperd [2] conducted a systematic review in which they identified up to 11 estimation approaches in 304 selected journal papers. These approaches fall into two major categories: parametric models, which are derived from the statistical and/or numerical analysis of historical project data, and machine learning (ML) models, which are based on a set of artificial intelligence techniques such as artificial neural networks (ANN), genetic algorithms (GA), analogy-based or case-based reasoning (CBR), decision trees, and genetic programming.

ML techniques are gaining increasing attention in software effort estimation research, as they can model the complex relationship between effort and software attributes (cost drivers), especially when this relationship is not linear and does not seem to have any predetermined form. Recently, Wen et al. [1] carried out a systematic literature review in which they identified eight types of ML techniques. ASEE and ANN-based effort estimation techniques are the most frequently used of these, 37% and 26% of the time respectively. Their SLR also showed that the CBR and ANN are more accurate in terms of the arithmetic mean of Preds(25) and arithmetic mean of MMREs, obtained from selected studies, than the other ML techniques (mPred(25) = 46% and mMMRE = 51% for CBR-based studies, and mPred(25) = 64% and mMMRE = 37% for ANN-based studies). This confirms the results of the study carried out in [2]: the use of ASEE techniques instead of other ML techniques (ANN, Classification and Regression Trees) is increasing over time (10% instead of 7% for ANN and 5% for classification and regression tress – CRT until the year 2004). Moreover, instead of ANNs which are often considered as black-box, ASEE techniques are claimed to be easily understood by users, as they are similar to human reasoning by analogy [1] (see Table C.13 of [1] in which more than 15 references are supporting this affirmation). Nevertheless, Section 4.3 discusses the numerous hard decisions and limitations that prevent ASEE techniques to be easily used in a given context.

In spite of these advantages, ASEE techniques are still limited by their inability to correctly handle categorical attributes (measured on a nominal or ordinal scale). Indeed, the commonly used way to assess the similarity between two software projects described by nominal attributes is to use the overlap measure which assigns a similarity of 1 if the values are identical and a similarity of 0 if the values are not identical [3–6]. For ordinal attributes, most studies map the ordinal values to their ranking numbers (or positions) and then assess the similarity using some arithmetic operations (addition, subtraction, etc.) that are not meaningful according to measurement theory [4,6,7]. Furthermore, inconsistent results have been reported regarding their accuracy, compared with other effort estimation techniques, both ML and non ML. For example, some studies [3,8–10] claim that ASEE techniques outperform regression models, while the results of others [11,12] indicate that regression models are superior to ASEE techniques. Based on these contradictory results, we see a need to systematically analyze the

Table 1
Mapping study questions.

ID	Mapping question	Main motivation
MQ1	Which (and how many) sources include papers on ASEE?	To provide effort estimation researchers with a list of relevant studies on ASEE
MQ2	What are the most frequently applied research approaches in the ASEE field, and how has this changed over time?	To identify research approaches and their trends over time in the ASEE literature
MQ3	What are the main types of contribution of ASEE studies?	To identify the different types of contribution of ASEE studies
MQ4	Which of the reported techniques are used the most frequently in combination with ASEE techniques?	To identify the techniques used in combination with analogy to improve the estimation accuracy of ASEE techniques
MQ5	Have the various steps of the analogy procedure received the same amount of attention on the part of researchers?	To classify the various steps of the analogy procedure based on the amount of attention they have received from researchers

Please cite this article in press as: A. Idri et al., Analogy-based software development effort estimation: A systematic mapping and review, Inform. Softw. Technol. (2014), http://dx.doi.org/10.1016/j.infsof.2014.07.013

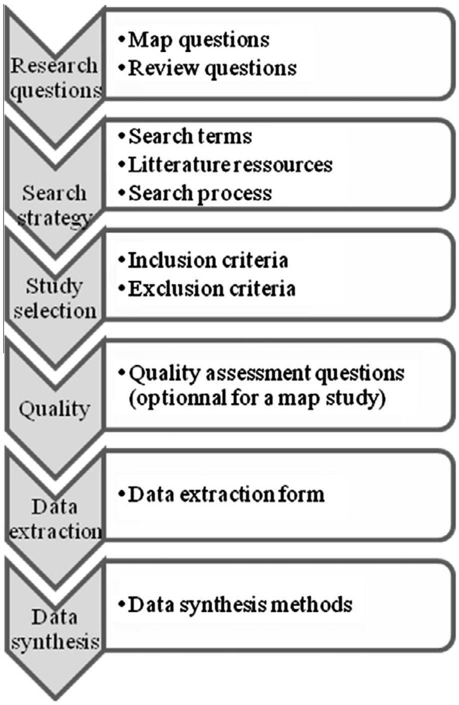

Fig. 1. Mapping and review process.

evidence reported on ASEE techniques, in order to understand and facilitate their application. We propose to achieve this by: (1) building a classification scheme and structuring the field of interest; and (2) summarizing the evidence of ASEE technique performance in current research.

To the best of our knowledge, no systematic mapping or review has been performed to date with a focus on software effort prediction using analogy: (1) the study [2] did not report the performance results of effort prediction techniques, and (2) the study [1] dealt only with the accuracy of effort prediction using analogy whereas this paper deals also with other issues of effort prediction using analogy such as the most investigated steps, the techniques used in combination with analogy, and impact on ASEE accuracy when combining other techniques with ASEE (see MQs and RQs of Tables 1 and 3). Consequently, in this paper, we aggregate the results of a set of selected studies on ASEE techniques, published in the period of 1990–2012, using systematic mapping and review procedures. The use of these procedures is motivated by the high quality and rigor of the methodology proposed by Kitchenham and Charters [13]. Our aims are the following:

- To provide a classification of ASEE studies with respect to: publication channels, research approach, contribution type, technique used in combination with analogy, and ASEE steps.
- To analyze evidence regarding: (1) the estimation accuracy of ASEE techniques; (2) the prediction accuracy of ASEE models compared with that of the other models; (3) favorable estimation contexts for using ASEE models; (4) the impact of incorporating other techniques into an ASEE model; and (5) the tools used to implement ASEE models.

The remainder of this paper is organized as follows. Section 2 describes the methodology adopted to conduct this systematic mapping and review. Section 3 reports and discusses the findings of the mapping. Section 4 presents and discusses the review results. Section 5 describes the implications for research and practice. Section 6 reports the limitations of this review. Finally, conclusions and future work are presented in Section 7.

2. Mapping and review process

Mapping studies use the same basic methodology as the systematic literature review (SLR), but they have different goals [14]. A systematic mapping study is a defined method for building a classification scheme and structuring a field of interest, and provides a structure for categorizing the type of research reports and results that have been published. An SLR is conducted to provide recommendations based on the strength of the evidence. We adopted the mapping and review process suggested by Kitchenham and Charters [13], comprising the following six steps: draw up mapping and review questions, carry out an exhaustive search for primary studies, select studies, perform a quality assessment of those studies, extract data, and finally synthesize data – see Fig. 1.

Table 2
Classification criteria.

Property	Categories
Research approach	History-based evaluation (HE), solution proposal (SP), case study (CS), experiment (EXP), theory (TH), review (RV), survey (SV), other (OT)
Contribution type	Technique, tool, comparison, validation, metric, model
Techniques used in combination with ASEE methods	Fuzzy logic (FL), genetic algorithm (GA), expert judgment (EJ), artificial neural network (ANN), least squares regression (LSR), statistical method (SM), grey relational analysis (GRA), collaborative filtering (CF), rough set analysis (RSA), bees algorithm (BA), multi-agent technology (MAT), model tree (MT)
Analogy step	Feature and case subset selection (FCSS), similarity evaluation (SE), adaptation (AD)

Table 3
Review study questions.

ID	Review question	Main motivation
RQ1	What is the overall estimation accuracy of ASEE techniques?	To identify to what extent ASEE techniques provide accurate estimates
RQ2	Do ASEE techniques perform better than other estimation models (ML and non ML)?	To compare ASEE techniques with other effort estimation models in terms of estimation accuracy
RQ3	What are the most favorable estimation contexts for ASEE techniques?	To identify the characteristics, strengths, and weaknesses of ASEE techniques
RQ4	What are the impacts of combining other techniques with an ASEE technique on its estimation accuracy?	To identify to what extent combining other techniques with an ASEE technique improves the accuracy of the estimates
RQ5	What are the most frequently used ASEE tools?	To support practitioners with ASEE tools

Please cite this article in press as: A. Idri et al., Analogy-based software development effort estimation: A systematic mapping and review, Inform. Softw. Technol. (2014), http://dx.doi.org/10.1016/j.infsof.2014.07.013

A detailed description of each of these steps is provided in the following subsections.

2.1. Mapping and review questions

Based on the focus of this study, we identified five mapping questions (MQs), which we list in Table 1. The MQs are related to the structuring of the ASEE research area with respect to the properties and categories described in Table 2. These categories are defined and explained in Tables A.18 and A.19 (Appendix A). Table 3 states the five questions, along with our main motivation for including them in the systematic review.

2.2. Search strategy

To find relevant ASEE studies to answer our research questions, we conducted a search composed of three steps. The first step was to define a search string. The second step was to apply this search string on a set of selected digital libraries to extract all the relevant papers. The third step was to devise a search procedure designed to ensure that no relevant paper had been left out. These three steps are described in detail below.

2.2.1. Search terms

We derived the search terms using the following series of steps [15]:

- Identify the main terms matching the mapping and review questions listed above.
- Search for all the synonyms and spelling variations of the main terms.
- Use the Boolean operator OR to join synonymous terms, in order to retrieve any record containing either (or all) of the terms.
- Use the Boolean operator AND to connect the main terms, in order to retrieve any record containing all the terms.

The complete set of search terms was formulated as follows:

> (analogy OR "analogy-based reasoning" OR "case-based reasoning" OR CBR) AND (software OR system OR application OR product OR project OR development OR Web) AND (effort OR cost OR resource) AND (estimat* OR predict* OR assess*).

2.2.2. Literature resources

To answer our research questions, we performed an automated search based on the preconstructed search terms using the following electronic databases:

- IEEE Digital Library.
- ACM Digital library.
- Science Direct.
- Google Scholar.

The IEEE, ACM and Science Direct Digital Libraries were chosen because most of the publication venues of selected papers in the previous SLRs on software development effort estimation [1,2] such as s (IST), IEEE Transactions on Software Engineering (TSE), Journal of Software and Systems (JSS), and Empirical Software Engineering (EMSE) are indexed by these three databases. Google Scholar was also used to seek other studies in the field because Google Scholar explores other digital databases. All the searches were limited to articles published between 1990 and 2012. They were conducted separately in the IEEE, ACM, and Science Direct databases based on title, abstract, and keywords. In Google Scholar, the search was restricted to paper titles, in order to avoid irrelevant studies. The search terms were used depending on the properties of the search engine of each electronic database.

2.2.3. Search process

To avoid leaving out any relevant paper and to ensure the quality of the search, a two-stage search process was adopted:

- The initial search stage

Here, we used the proposed search terms to search for primary candidate studies in the four electronic databases. The retrieved papers were grouped together to form a set of candidate papers.

- The secondary search stage

The reference lists of relevant studies (candidate studies that meet the inclusion and exclusion criteria) were reviewed to identify papers related to ASEE based on their title. Whenever a highly relevant article was found, we added it to the set of primary relevant studies. Besides, existing relevant papers that we were already aware of were used to control the quality of the search. Table B.20 of Appendix B shows, for each existing paper, the databases from which it was retrieved before and after the search. Note that in most cases, the databases are the same except for 6 cases due to the sequence of database search (IEEE, ACM, Science Direct, and then Google Scholar). In this way, we were able to assess whether or not the initial search stage had missed any highly relevant papers and to ensure that the search covered the maximum number of available ASEE studies.

2.3. Study selection procedure

The aim of this step was to identify the relevant studies that addressed the research questions based on their title, abstract, and keywords. To achieve this, each of the candidate papers identified in the initial search stage was evaluated by two researchers, using the inclusion and exclusion criteria, to determine whether it should be retained or rejected. If this decision could not be made using its title and/or its abstract alone, the full paper was reviewed. The inclusion criteria as well as exclusion criteria are linked using the OR Boolean operator.

Inclusion criteria:

- Use of an ASEE technique to predict software effort, and possibly comparison of the performance of this technique with that of other software effort estimation techniques (not the opposite).
- Use of a hybrid model that combines analogy with another technique (e.g. GA, ANN, or FL) to estimate software development effort.
- Comparison of two or more ASEE techniques.

Exclusion criteria:

- Duplicate publications of the same study (where several publications of the same study exist, only the most complete one is included in the review).
- Estimation of maintenance or testing effort.
- Estimation of software size or time without estimating effort.
- Study topic is software project control.

Each paper was evaluated by two researchers using the above criteria. Prior to applying the exclusion and inclusion criteria, the researchers discussed the criteria and reached agreement on which ones to retain. Then, each researcher went through the titles and abstracts, and categorized each candidate paper as "Include" (the

Table 4
Quality assessment questions.

ID	Question
QA1	Are the objectives of the study clearly defined?
QA2	Is the solution proposed well presented?
QA3	Is there a description of the estimation context?
QA4	Does the study report results that support the findings of the paper?
QA5	Does the study make a contribution to academia or to industry?
QA6	Has the study been published in a recognized and stable journal, or at a recognized conference/workshop/symposium?

researcher is sure that the paper meets at least one of the inclusion criteria and none of the exclusion criteria), "Exclude" (the researcher is sure that the paper meets at least one of the exclusion criteria and none of the inclusion criteria), or "Uncertain" in all other situations. If both researchers categorized one paper as "Include", the paper was considered to be relevant; if both researchers categorized one paper as "Exclude", the paper was excluded; otherwise, the paper was labeled "Uncertain", which means that the researchers disagreed on its relevance. The results show a high level of agreement between the two researchers and only six cases of disagreement. The high level of agreement indicates the relevance of the inclusion and exclusion criteria used. In cases of disagreement, the two reviewers discussed the papers, using either the partial text or the full text, until they came to an agreement. Of the six papers on which there was disagreement, four were retained and two were excluded.

The application of the selection criteria to the candidate articles in the initial search stage resulted in 104 relevant papers. Scanning of the reference lists of these papers that we compiled revealed no additional relevant papers.

2.4. Study quality assessment

Quality assessment is usually carried out in SLRs, but less often in systematic mapping studies. However, in order to enhance our study, we designed a questionnaire to assess the quality of the 104 relevant papers and used it in both the systematic mapping and the review studies. Quality assessment (QA) is necessary in order to limit bias in conducting the mapping and review studies, to gain insight into potential comparisons, and to guide the interpretation of findings [16].

The quality of the relevant papers was evaluated based on the 6 questions presented in Table 4. Questions 1–5 have three possible answers: "Yes", "Partially", and "No". These answers are scored as follows: (+1), (+0.5), and (0) respectively. Question 6 was rated based on the 2011 Journal Citation Reports (JCR) and the computer science conference rankings (CORE) [17]. The possible answers to this question were the following:

- For journals: (+1) if the journal ranking is Q1, (+0.5) if the journal ranking is Q2, and (0) if the journal ranking is Q3 or Q4.
- For conferences, workshops, and symposiums: (+1) if the conference/workshop/symposium is CORE A, (+0.5) if the conference/workshop/symposium is CORE B, and (0) if the conference/workshop/symposium is CORE C.

Even though the quality assessment criteria and their evaluation scales may be subjective, they do provide a common framework for comparing the selected papers. Similar criteria were used in [1,13,18]. However, the score for question 6 reflects whether or not the study has been published in a recognized and stable journal, or at a recognized conference, workshop, or symposium. Recognizable and stable journals/conferences means journals ranked in JCR 2011 and conferences ranked in CORE 2012 respectively. These two ranking sources (JCR and CORE) are largely accepted within the community as providing high quality papers.

The quality assessment of the relevant studies was performed by two researchers independently. All disagreements were discussed until a final consensus was reached. In order to ensure the validity of the selected papers and the reliability of our findings, an article was selected if its quality score exceeded 3 (50% of the perfect quality score of an article: 6). Note that the same strategy has been adopted by Wen et al. [1]. We selected 65 relevant articles with an acceptable quality score and rejected 39 articles with quality score of less than 3. The quality scores of the 65 selected articles are presented in Table B.21 in Appendix B.

2.5. Data extraction and synthesis

A data extraction form was created and completed for each of the selected papers for addressing the research questions of both the systematic mapping and the systematic review. The data extracted from each of these papers are listed in Table 5.

The data extraction was performed independently by two researchers who read the full text (for the systematic review in particular) of all selected papers, and collected the data necessary to address the research questions raised in this review. The extracted data were compared and disagreements were resolved by consensus between the two researchers. The number of disagreements depends on each MQ/RQ. Tables of Appendices B–D provided the final data extraction results to allow the readers to check their validity. Note that not all the selected papers necessarily answer all the review questions listed in Table 3 explicitly, that is, RQ1, RQ2, and RQ4. The solution suggested in [1] was adopted for those questions, which is that the optimal configuration results were used if there were different model configurations involved (optimal configuration means the best performance in terms of MMRE and Pred(25), and the average of the accuracy values when different dataset samplings were used.

Once the data had been extracted from the included studies, they were synthesized and tabulated in a manner consistent with the research questions addressed, in order to aggregate evidence to answer them. Since these data include both quantitative and qualitative data, and because the review addresses different kinds of research questions, various data synthesis approaches were used:

- *Narrative synthesis*: In this method, a narrative summary of the findings of the selected papers is created. To enhance the presentation of these findings, we used visualization tools such as bar charts, bubble plots, and box plots.
- *Vote counting*: This approach consists of calculating the frequency of various kinds of results across selected studies. Although it has been criticized by some researchers [19], the method is useful in addressing some review questions (e.g. RQ2).
- *Reciprocal translation*: This technique was used in this review to analyze and synthesize the qualitative data extracted from the selected papers (e.g. RQ3). It consists of a process of translation of the main concepts or themes reported across multiple studies to identify the similarities or differences between them.

2.6. Threats to validity

The main threats to the validity of our review are: exclusion of relevant articles, publication bias, and data extraction bias.

Exclusion of relevant articles: One of the major issues we faced in this review was finding all the relevant papers that addressed

Table 5
Data extraction form.

- Data extractor
- Data checker
- Study identifier
- Publication year
- Name(s) of the author(s)
- Title
- Source
- MQ2 – Research approach (see Tables 2 and A.18)
- MQ3 – Contribution type (see Table A.19)
- MQ4 – Techniques used in combination with analogy (see Table 2)
- MQ5 – ASEE steps investigated
  - Steps investigated
  - Author(s)
  - Purposes
- RQ1 – Estimation accuracy of the ASEE technique
  - Datasets employed for validation (name, size, number of used projects)
  - Evaluation criteria used to measure estimate accuracy (Pred(25), MMRE, MdMRE, other)
  - Validation method used in the study (leave-one-out cross validation, holdout, n-fold cross validation, other)
  - Estimation accuracy according to each evaluation criterion
- RQ2 – Performance of the ASEE technique compared to that of the other estimation models
  - Estimation techniques compared with the ASEE technique
  - Estimation accuracy of each technique used for comparison according to each evaluation criterion
- RQ3 – Favorable estimation contexts for ASEE techniques
  - Advantages of ASEE techniques
  - Limitations of ASEE techniques
  - Other characteristics of ASEE techniques
- RQ4 – Impact of combining ASEE methods with other techniques
  - Degree of improvement based on each evaluation criterion
  - Motivations for combining analogy with another technique
- RQ5 – ASEE tools used to generate estimates
  - Name of the ASEE tool
  - Author(s)
  - Year
  - Description

the research questions. To achieve this objective, we conducted a search on the four electronic databases listed in Section 2.2.2, using our search string on their search engines. However, we recognized the probability that some relevant studies would not be returned by the search terms we used. To reduce this threat, we manually checked the reference list of each of the relevant studies to look for any relevant studies that were missed in the automated search. To further reduce the risk of incorrectly excluding relevant papers, we took the following actions:

- Two researchers conducted the process of selecting the relevant studies separately, using the inclusion and exclusion criteria based on title, abstract, and keywords. If there was any doubt, the full article was read. All disagreements between researchers were discussed until a final consensus was reached.
- Minimum criteria were defined in the quality assessment to make the decision objective. Moreover, there were three possible answers to the questions posed in Table 4 (yes, partially, and no), rather than only two (yes and no), which minimizes the risk of disagreement.
- Two researchers conducted the quality assessment based on the quality questions posed. They discussed any disagreement that arose until the issue was resolved.

Data extraction bias: Next to finding and selecting all the relevant studies, data extraction was the most critical task in this study. To correctly extract data from these studies, two researchers read each paper independently and collected the data presented in Table 5 that are required to answer the research questions posed. The data extracted for each paper were compared and all disagreements were discussed by the researchers. However, data extraction bias may occur, especially when the accuracy values are extracted from a study using different model configurations. We believe that using the optimal configuration was a good choice.

Publication bias: Our review takes into account only ASEE studies, which means that the authors of the selected studies may have some bias towards ASEE. Consequently, there is a risk of overestimating the performance of ASEE methods, given that some authors might wish to show that their methods perform better than those of others.

3. Mapping results

This section presents and discusses the results related to the systematic mapping questions listed in Table 2. The classification schemes in this table that we used are defined as: (1) orthogonal (there are clear boundaries between categories, which makes classification easy); (2) based on an exhaustive analysis of existing literature in the ASSE field; or (3) complete (no categories are missing, and so existing papers can be classified). The classification of each of the selected papers can be found in Table C.22 in Appendix C.

3.1. Overview of the selected studies

Fig. 2 shows the number of articles obtained at each stage of the selection process. As can be seen in Fig. 2, the search in the four electronic databases resulted in 1657 candidate papers. Our inclusion and exclusion criteria were applied to identify those that were relevant, as many of the papers would not prove to be useful for addressing the research questions. This process left us with 104 relevant articles. As mentioned in Section 2.3, the selection was performed based on title, abstract, and keywords. If there was any doubt, the full article was read. Scanning of the reference lists of the selected papers revealed no additional relevant papers. At this point, we applied the quality assessment criteria to the remaining 104 relevant articles. This resulted in 65 articles of acceptable quality, almost 88% of them (57 out of 65) of high or very high quality – see Table 6.

3.2. Publication sources of the ASEE studies (MQ1)

Of the 65 selected papers, 27 (42%) were published in journals, 24 (37%) were presented at conferences, 12 (18%) were presented in symposiums, and 2(3%) were published in workshops. Table 7 shows the distribution of the selected papers across the publication sources. Sources with 4 or more papers on ASEE techniques were: the Empirical Software Engineering (EMSE) journal, the International Symposium on Empirical Software Engineering and Measurement (ESEM), IEEE Transactions on Software Engineering (IEEE TSE), the Journal of Systems and Software (JSS), the International Conference on Predictive Models in Software Engineering (PROMISE), and the International Symposium on Empirical Software Engineering (ISESE). If we consider that ESEM is the fusion of ISESE and METRICS conferences, ESEM will be the first publication source with 12 studies, followed by EMSE with 9 studies; hence, 32% of the papers included in our research were retrieved from these two sources.

3.3. Research approaches (MQ2)

As shown in Table 8, we identified five main research approaches that were applied in the selected studies: history-based evaluation (HE), solution proposal (SP), experimental (EXP), theoretical (TH), and review (RV). Other approaches are denoted OT. Table 8 shows that HE and SP were the most frequently employed approaches. Furthermore, the number of papers using these two approaches is increasing over time. Note that only 5% (3 out of 65) of selected studies are theoretical or review and the rest are empirically validated through history-based (94%) or experiment (5%) evaluations. According to Kitchenham et al. [14], all papers related to a topic area may be included in a systematic mapping study but only classification data about these are collected whereas in a SLR only empirical studies are considered. Hence, we have used the three theoretical/review paper (S21, S38, 45) to answer only the mapping questions of Table 1.

We investigated the use of the HE approach in the selected papers: 15 papers employed historical data to analyze the impact of dataset properties, such as missing data and outliers, on the accuracy of ASEE methods, while the remaining 46 papers used historical data to evaluate or compare the performance of ASEE methods with other estimation techniques. Regarding the type of historical data, most of the papers used professional or industrial software project datasets, such as Desharnais, ISBSG, Albrecht, and COCOMO. Student project data are rarely used. From the 61 papers included in the HE category, 23 datasets were used in 111 evaluations. Fig. 3 shows the distribution of the number of studies using HE over the datasets. Note that one study may involve more than one dataset. As can be seen, Desharnais (24 studies) was the dataset most frequently employed, followed by ISBSG (15 studies) and Albrecht (14 studies). Note, too, that we include: (1) studies that use industrial/professional projects, rather than student projects; and (2) studies that use MMRE, MdMRE, and/or Pred(25) to evaluate estimation accuracy (see Section 4.1 for more details).

From the results obtained, we can conclude that few research works deal with dataset properties such as categorical data and missing values. As well, there is a lack of in-depth studies on real-life evaluations of ASEE methods (i.e. evaluations in industrial settings). Moreover, most of the selected papers use historical data to evaluate ASEE methods, i.e. there was no research on how to evaluate ASEE methods in real-life contexts.

3.4. Contributions of the ASEE studies (MQ3)

Fig. 4 shows the classification of the selected studies based on their contribution type. Note that most of the papers are classified in the Technique contribution category (66%). As shown in Fig. 5 and 77% of these propose improvements to existing ASEE techniques (the improvement may target feature and case subset selection, feature weighting, outlier detection, or effort adjustment), while 23% develop a novel technique for predicting software effort using analogy (either alone or in combination with another technique). This illustrates that, in general, the analogy process is well defined for software effort estimation, but still needs improvement and refinement.

Table 6
Quality levels of the selected studies.

Quality level	Number of studies	Proportion (%)
Very high ($5 < \text{score} \leqslant 6$)	22	21.1
High ($4 < \text{score} \leqslant 5$)	35	33.7
Medium ($3 \leqslant \text{score} \leqslant 4$)	8	7.7
Low ($0 \leqslant \text{score} < 3$)	39	37.5
Total	104	100.0

In 14% of the selected papers, researchers compared their ASEE technique with other techniques in response to the inconsistent results reported in the ASEE literature on estimation accuracy. These conflicting results may be generated by a number of issues, including dataset sampling, analogy parameter configuration (feature selection, number of analogies, etc.), and evaluation techniques (jackknife method, *n*-fold cross validation, etc.). Note that, in addition to the 9 studies included in the Comparison contribution category, there are other studies in which this comparison is made, but they were included in the Technique contribution category, since their main focus is the development of new techniques or the improvement of existing ones.

Fig. 4 shows that there are few tools available for estimating software effort using analogy. In fact, of the 65 papers selected, only 9 studies (14%) propose new tools to implement ASEE techniques. This lack of ASEE tools may limit the use of ASEE in industry, given the need for such tools to make the ASEE process easier for practitioners. Note that some of the tools that have been developed are not available, and most only implement the classical ASEE techniques.

When investigating the relationship between research approaches and the contribution types of the selected studies, we observed that:

- 43 of the 47 selected studies in the SP approach category developed ASEE techniques, and only 9 of them proposed new tools to support their techniques;
- 42 of the 43 selected studies in the Technique contribution category were empirically validated using HE, and only 1 of them was validated by EXP;
- 7 of the 9 selected studies in the Comparison contribution category were empirically validated using HE, and only 1 of them was validated by TH and only 1 by OT (a survey).

This extensive use of historical data to evaluate ASEE techniques is encouraging for investigating the SLR questions in Table 3, which are, in general, answered through empirical research.

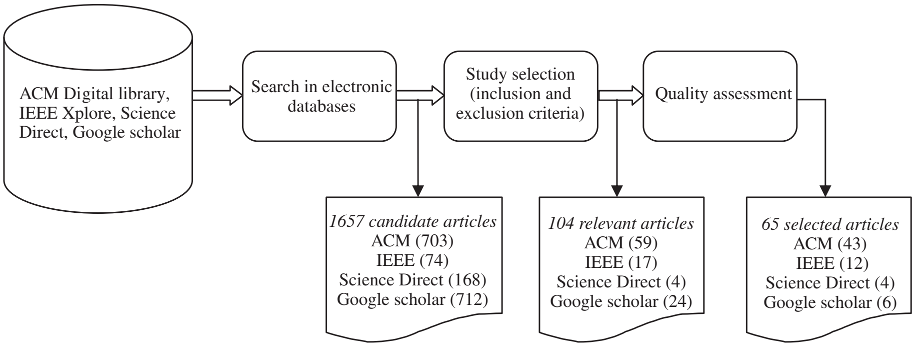

Fig. 2. Study selection process.

Please cite this article in press as: A. Idri et al., Analogy-based software development effort estimation: A systematic mapping and review, Inform. Softw. Technol. (2014), http://dx.doi.org/10.1016/j.infsof.2014.07.013

Table 7
Publication sources and distribution of the selected studies.

Publication source	Type	Number	Proportion (%)
Empirical Software Engineering (EMSE)	Journal	9	14
International Symposium on Empirical Software Engineering and Measurement (ESEM)	Conference	6	9
IEEE Transactions on Software Engineering (IEEE TSE)	Journal	5	8
Journal of Systems and Software (JSS)	Journal	4	6
International Conference on Predictive Models in Software Engineering (PROMISE)	Conference	4	6
International Symposium on Empirical Software Engineering (ISESE)	Conference	4	6
Information and Software Technology (IST)	Journal	3	5
Expert Systems with Applications (ESA)	Journal	2	3
Asia–Pacific Software Engineering Conference APSEC	Conference	3	5
International Conference on Evaluation and Assessment in Software Engineering (EASE)	Conference	2	3
International Software Metrics Symposium (METRICS)	Conference	2	3
Other		21	32

Table 8
Distribution of ASEE research approaches over the years.

Research approach	1992–1998	1999–2005	2006–2012	Total
HE	3	20	38	61
SP	3	11	33	47
EXP	0	3	0	3
TH	0	0	2	2
RV	0	1	0	1
OT	0	0	1	1

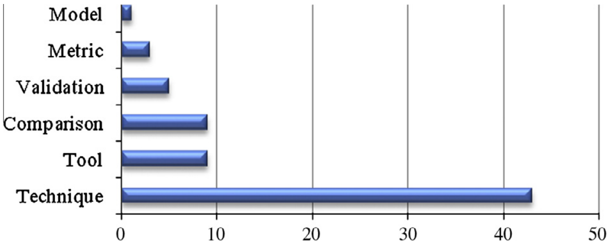

Fig. 4. Number of studies per contribution type.

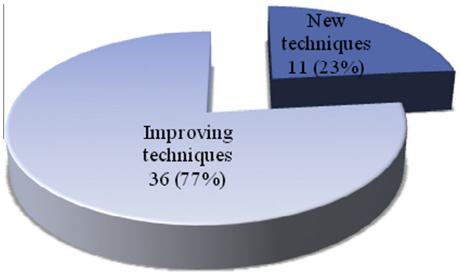

Fig. 5. Distribution of studies of the ‘Technique’ contribution type.

3.5. Techniques used in combination with ASEE methods (MQ4)

Various paradigms were used in combination with the ASEE techniques to overcome several challenges related to feature and case selection, similarity measures, and adaptation strategies. Fig. 6 shows that statistical methods (SM) and fuzzy logic (FL) are the most frequently used techniques, in combination with analogy (18% each), followed by genetic algorithms (GA) with 8%. Other paradigms were used less often, such as EJ, LSR, and GRA (3% each).

The most frequently used statistical methods were the following:

- Mantel test.
- Bootstrap method.
- Monte Carlo simulation.
- Principal Components Analysis.
- Regression toward the mean.
- Kendall’s coefficient of concordance.
- Pearson’s correlation.

The statistical methods most often used were the Mantel test and the Bootstrap method. The former was used to assess the appropriateness of ASEE techniques for a specific dataset and to address the problem of feature and case selection [20–23]. The latter was usually applied for model calibration and the computation of prediction intervals [24–27]. We investigated the use of FL in combination with ASEE in the selected studies: the main purpose of using FL was to handle linguistic attributes and to deal with

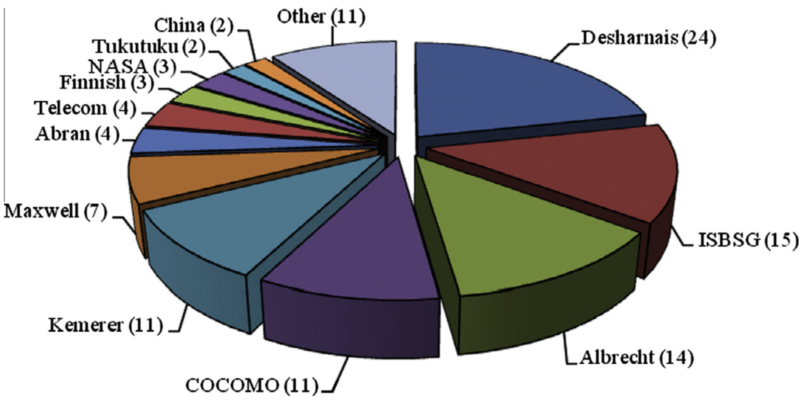

Fig. 3. Distribution of the HE research approach over the datasets.

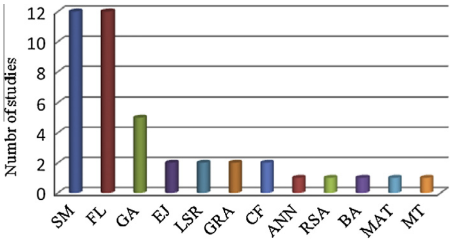

Fig. 6. Distribution of techniques used in combination with ASEE methods.

imprecision and uncertainty. Note that FL was employed in three phases: feature subset selection, similarity measurement, and case adaptation [4,28–37]. GA, which are based on the mechanism of natural evolution and the Darwinian theory of natural selection, were used in combination with analogy, especially for feature weighting, project selection, and effort adjustment [5,10,38,39]. Table 9 shows in detail the reasons why each technique was combined with analogy in the selected studies.

3.6. ASEE step classification (MQ5)

The ASEE process is generally composed of three steps:

- *Feature and case subset selection (FCSS)*: Feature and project selection, feature weighting, and the selection of other dataset properties, such as dataset size, outliers, feature type, and missing values.
- *Similarity evaluation (SE)*: Retrieval of the cases that are the most similar to the project under development using similarity measures, in particular the Euclidean distance.
- *Adaptation (AD)*: Prediction of the effort of the target project based on the effort values of its closest analogs. This requires choosing the number of analogs and the adaptation strategy. The number of analogs refers to the number of similar projects to consider for generating the estimates. Based on the closest analogs, the effort of the new project is derived using an adaptation strategy.

Fig. 7 shows the number of selected studies in which each of the above steps was performed. Note that one study may perform more than one step. Note, too, that the FCSS step was performed the most (63%), followed by the AD step (57%), and, finally, the SE step (34%). Regarding the FCSS step, there was significant interest on the part of study authors as to how to deal with missing values and category attributes. Case selection has also attracted considerable attention, since estimation accuracy may be influenced by outliers. As a result, several researchers have looked at feature selection and feature weighting in terms of improving estimation accuracy by considering the degree of relevance of each feature to the project effort. For the AD step, new effort adjustment techniques were investigated in most of the studies to capture the difference between the project being estimated and its closest analogs. In studies of the SE step, the authors were interested in how to measure the level of similarity between two software projects, especially when they are described by both numerical and categorical features. Table 10 shows in detail which steps of the ASEE process were performed in which studies and why.

In Fig. 8, the relationship between the technique used in combination with ASEE techniques and the targeted step is investigated. Our findings are summarized as follows:

- FL and SM were the most frequently used techniques in the FCSS step, followed by GA. FL was mainly used to handle categorical attributes and to deal with imprecision and uncertainty when describing software projects, whereas SM and GA were used to address different feature and case subset selection issues, such as feature weighting and case selection.
- To assess the similarity between software projects, most studies used FL and GRA to model and tolerate imprecision and uncertainty, in order to adequately handle both numerical and categorical data.
- The techniques most frequently used in combination with analogy in the AD step were FL and SM, followed by LSR. The main purpose of using FL in the third step was to propose new adaptation techniques. SM was used to model the calibration and computation of prediction intervals, whereas LSR was incorporated into ASEE techniques to deal with attributes that are linearly correlated with effort.

4. Review results

This section describes and discusses the results related to the systematic review questions listed in Table 3. These questions were aimed at analyzing ASEE studies from five perspectives: estimation accuracy, relative prediction accuracy, estimation context, impact of the techniques used in combination with ASEE methods, and ASEE tools. We discuss and interpret the results related to each of these questions in the subsections below.

4.1. Estimation accuracy of ASEE techniques (RQ1)

From the results of MQ2, ASEE technique evaluation is mainly based on historical software project datasets, rather than the use of a case study or an experiment (61 of 65). Their accuracy may therefore depend on several categories of parameters: (1) the characteristics of the dataset used (size, missing values, outliers, etc.); (2) the configuration of the analogy process (feature selection, similarity measures, adaptation formula, etc.); and (3) the evaluation method used (leave-one-out cross validation, holdout, *n*-fold cross validation, evaluation criteria, etc.). In the following subsections, we discuss the first and third categories of parameters, and those related to the analogy process are discussed in connection with RQ4 (Section 4.4).

Various datasets were used to construct and evaluate the performance of ASEE techniques in the 65 selected studies. Table 11 summarizes the most frequently used datasets, along with their description, including the number and percentage of selected studies that use the dataset, the size of the dataset, and the source of the dataset. Note that the Desharnais dataset is the most frequently used (35%), followed by the ISBSG dataset (15%). Note that the review takes into account only industrial/ professional datasets, that is, no in-house or student datasets were included. Table 11 is extracted from Fig. 3, the datasets for which there are fewer than 4 studies having been discarded.

Regarding evaluation techniques, the selected studies use several methods to assess the estimation accuracy of ASEE approaches. The most popular of these were leave-one-out cross validation (LOOCV) and *n*-fold cross validation ($n > 1$). LOOCV was applied in 58% of the studies, and $n > 1$ in 11% of the studies. The selection of criteria for defining an accuracy evaluation method for ASEE techniques is very challenging. In the selected studies, various criteria were used; in particular, the Mean Magnitude of Relative Error (MMRE), the Median Magnitude of Relative Error (MdMRE), and the percentage of predictions with an MRE that is less than or equal to 25% (Pred(25)). MMRE was used in 47 of the studies (72%), Pred(25) was used in 37 of the studies (57%), and MdMRE was used in 23 of the studies (35%). Consequently, we selected these criteria to answer RQ1.

Table 9
Purposes of using other techniques in combination with analogy.

Techniques used in combination with ASEE methods	Paper ID	Purpose
ANN: Artificial neural networks	S44	For non linear adjustment with learning ability and including categorical features
BA: Bees algorithms	S5	For effort adjustment (optimization of the number of analogies (K) and the coefficient values used to adjust feature similarity degrees from new case and other K analogies)
CF: Collaborative filtering	S39, S40	• To support non quantitative attributes • For missing value tolerance • For estimation at different object levels: requirement (RQ), feature (FT), and project (PJ).
EJ: Expert judgment	S56	To test whether or not tools perform better than people aided by tools
	S64	To test whether or not people are better at selecting analogs than tools
FL: Fuzzy logic	S8, S16	• To handle categorical and numerical attributes and deal with uncertainty • To propose a new approach to measure software project similarity (2 projects)
	S9	For feature subset selection
	S17, S18, S19, S21 S12	• To handle linguistic values and deal with imprecision and uncertainty • To propose a new ASEE technique using fuzzy sets theory • To deal with attribute measurement and data availability uncertainty • To propose a new similarity measure and adaptation technique
	S58	To identify misleading projects
FL + GRA: Fuzzy logic and grey relational analysis	S11	To reduce uncertainty and improve both numerical and categorical data handling in similarity measurement
	S10	To model and tolerate software project similarity measurement uncertainty (2 projects), when they are described by both numerical and categorical data
FL + GA: Fuzzy logic and genetic algorithms	S20	To deal with linguistic values and build fuzzy representations for software attributes
GA: Genetic algorithms	S43	For optimizing feature weights and project selection
	S14	For effort adjustment
	S51	For selecting the optimal CBR configuration (attribute weighting)
	S15	For deriving suitable effort driver weights for similarity measures
LSR: Least squares regression	S52, S53	To deal with variables that are linearly correlated with the effort
MT: Model tree	S6	As an adaptation technique (to deal with categorical attributes, minimize user interaction, and improve the efficiency of model learning through classification)
MAT: Multi-agent technology	S1	To address the problem of obtaining data from different companies
RSA: Rough set analysis	S39	For attribute weighing
Statistical method with Mantel correlation	S25, S27, S28, S29	• To provide a mechanism to assess the appropriateness of ASEE techniques for a specific dataset • To identify abnormal projects • To address the problem of feature subset selection
	S28	To incorporate joint effort and duration estimation into the analogy
Statistical method with Bootstrap method	S2	For calibrating the process of estimation by analogy and the computation of prediction intervals
	S61	For model calibration
	S54	To reduce the prediction error of ASEE techniques
Statistical method with Bootstrap method and Monte Carlo simulation	S60	To calculate confidence intervals for the effort needed for a project portfolio
Statistical method with Principal Components Analysis	S62	For feature weighting
Statistical method with Principal Component Analysis and Pearson correlation coefficients	S65	For feature selection and feature weighting
Statistical method with Kendall's coefficient of concordance	S10	For attribute weighting
Statistical method with Regression toward the mean	S22	To adjust the estimates when the selected analogs are extreme and the estimation model is inaccurate

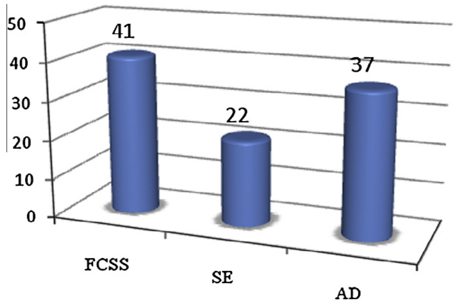

Fig. 7. Number of studies per step of an ASEE process.

The values of MMRE, MdMRE, and Pred(25), as extracted from the selected studies, are shown in Table D.23 of Appendix D. As mentioned above, for some studies, where we could not extract the values corresponding to each of these three criteria directly, we used the values of the optimal configuration (configuration with the best accuracy values) if there were different model configurations, and the means of the accuracy values if there were different dataset samplings.

To analyze the distribution of the MMRE, MdMRE, and Pred(25) of ASEE techniques, we drew box plots corresponding to each of these criteria using the estimation accuracy values of each selected study. As can be seen in Fig. 9, the medians of the accuracy values of ASEE techniques are around 42% for MMRE, 28% for MdMRE, and 49% for Pred(25). We recall that, unlike MMRE and MdMRE, a higher value of Pred(25) indicates better estimation accuracy. It can also be seen in Fig. 9 that, according to the MdMRE criterion, ASEE techniques are symmetrically distributed around the median, while the distribution of MMRE and Pred(25) indicates a positive skewness, since the medians are closer to the lower quartile. In addition, the Pred(25) and MdMRE values have high variations than those of MMRE, since the lower and upper quartiles are far from one another. Therefore, the boxes corresponding to Pred(25) and MdMRE are taller than that of MMRE. This is because the values used to draw the box plots come from different ASEE techniques applied on a variety of datasets using different configurations and evaluation methods.

Table 10
Steps performed and why.

Step	Paper ID	Purpose
1	S3, S4, S15, S27, S39, S43, S62, S65, S10, S51	Technique for feature weighting
	S9, S27, S28, S65, S29, S42, S11	Technique for feature subset selection
	S7, S23, S49, S46, S47	Impact of feature selection on accuracy
	S31	To compare 3 search techniques to obtain the optimal feature subset
	S20	To build a fuzzy representation for software attributes
	S12	To represent software attributes using fuzzy numbers
	S25	To compare the results of the feature selection procedure of Analogy-X with that of ANGEL
	S27, S28, S29	Technique for outlier detection
	S35, S43, S58	Approach for project selection
	S33	To apply the easy path principle to design a new method for project selection
	S37	Impact of missing values on accuracy
	S24	To develop a new method to generate synthetic project cases to improve the performance of ASEE
2	S10, S11, S8, S16, S17, S12	To develop an approach to measure similarity
	S48, S49, S50	To compare different similarity measures
	S2	To choose an appropriate distance metric
3	S33, S34, S2	Approach for choosing the optimal number of analogies
	S7, S23, S36, S46, S47, S48, S49, S50	To compare the use of different numbers of analogies
	S6, S11, S12, S30	To develop an adaptation technique
	S7, S23, S48, S49, S50	To compare the use of various adaptation strategies
	S5, S14, S22, S44	Technique for effort adjustment
	S2, S41, S19, S60	Uncertainty assessment
	S54	Use of an iterated bagging procedure to reduce the prediction error of ASEE
	S57	Impact of using homogeneous analogs on estimation reliability
	S63	Method of eliminating outliers from the neighborhoods of a target project when the effort is extremely different from that of other neighborhoods
All	S21	To compare Radial Basis Function neural networks and Fuzzy Analogy
	S1, S13, S18, S40, S52, S53, S55, S59, S64	To develop a new ASEE technique, or a tool implementing an ASEE technique
	S38	Model development
	S61	Calibration of the ASEE method, detection of the best configuration of the ASEE method options

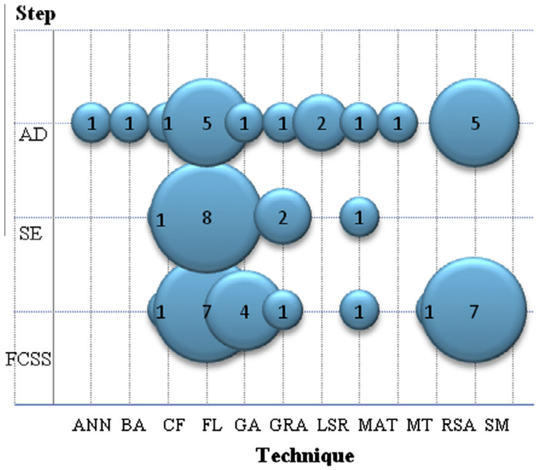

Fig. 8. Techniques used in combination with analogy for each step.

Table 11
Datasets used for ASEE validation.

Dataset	Number of studies	Proportion	Number of projects	Source
Desharnais	24	37	81	[40]
ISBSG	15	23	>1000	[41]
Albrecht	14	21	24	[42]
COCOMO	11	17	63	[43]
Kemerer	11	17	15	[44]
Maxwell	7	11	63	[45]
Abran	4	6	21	[46]
Telecom	4	6	18	[47]

To further analyze the estimation accuracy of ASEE methods, Table 12 provides the detailed statistics of MMRE, MdMRE, and Pred(25) for each of the most frequently used datasets. In general, for all the datasets except Maxwell, the mean of the prediction accuracy values varies from 37% to 52% for MMRE, from 19% to 35% for MdMRE, and from 45% to 62% for Pred(25). This indicates that ASEE methods tend to yield acceptable estimates.

4.2. Accuracy comparison of ASEE techniques with other ML and non-ML models (RQ2)

The ASEE techniques were compared with eight ML and non-ML models: Regression (SR), COCOMO model (CCM), Expert Judgment (EJ), Function Point Analysis (FP), Artificial Neural Networks (ANN), Decision Trees (DT), Support Vector Regression (SVR), and Radial Basis Function neural networks (RBF). Figs. 10–12 show the results of comparing these eight models with ASEE techniques with respect to the MMRE, MdMRE, and Pred(25) criteria respectively. This was achieved by counting the number of evaluations in which an ASEE technique outperforms (or underperforms) one of these eight techniques based on a specific evaluation criterion. Note that for Figs. 10–12, the blue[1] bars indicate the number of evaluations suggesting that ASEE techniques are more accurate, and the green bars indicate the number of evaluations suggesting that ASEE techniques are less accurate. The details of the comparison can be found in Tables D.24 and D.25 of Appendix D.

Regarding the comparison with non ML techniques, most studies compared ASEE methods with the regression model (38 evaluations). As can be seen from Figs. 10–12, ASEE methods outperform regression based on the three criteria used. With respect to ML techniques, ANN was the most frequently compared with ASEE methods (16 evaluations), followed by DT (11 evaluations). Similarly, the results suggest that ASEE methods are more accurate than ANN and DT in terms of MMRE, MdMRE, and Pred(25). These

[1] For interpretation of color in Figs. 10–12, the reader is referred to the web version of this article.

Please cite this article in press as: A. Idri et al., Analogy-based software development effort estimation: A systematic mapping and review, Inform. Softw. Technol. (2014), http://dx.doi.org/10.1016/j.infsof.2014.07.013

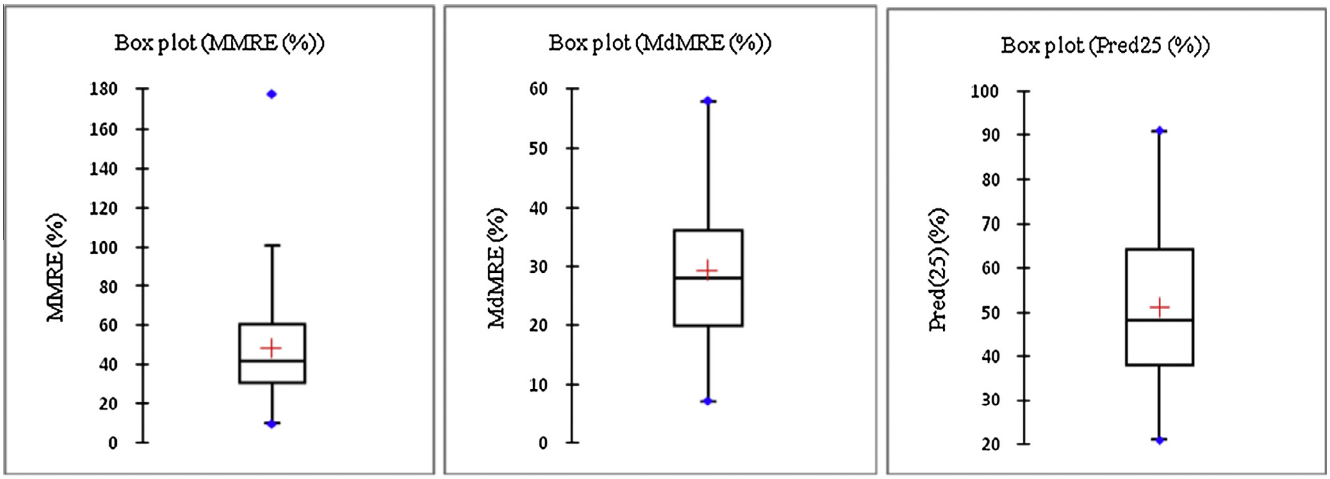

Fig. 9. Box plots of MMRE, MdMRE, and Pred(25).

Table 12
Statistics related to MMRE, MdMRE, and Pred(25) for each dataset.

Dataset	MMRE					MdMRE					Pred(25)				
	No. of values	Min (%)	Max (%)	Mean (%)	Median (%)	No. of values	Min (%)	Max (%)	Mean (%)	Median (%)	No. of values	Min (%)	Max (%)	Mean (%)	Median (%)
Desharnais	24	11.30	71.00	44.17	41.95	12	7.20	37.08	26.47	31.04	18	32.00	91.10	48.81	43.50
ISBSG	15	13.55	177.79	51.98	28.70	11	17.80	57.98	34.70	34.00	14	22.73	84.00	54.47	57.30
Albrecht	14	30.00	100.60	49.50	46.60	7	19.90	48.00	27.81	25.00	12	28.60	70.00	48.59	50.00
COCOMO	10	18.38	151.00	48.73	41.37	6	13.90	35.04	22.60	21.13	9	21.00	89.41	57.10	61.00
Kemerer	11	14.00	68.10	43.04	40.20	3	24.24	33.20	27.85	26.10	8	33.40	83.33	54.92	49.80
Maxwell	**7**	**28.00**	**120.59**	**68.13**	**69.80**	**5**	**18.60**	**53.15**	**36.72**	**45.00**	**5**	**29.00**	**67.00**	**43.31**	**35.00**
Abran	4	19.72	52.00	37.07	38.29	3	9.09	36.00	19.77	14.23	4	43.00	71.43	61.96	66.71
Telecom	4	36.70	60.30	43.60	38.70	0	N	N	N	N	2	44.00	46.67	45.33	45.33

Bold values indicate the low obtained accuracy values on the Maxwell dataset.

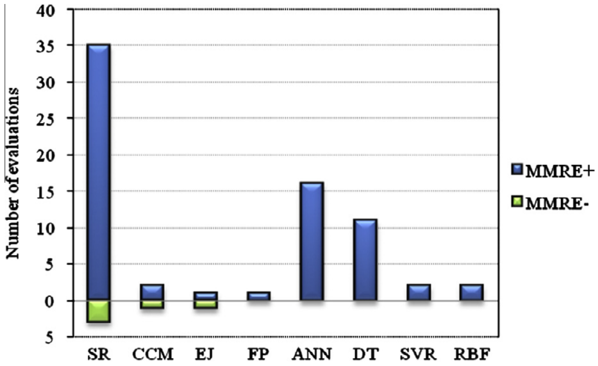

Fig. 10. Comparison of the MMRE of ASEE techniques with that of the other models ("MMRE+" indicates that ASEE techniques are more accurate, "MMRE−" indicates that the other model is more accurate).

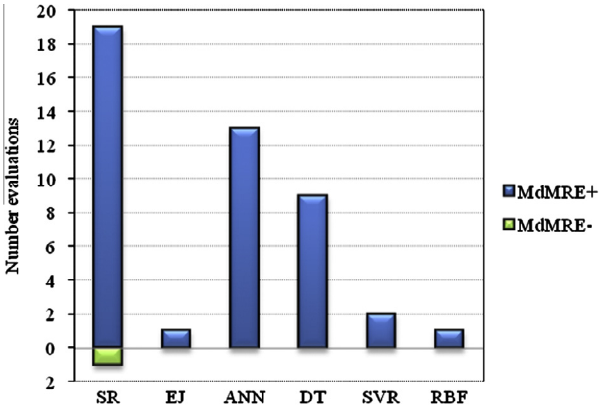

Fig. 11. Comparison of the MdMRE of ASEE techniques with that of the other models ("MdMRE+" indicates that ASEE techniques are more accurate, "MdMRE−" indicates that the other model is more accurate).

findings are highly consistent with the results reported in [1], which suggests that ASEE methods outperform Regression, ANN, and DT.

Unlike the comparison with Regression, ANN, and DT, few studies have compared ASEE methods with the remaining five techniques (i.e. COCOMO, FP, EJ, SVR, and RBF). In fact, fewer than 5 evaluations compare ASEE methods with these techniques, making it difficult to generalize the results obtained.

In general, the overall picture suggests that ASEE techniques outperform the eight techniques based on, MMRE, MdMRE, and Pred(25) criteria, especially for Regression, ANN, and DT, for which there were enough evaluations. Note that the results in this review are taken from ASEE studies, which means that their authors could have a favorable bias towards ASEE techniques. However, except for SVR, the same results were obtained in [1] by Wen et al., who conducted their systematic review based on eight ML studies.

4.3. Estimation context of ASEE techniques (RQ3)

Since software effort estimation studies using different techniques have produced varying results, it is of greater interest to identify the favorable estimation context of each technique, rather than to look for the best prediction model. Wen et al. have studied and compared the estimation contexts of different ML effort

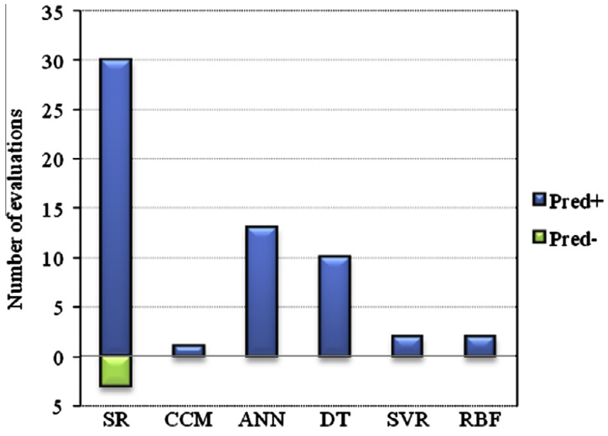

Fig. 12. Comparison of the Pred(25) of ASEE techniques with that of the other models ("Pred+" indicates that ASEE techniques are more accurate, "Pred−" indicates that the other model is more accurate).

estimation techniques, including ASEE techniques, based mainly on four characteristics related to the dataset used: dataset size, outliers, categorical features, and missing values. They found that while ASEE techniques deal quite well with small datasets that may contain outliers, they do not deal well with categorical attributes and missing data. Our study focuses on these issues, in order to confirm or refute the findings of Wen et al. With this objective, we extracted and investigated the strengths and weaknesses reported in the selected studies on ASEE techniques – see Tables 13 and 14 for details. We found that the information reported is mainly related to dataset properties, which seem to have a significant impact on the prediction accuracy of ASEE techniques.

Among the dataset properties, size is considered to be an influential factor in an ASEE technique, and several studies (S4, S59) have investigated its effect on prediction accuracy. However, contradictory results were obtained. For example: Briand et al. [48] found that ASEE techniques are less robust than other models when large heterogeneous datasets are used, whereas Shepperd and Kadoda [49] claim that ASEE techniques benefit from larger training sets. Considering the results obtained in Section 4.1 (Table 12), it seems difficult to claim that ASEE techniques should be favored in either case, since acceptable estimates were obtained for all eight datasets, which vary greatly in size.

One of the major challenges for ASEE techniques is to produce accurate estimates when the dataset contains categorical features or missing values, or both. In fact, classical ASEE methods can only correctly handle categorical data that consists of binary valued variables, and cannot tolerate missing values. As a result, several techniques have been proposed to extend the traditional ASEE method. Li et al. [7], for example, developed a new technique, called AQUA, which combines CBR and collaborative filtering. Their method supports non quantitative attributes and can tolerate missing values. Idri et al. [31] have also proposed a new technique, Fuzzy Analogy, which extends the classical ASEE method by integrating FL to handle categorical features. Similarly, Azzeh et al. [28] have proposed two approaches to measure the similarity between (two) software projects by describing them in terms of either numerical features or categorical features, or both, using fuzzy C-means clustering and FL.

An important aspect of ASEE techniques is that they can be applied even if the dataset contains outliers, and several techniques have been proposed for project selection in the ASEE process. For example, Keung et al. [23] developed a new method, called Analogy-X, to identify abnormal cases in a dataset using Mantel's correlation and randomization test.

There are characteristics other than dataset characteristics to be considered when applying an ASEE technique. We summarize these in Tables 13 and 14. For example, an ASEE technique is the better choice when the relationship between effort and software attributes is not strongly linear. This is because ASEE are intuitive methods that can be easily understood and explained to practitioners and other users; they can be used with partial knowledge of the target project at an early stage of a project; they allow a number of design decisions to be made; and they cannot generate an estimate without a historical dataset.

To summarize, one ASEE technique alone may not be the best estimation method in all contexts. However, in any context, an appropriate effort estimation model can be built by combining an ASEE technique with other techniques to overcome the weaknesses listed in Table 14. The benefit of combining different models is supported by many studies. In [50], Shepperd recommends combining techniques if no dominant technique can be found. In [51], Jørgensen argues that there is a potential benefit to using more than one model. In [5,8,9,30,10,38], it is shown that combining ASEE methods with other techniques may generate better estimates than using other estimation models alone. Below, we discuss

Table 13
Advantages of an ASEE technique.

Advantages	Supporting study
Can model the complex relationship between effort and other software attributes	S1, S9, S10, S11, S18, S19, S20, S23, S33, S59
Solutions from analogy-based techniques more readily accepted by users	S2, S10, S14, S15, S34, S36, S40, S53, S59
Transparent by nature, with a process that can be easily understood and explained to practitioners and other users	S10, S19, S20, S21, S35, S49, S55, S64, S65
Intuitive	S29, S36, S45, S53, S54, S55, S57, S59
Mimics the human problem solving approach	S1, S11, S14, S15, S34, S35, S55, S59
Simple and flexible	S1, S3, S4, S36, S53, S54
Can handle both quantitative and qualitative data	S1, S10, S36, S53, S54
Can be used with partial knowledge of a target project at an early stage of the project	S13, S29, S40, S59, S64
Can deal with poorly understood domains	S59, S38, S40, S64, S65
Has the potential to mitigate problems with outliers	S5, S40, S64, S65
Can handle failed cases (i.e. those for which an accurate prediction was not made)	S1, S59
Can use an existing solution and adapt it to the current situation (even providing accurate estimates even with another organization's data)	S1, S64
Can be implemented very quickly	S1
May be better for relatively small datasets	S2
Particularly helpful for cross source studies, as it is based on distances between individual project instances	S32
Avoids the problems associated with knowledge elicitation, and with extracting and codifying it	S59
Makes no assumptions about data distributions or an underlying model, unlike other predictors	S33

Table 14
Limitations of an ASEE technique.

Limitations	Supporting study
Potentially vulnerable to erroneous, irrelevant, or redundant data	S7, S15, S16, S31
A classical ASEE method cannot handle categorical variables	S8, S16, S18, S40
A classical ASEE method cannot handle missing values	S8, S37, S40
Cannot deal with imprecision and uncertainty	S18, S19, S20
Has no means of assessing dataset quality and will always endeavor to predict, no matter what the circumstances	S27, S28, S29
Use involves several design decisions	S23, S53
Cannot estimate without a stored software project dataset	S40, S63
Application requires datasets maintained and updated according to changes in the development process	S2
Computationally intensive	S16
A more complex technology	S23
Quality of the estimates for a target project strongly reliant on the quality of the historical data	S40
Accuracy of the method dependent on the ability to find analogies from the dataset through appropriate similarity measures	S40
Requires specific adjustments that have to be examined in order to calibrate the procedure and produce accurate predictions	S53

the improvement in accuracy achieved by combining other techniques with ASEE methods (RQ4).

4.4. Impact of combining an ASEE with another technique (RQ4)

In this section, we analyze the impact on estimate accuracy when ASEE methods are combined with the techniques identified in Section 3.5. Table 15 provides the accuracy improvement statistics with respect to MMRE, MdMRE, and Pred(25) for the techniques used in combination with ASEE methods. The original values, showing the accuracy improvement in terms of MMRE, MdMRE, and Pred(25), are presented in Table D.26 of Appendix D. It is worth noting that there were some studies in which the accuracy of ASEE combination techniques was compared without taking into account their performance relative to that of an ASEE technique alone, and so no accuracy improvement values could be provided for these studies.

Table 16 shows the number of studies investigating each technique used in combination with ASEE methods (also shown in Fig. 6), the number of studies providing an accuracy comparison, and the number of evaluations carried out in the studies. For example, of the 12 selected studies on SM-ASEE techniques, only 3 of them compared the prediction accuracy of an SM-ASEE technique with that of an ASEE technique alone, and only 6 of them evaluated the estimation accuracy of the combined model. In contrast, there were some techniques used in combination with ASEE methods for which the number of evaluations conducted was much higher than the number of studies on ASEE methods incorporating these techniques. This was mainly the case for MT and BA, for which there was only 1 study for each (S6 and S5 respectively) comparing the accuracy of an ASEE technique with that of an MT-ASEE technique and a BA-ASEE respectively, but 7 and 6 evaluations were conducted respectively. Note that, in order to adequately evaluate the impact of each technique used in combination with ASEE methods, we have distinguished cases where more than one technique is combined with an ASEE method from those where only one technique is combined. For example, the FL line in Table 15 indicates accuracy values when combining only FL with an ASEE technique, whereas the FL + GRA line indicates accuracy values when combining both FL and GRA with an ASEE technique. Finally, note that the EJ technique is used least in combination with ASEE methods (1 study with 1 evaluation).

As can be seen from Table 15, taking into consideration the number of evaluations and based on the median of the MMRE, MT is the technique that improves the accuracy of ASEE methods the most (59.42% improvement), followed by CF combined with RSA (51.85%) and LSR (41.03%). Based on the median of the MdMRE, MT has the greatest impact (67.75%), followed by FL

Table 15
Descriptive statistics of accuracy improvement in terms of MMRE, MdMRE, and Pred(25) for each technique used in combination with ASEE methods.

Technique	MMRE improvement					MdMRE improvement					Pred(25) improvement				
	No. of values	Min (%)	Max (%)	Mean (%)	Median (%)	No. of values	Min (%)	Max (%)	Mean (%)	Median (%)	No. of values	Min (%)	Max (%)	Mean (%)	Median (%)
ANN	4	13.33	52.87	28.44	23.78	4	23.81	41.86	30.48	28.12	4	5.88	66.67	29.50	22.73
BA	6	27.21	75.37	43.22	32.48	0	N	N	N	N	6	16.75	219.69	89.51	63.64
CF	2	−35.54	77.42	20.94	20.94	0	N	N	N	N	1	108.33	108.33	108.33	108.33
CF + RSA	3	7.81	69.35	43.00	51.85	0	N	N	N	N	3	16.67	107.5	61.39	60.00
EJ	1	11.69	11.69	11.69	11.69	1	1.92	1.92	1.92	1.92	0	N	N	N	N
FL	**8**	**2.38**	**77.19**	**30.12**	**26.23**	**7**	**−5.72**	**41.12**	**24.69**	**27.27**	**9**	**0**	**181.61**	**62.86**	**50.76**
FL + GRA	7	19.90	70.42	34.04	31.38	7	−23.39	76.16	34.12	40.80	7	−14.11	112.55	40.04	34.31
GA	**7**	**27.12**	**58.40**	**40.50**	**38.78**	**7**	**19.70**	**45.95**	**34.23**	**37.93**	**7**	**56.41**	**400.00**	**172.57**	**100.00**
LSR	4	11.81	65.87	39.93	41.03	4	15.13	59.74	32.93	28.43	4	33.34	57.16	41.45	37.65
MT	7	32.78	72.95	54.58	59.42	7	−3.98	75.42	56.30	67.75	7	0	409.01	165.89	129.01
SM	**6**	**4.62**	**35.00**	**17.43**	**15.94**	**2**	**8.82**	**27.78**	**18.30**	**18.30**	**4**	**10.26**	**23.53**	**17.20**	**17.51**

Bold values indicate the best accuracy improvement obtained when combining techniques with analogy-based effort estimation methods.

Table 16
Number of studies with accuracy comparison, and number of evaluations for each technique used in combination with ASEE methods.

	ANN	BA	CF	CF + RSA	EJ	FL	FL + GRA	GA	LSR	MT	SM
No. of studies	1	1	2	1	1	10	2	5	2	1	12
No. of studies with accuracy comparison	1	1	1	1	1	**4**	2	**4**	2	1	**3**
No. of evaluations	4	6	2	3	1	**9**	7	**7**	4	7	**6**

Bold values indicate the techniques that have been frequently combined with analogy-based effort estimation methods.

Table 17
ASEE tools.

Tool	Author(s)	Year	Studies using the tool	Description	References
ANGEL (ANaloGy softwarE tooL)	Shepperd, Schofield, and Kitchenham	1996	S7, S9, S10, S12, S23, S25, S28, S31, S36, S39, S40, S42, S46, S47, S49, S56, S59, S64, S65	This tool uses a brute-force approach or an exhaustive search of all possible permutations to select the optimal subset of features based on the overall performance evaluation criteria, like MMRE. The similarity between projects is calculated using the Euclidean distance. The adaptation strategies implemented in the tool are: simple average, distance weighted, rank weighted, maximum distance, and adjusted distance	[52]
ESTOR	Mukhopadhyay, Vicinanza, and Prietula	1992	S39, S40, S55, S64	This tool is an early implementation of the ASCE system. It was developed to examine the feasibility of CBR in software cost estimation. The features used in this tool are function point components and the inputs of the intermediate COCOMO model. The attribute values of each project are manually typed into the system. The similarity between projects is calculated using the Euclidean distance. The effort values of the closest analogs are adjusted to take into account the differences between the new project and its closest analogs	[53]
CBR-WORKS	Schulz	1999	S48, S49, S50	This tool is a commercial CBR environment providing important features for modeling, maintaining, and consulting a case base. CBR-Works does not provide the feature subset selection option. An important feature of the tool is that it offers various retrieval algorithms such as Euclidean distance, average similarity, and maximum distance. In addition, a variety of adaptation strategies can be used, such as the mean of the closest cases, the median of the closest cases, and the inverse rank weighed mean	[54]
F_ANGEL	Idri and Abran	2001	S17, S18, S20	This tool is a software prototype developed with Matlab 7.0. It implements the Fuzzy Analogy approach, which is based on estimation by analogy and fuzzy set theory. The tool does not offer the feature subset selection option. The attributes describing software projects are represented by fuzzy sets, rather than classical intervals using the fuzzy C-means clustering algorithm and a real coded GA. To measure the similarity between two projects, the tool employs a set of new measure based on FL. Thereafter, the effort value of the new project is calculated using the weighted mean of its closest analogs. The weights used in the case adaptation use fuzzy set theory	[32]
BRACE (bootstrap based analogy cost estimation)	Stamelos, Angelis, and Sakellaris	2001	S2, S61	This tool supports the practical application of the analogy-based method using a Bootstrap approach. Bootstrap is used for method calibration and the calculation of confidence intervals. The calibration of the ASCE method is aimed at choosing the best combination of distance metrics (e.g. Euclidean distance, Manhattan distance), the number of analogies (one or more), the adaptation strategy (mean or median), and size adjustment (yes or no)	[55]
AMBER	Auer and Biffl	2004	S3, S4	This is a Java command line tool which facilitates batch processing. It implements Auer's brute-force approach for weighting project feature dimensions for analogy. The principle of AMBER's feature weighting approach is similar to the brute force feature selection algorithm implemented in the ANGEL tool. AMBER selects the optimal subset of features based on the overall performance evaluation criteria, such as MMRE	[56]
TEAK (Test Essential Assumption Knowledge)	Kocaguneli, Menzies, Bener, and Keung	2012	S32, S33	This an ASCE system which uses an easy path principle. It was designed to avoid high computational cost and to find the insights that simplify effort estimation. TEAK's design applies the easy path in five steps [57]: (1) select a prediction system; (2) identify the predictor's essential assumption(s); (3) recognize when those assumption(s) are violated; (4) remove those situations; and (5) execute the modified prediction system	[57]
FACE (Finding Analogies for Cost Estimation)	Bisio and Malabocchia	1995	S13	This tool was implemented based on the commercial tool, CBR-Express. In FACE, each case is assigned a similarity score between 0 and 100, according to its degree of similarity with the target project. To determine the closest analogs to the new project, the tool identifies the projects with a score higher than a given threshold (θ). These projects (called θ-cases) are used to estimate the effort for the new project using the size/effort ratio. The tool was assessed using the COCOMO dataset	[58]
ACE (Analogical and Algorithmic Cost Estimator)	Walkerden and Jeffery	1999	S64	This tool estimates the effort of the target project by selecting its closest analogs. Thereafter, the effort value of the most similar project is adjusted to take into account the difference in size between the target project and its closest analog. To determine the closest analog to the new project, ACE ranks each project in the dataset across the set of the search features based on the difference between the new project and each historical project. The closest analog is the project with the lowest rank over all the search features	[59]

combined with GRA (40.80%) and GA (37.93%). Based on the arithmetic median of Pred(25), ASEE techniques are improved the most by MT (129.01% improvement), followed by CF (108.33%) and GA (100.00%).

In order to avoid bias stemming from the use of many evaluations from the same study, we analyzed the accuracy improvement of the techniques used in combination with ASEE methods taking into consideration the number of studies, rather than the number of evaluations. As shown in Table 16, SM, FL, and GA are the techniques most often combined with ASEE methods. The FL, GA, and SM lines in Table 15 show that, for the three accuracy criteria MMRE, MdMRE, and Pred(25), GA is the technique that improves the accuracy of ASEE methods the most, followed by FL and SM.

In summary, our results suggest overall that all the techniques listed in Section 3.5 improve the estimation accuracy of ASEE methods, especially GA and FL, which are supported by 4 studies each. There is much less improvement in the accuracy of ASEE techniques when combined with SM. This may be caused by the complexity of relationships between software project attributes, which would indicate that using ML rather than non ML techniques to address these issues would be preferable. Moreover, from the findings in Fig. 8, FL seems to be a promising technique to be combined with ASEE methods to improve their performance, since it could be used in all three steps of the analogy process (FCSS, SE, and AD). In contrast, GA was mainly used in the selected studies to solve problems in the FCSS step. However, owing to the insufficient number of studies evaluating the impact of all the techniques used in combination with ASEE methods, these results need to be investigated in further research.

4.5. ASEE tools (RQ5)

ASEE techniques are computationally intensive, and they require software tools for their use. Nine ASEE tools were identified in the selected studies. Table 17 lists these tools with a short description of each. ANGEL is the tool used most often, followed by ESTOR.

ANGEL was developed by Shepperd et al. (1996) at Bournemouth University. This tool uses Euclidean distance to find the projects closest to the target project. An important feature of ANGEL is its ability to identify the optimal subset of features to use to generate estimates. However, this task can be time-consuming, especially when a large number of attributes is involved, since ANGEL uses either a brute force algorithm or an exhaustive search of all possible combinations.

ESTOR was developed by Mukhopadhyay et al. (1992). This tool also assesses the similarity between two projects using the Euclidian distance. However, unlike ANGEL, ESTOR assumes that the estimator should choose a specific set of features to use for the estimation process. Indeed, the features used in ESTOR are function point components and the inputs of the intermediate COCOMO model.

There seem to be few ASEE tools in use, based on the results we obtained. This scarcity of ASEE tools may limit the use of ASEE techniques by practitioners, given that ASEE tools are required in order to apply ASEE techniques. Furthermore, most of the available tools implement the classical ASEE methods, which have not incorporated other techniques, such as FL and GA, to overcome the weaknesses of these methods.

5. Summary and implications for research and practice

A summary of the obtained results as well as our recommendations for researchers are given as follows:

Research approaches: Our review has revealed that the history-based evaluation of ASEE techniques is the most frequently applied approach. History-based evaluation is used either to analyze the impact of dataset properties on the accuracy of ASEE techniques or to evaluate or compare the performance of ASEE techniques with other effort estimation techniques. The review has found that there is a lack of in-depth studies on how to evaluate ASEE techniques in real-life contexts. It is, therefore, hoped that case studies and real-life evaluations of ASEE techniques in industry will become more attractive for software effort estimation researchers. In addition, most of the datasets used are too obsolete to be representative of recent trends in software development. Consequently, we suggest that ASEE researchers take into account not only the availability of the datasets, but also how representative they are.

Contributions of the ASEE studies: As has been observed, the main contribution of most papers is the development of new techniques, especially to improve the prediction accuracy of existing ASEE methods. Few tools implementing ASEE techniques were developed. It is perhaps not surprising that the use of ASEE techniques among practitioners is so limited. To address this issue, we recommend that researchers implement their ASEE techniques and provide guidelines on how to use these tools in industry.

Techniques *used in combination with ASEE methods*: This review has shown that statistical methods and fuzzy logic are the most frequently used techniques, in combination with analogy, followed by genetic algorithms. Some other techniques, such as association rules and Bayesian networks were not used in combination with analogy. Therefore, researchers are encouraged to investigate the impact that these techniques may have when used in combination with ASEE techniques.

ASEE step classification: FCSS was the most investigated step followed by AD and SE steps. Several techniques were used to address some issues related to each step. This review recommends more research on the use of FL to deal with problems related to the three steps of an ASEE method, GA for the FCSS step, and SM for the FCSS and AD steps. Regarding techniques such as ANN, RSA, BA, MAT, and MT, more studies are needed to determine in which ASEE steps they may be useful.

Estimation accuracy *of ASEE techniques*: The overall picture suggests that ASEE techniques tend to yield acceptable results. However, the obtained results are mainly based on historical datasets of software project. It is therefore, recommended to perform further research works using case studies, experiments and real-life evaluations of ASEE techniques in industry.

Accuracy comparison of ASEE techniques with other ML and non-ML models: We have determined that ASEE techniques are usually more accurate than eight other models, both ML and non ML, especially when techniques like FL and GA are incorporated; however accuracy comparisons are still a challenge. The limited number of studies on ASEE methods combined with these techniques may account for these inconclusive results. Researchers are encouraged to conduct further studies and experiments to address this issue.

Estimation context of ASEE *techniques*: Researchers should be aware of the impact that dataset properties may have on the results of constructing and evaluating ASEE techniques. Although we have determined in this review that ASEE techniques deal adequately with both small and large datasets that may contain outliers, other dataset properties still represent serious challenges for ASEE techniques. For example, few research works have studied the limitations of categorical features and missing data. It would be beneficial for the ASEE research community to address these limitations, since most of the available datasets contain a number of categorical data and missing values.

Impact of combining an ASEE with another technique: The results suggest overall that the estimation accuracy of ASEE methods is improved when used in combination with other techniques. As

Table A.18
Research approaches.

Research approach	What it is
HE	A study evaluating an existing ASEE technique, or one of its specific steps (e.g. similarity measurement)
SP	A study in which a new ASEE technique or tool is developed. A new technique to predict software effort using analogy (either alone, or in combination with other techniques), or to improve an existing ASEE technique
CS	An empirical evaluation of an ASEE technique based on a case study (real-life evaluation)
EXP	An empirical method applied under controlled conditions to evaluate an existing ASEE technique
TH	A study using a non empirical research approach, or evaluating the properties of ASEE techniques theoretically
RV	A primary study in which ASEE papers are reviewed
SV	A study providing a comprehensive survey of ASEE techniques
OT	A study using another research approach

Table A.19
Contribution types.

Contribution	What it is
Technique	A new ASEE technique, or an existing ASEE technique which has been improved
Tool	A new tool implementing an ASEE technique
Comparison	A comparison of different ASEE configurations, or a comparison of an existing ASEE technique with other software effort estimation techniques
Validation	An evaluation of the performance of an existing ASEE technique using one historical dataset
Metric	A new means of evaluating the performance of an ASEE technique, or to measure project similarity
Model	A new analogy-based method of software effort evaluation, e.g. a decision-centric model
Other	Another type of contribution

Table B.20
List of known existing papers used to validate the search string.

Id of existing paper	Database before search	Database after search	Id of existing paper	Database before search	Database after search
S2	ACM	ACM	S27	ACM	ACM
S4	ACM	ACM	S28	ACM	ACM
S8	ACM	ACM	S29	ACM	ACM
S12	ACM	ACM	S31	Google Scholar	ACM
S16	IEEE Xplore	IEEE Xplore	S39	ACM	ACM
S17	Google Scholar	IEEE Xplore	S40	ACM	ACM
S18	Google Scholar	IEEE Xplore	S48	Google Scholar	IEEE Xplore
S19	Google Scholar	Google Scholar	S49	IEEE Xplore	IEEE Xplore
S20	Google Scholar	Google Scholar	S50	Google Scholar	ACM
S21	IEEE Xplore	IEEE Xplore	S55	Google Scholar	ACM
S23	Google Scholar	Google Scholar	S59	ACM	ACM
S26	ACM	ACM			

has been found, SM improves the accuracy of ASEE techniques much less than the other techniques. This suggests that using ML rather than non ML techniques in combination with analogy would be preferable, in particular, fuzzy logic, genetic algorithms, the model tree, and the collaborative filtering. It is worth noting that, before making any decision on the use of an ASEE technique, practitioners need to determine which techniques should be combined with ASEE methods to overcome their limitations (categorical data, missing values, features selection, etc.), in order to adapt the ASEE method to their context.

ASEE tools: The review has identified nine tools to predict software effort using ASEE techniques. Among them, ANGEL and ESTOR are the tools most frequently employed. Based on the obtained results, most of the existing tools implement classical ASEE techniques. Therefore, it is suggested to the researchers to implement their ASEE techniques incorporating other techniques, such as FL and GA, to facilitate and encourage the use of ASEE among practitioners.

6. Study limitations

In this review MMRE, MdMRE, and Pred(25) were used as prediction accuracy indicators. These three indicators are all derived using the magnitude of the relative error (MRE). There has been some criticism of these indicators, in particular that they ignore the importance of the dataset quality, and implicitly assume that the prediction model can predict with up to 100% accuracy at its maximum for a specific dataset [60]. In addition, the MMRE criterion has been criticized for being unbalanced in many validation circumstances and for penalizing overestimates more than underestimates [3,61]. Nevertheless, we adopted these three criteria in our study, as they are the most commonly used in the selected studies. This allowed us to synthesize and compare the results obtained in the selected papers.

The estimation accuracy values were extracted from studies using different ASEE techniques (the traditional ASEE technique and its extensions). In addition, these values were obtained in different experimental designs. These are designs that involve design decisions (project selection, feature selection, distance measurement, number of analogies, and adaptation rules) and validation techniques (jackknife method, *n*-fold cross validation, etc.). Therefore, it is difficult to define the conditions under which they were obtained. However, we believe that the results obtained using different experimental designs are more robust than those obtained using a single experimental design.

Only ASEE studies are considered in this review. Therefore, the reported performances of ASEE techniques may have been

Table B.21
Selected studies with their quality scores.

Paper ID	Author	Reference	QA1	QA2	QA3	QA4	QA5	QA6	Score
S1	H. Al-Sakran et al.	[62]	1	0.5	0.5	0	1	0	3
S2	L. Angelis et al.	[24]	1	1	1	1	1	1	6
S3	M. Auer et al.	[56]	1	1	1	1	1	1	6
S4	M. Auer et al.	[63]	1	1	1	1	1	1	6
S5	M. Azzeh	[64]	1	0.5	1	1	1	0.5	5
S6	M. Azzeh	[65]	1	1	1	1	1	0	5
S7	M. Azzeh	[6]	1	1	0.5	1	1	1	5.5
S8	M. Azzeh et al.	[28]	1	1	1	1	1	1	6
S9	M. Azzeh et al.	[29]	1	0.5	0.5	1	1	0	4
S10	M. Azzeh et al.	[4]	1	1	1	1	1	0	5
S11	M. Azzeh et al.	[30]	1	1	0.5	1	1	1	5.5
S12	M. Azzeh et al.	[8]	1	1	1	1	1	0	5
S13	R. Bisio et al.	[58]	0.5	0.5	0	1	1	0	3
S14	N.-H. Chiu et al.	[10]	1	1	0.5	1	1	0	4.5
S15	S.-J. Huang et al.	[38]	1	1	0.5	1	1	0.5	5
S16	A. Idri et al.	[31]	1	1	1	0.5	1	0	4.5
S17	A. Idri et al.	[32]	1	1	1	0.5	1	0.5	5
S18	A. Idri et al.	[33]	1	1	1	1	1	0	5
S19	A. Idri et al.	[34]	1	1	1	1	1	1	6
S20	A. Idri et al.	[35]	1	1	0.5	1	1	0	4.5
S21	A. Idri et al.	[36]	1	1	0	0	1	0	3
S22	M. Jørgensen et al.	[66]	1	1	0.5	1	1	0	4.5
S23	G. Kadoda et al.	[67]	1	1	0	1	1	1	5
S24	Y. Kamei et al.	[68]	1	0.5	0.5	1	1	1	5
S25	J.W. Keung	[20]	1	0.5	0.5	1	1	1	5
S26	J.W. Keung	[60]	1	1	0.5	1	1	0.5	5
S27	J.W. Keung et al.	[21]	1	1	0.5	1	1	0.5	5
S28	J.W. Keung et al.	[22]	1	1	1	0.5	1	0.5	5
S29	J.W. Keung et al.	[23]	1	1	1	0.5	1	1	5.5
S30	C. Kirsopp et al.	[69]	1	1	1	1	1	0	5
S31	C. Kirsopp et al.	[70]	1	1	0.5	1	1	0.5	5
S32	E. Kocaguneli et al.	[71]	1	0.5	0.5	1	1	1	5
S33	E. Kocaguneli et al.	[57]	1	0.5	0.5	1	1	1	5
S34	M.V. Kosti et al.	[72]	1	1	0.5	1	1	0	4.5
S35	T.K. Le-Do et al.	[73]	1	1	0.5	1	1	0.5	5
S36	S. Letchmunan et al.	[74]	1	1	0.5	1	1	1	5.5
S37	J. Li et al.	[75]	1	1	0.5	0.5	1	1	5
S38	J. Li et al.	[76]	1	1	0.5	0	1	0	3.5
S39	J. Li et al.	[77]	1	1	1	1	1	1	6
S40	J. Li et al.	[7]	1	1	1	1	1	1	6
S41	Y.F. Li et al.	[78]	1	0.5	0.5	1	1	0	4
S42	Y.F. Li et al.	[79]	1	1	0.5	1	1	0	4.5
S43	Y.F. Li et al.	[5]	1	1	0.5	1	1	0	4.5
S44	Y. F. Li et al.	[9]	1	1	1	1	1	1	6
S45	C. Mair et al.	[80]	1	1	0.5	1	1	1	5.5
S46	E. Mendes et al.	[81]	1	1	0	1	1	1	5
S47	E. Mendes et al.	[82]	1	1	0.5	1	1	0	4.5
S48	E. Mendes et al.	[83]	1	1	0.5	1	1	1	5.5
S49	E. Mendes et al.	[84]	1	1	0	1	1	1	5
S50	E. Mendes et al.	[85]	1	1	0.5	1	1	1	5.5
S51	D. Milios et al.	[39]	1	1	1	1	1	0	5
S52	N. Mittas et al.	[86]	1	1	0.5	1	1	1	5.5
S53	N. Mittas et al.	[87]	1	1	1	1	1	1	6
S54	N. Mittas et al.	[25]	1	1	1	1	1	0.5	5.5
S55	T. Mukhopadhyay et al.	[53]	1	1	1	1	1	1	6
S56	I. Myrtveit et al.	[12]	1	1	1	1	1	1	6
S57	N. Ohsugi et al.	[88]	1	1	0	1	1	1	5
S58	R. Premraj et al.	[37]	1	0.5	0.5	0.5	1	0	3.5
S59	M. Shepperd et al.	[89]	1	1	1	1	1	1	6
S60	I. Stamelos et al.	[26]	1	1	1	1	1	0.5	5.5
S61	I. Stamelos et al.	[27]	1	0.5	0	1	1	1	4.5
S62	A. Tosun et al.	[90]	1	1	0.5	1	1	0	4.5
S63	M. Tsunoda et al.	[91]	1	1	1	1	1	0	5
S64	F. Walkerden et al.	[59]	1	1	1	1	1	1	6
S65	J. Wen et al.	[92]	1	1	0.5	0.5	1	0.5	4.5

overestimated. Furthermore, it is possible that the extracted advantages and limitations of ASEE methods reflect only the authors' opinions. Being aware of this limitation, we listed the supporting studies for each of the extracted advantages and limitations. However, the reader must also be aware of the possible impact of authors' interests and opinions on these findings.

7. Conclusion

This systematic mapping and review summarizes the existing studies with their focus on analogy-based software effort estimation (ASEE). The paper provides a library of ASEE papers classified according to research source, research approach, contribution type,

Table C.22
Classification of the selected studies.

Paper ID	Research approach	Contribution	Techniques used in combination with ASEE methods	Investigated step
S1	DEM	Technique	Multi-agent technology	Steps 1, 2, and 3
S2	DEM + HE	Technique + tool	Statistical method	Steps 2 and 3
S3	DEM + HE	Technique + tool	None	Step 1
S4	DEM + HE	Technique + tool	None	Step 1
S5	DEM + HE	Technique	Bees algorithm	Step 3
S6	DEM + HE	Technique	Model tree	Step 3
S7	HE	Comparison	None	Steps 1 and 3
S8	DEM + HE	Metric	Fuzzy logic	Step 2
S9	DEM + HE	Technique	Fuzzy logic	Step 1
S10	DEM + HE	Technique	Fuzzy logic + grey relational analysis + statistical method	Steps 1 and 2
S11	DEM + HE	Technique	Fuzzy logic + grey relational analysis	Steps 1, 2, and 3
S12	DEM + HE	Technique	Fuzzy logic	Steps 1, 2, and 3
S13	DEM + HE	Technique + tool	None	Steps 1, 2, and 3
S14	DEM + HE	Technique	Genetic algorithm	Step 3
S15	DEM + HE	Technique	Genetic algorithm	Step 1
S16	DEM + HE	Metric	Fuzzy logic	Step 2
S17	DEM + HE	Metric	Fuzzy logic	Step 2
S18	DEM + HE	Technique + tool	Fuzzy logic	Steps 1, 2, and 3
S19	DEM + HE	Technique	Fuzzy logic	Step 3
S20	DEM + HE	Technique	Fuzzy logic + genetic algorithm	Step 1
S21	TH	Comparison	Fuzzy logic	Steps 1, 2, and 3
S22	DEM + EXP + HE	Technique	Statistical method	Step 3
S23	HE	Validation	None	Steps 1 and 3
S24	DEM + HE	Technique	Other	Step 1
S25	HE	Validation	Statistical method	Step 1
S26	DEM + HE	Other	None	_
S27	DEM + HE	Technique	Statistical method	Step 1
S28	DEM + HE	Technique	Statistical method	Step 1
S29	DEM + HE	Technique	Statistical method	Step 1
S30	DEM + HE	Technique	None	Step 3
S31	HE	Comparison	None	Step 1
S32	HE + OT	Other	None	_
S33	DEM + HE	Technique + tool	None	Steps 1 and 3
S34	DEM + HE	Technique	None	Step 3
S35	DEM + HE	Technique	None	Step 1
S36	HE	Validation	None	Step 3
S37	HE	Other	None	Step 1
S38	TH	Model	None	Steps 1, 2, and 3
S39	DEM + HE	Technique	Collaborative filtering + rough set analysis	Step 1
S40	DEM + HE	Technique	Collaborative filtering	Steps 1, 2, and 3
S41	DEM + HE	Technique	None	Step 3
S42	DEM + HE	Technique	Other	Step1
S43	DEM + HE	Technique	Genetic algorithm	Step1
S44	DEM + HE	Technique	Artificial neural network	Step 3
S45	RV	Comparison	None	_
S46	HE	Validation	None	Steps 1 and 3
S47	HE	Comparison	None	Steps 1 and 3
S48	HE	Comparison	None	Steps 2 and 3
S49	HE	Comparison	None	Steps 1, 2, and 3
S50	HE	Comparison	None	Steps 2 and 3
S51	DEM + HE	Technique	Genetic algorithm	Step 1
S52	DEM + HE	Technique	Least squares regression	Steps 1, 2, and 3
S53	DEM + HE	Technique	Least squares regression	Steps 1, 2, and 3
S54	DEM + HE	Technique	Statistical method	Step 3
S55	DEM + HE	Technique + tool	None	Steps 1, 2, and 3
S56	HE + EXP	Comparison	Expert judgment	_
S57	HE	Other	None	Step 3
S58	DEM + HE	Technique	Fuzzy logic	Step 1
S59	DEM + HE	Technique + tool	None	Steps 1, 2, and 3
S60	DEM + HE	Technique	Statistical method	Step 3
S61	HE	Validation	Statistical method	Steps 1, 2, and 3
S62	DEM + HE	Technique	Statistical method	Step 1
S63	DEM + HE	Technique	None	Step 3
S64	DEM + HE + EXP	Technique + tool	Expert judgment	Steps 1, 2 and 3
S65	DEM + HE	Technique	Statistical method	Step 1

techniques used in combination with ASEE methods, and ASEE steps. In addition, this study has investigated ASEE techniques from five perspectives: estimation accuracy, relative prediction accuracy, estimation context, impact of the techniques used in combination with ASEE methods, and ASEE tools. In total, 65 relevant articles were identified in the 1992–2012 period. The main findings of the systematic mapping and review process are the following, in summary form:

What are the approaches most frequently applied in ASEE research, and how has their frequency changed over time? Most ASEE studies apply the history-based evaluation and solution proposal approaches. The number of papers using these two approaches is increasing over time.

What are the main contributions of ASEE studies? The majority of ASEE researchers focus on the development of techniques, in particular, the enhancement of existing techniques, to improve the

Table D.23
Estimation accuracy values of ASEE techniques.

ID	MMRE (%)	MdMRE (%)	Pred(25) (%)	Dataset	ID	MMRE (%)	MdMRE (%)	Pred(25) (%)	Dataset
S2	73.00	_	33.00	Albrecht	S30	63.10	_	_	BT
S2	40.00	_	62.00	Abran-Robillard	S30	41.20	_	_	Desharnais
S3	48.20	_	50.00	Albrecht	S30	71.20	_	_	Finnish
S3	58.20	_	33.40	Kemerer	S34	120.59	53.15	_	Maxwell
S3	30.10[a]	_	49.97[a]	Desharnais	S34	64.80	37.08	_	Desharnais
S5	51.68	_	54.20	Albrecht	S34	54.93	35.04	_	Cocomo-Nasa
S5	40.20	_	46.70	Kemerer	S35	53.86	30.98	42.86	Desharnais
S5	42.70	_	44.20	Desharnais	S35	71.31	47.86	29.03	Maxwell
S5	57.00	_	40.60	COCOMO	S35	86.62	36.28	37.42	ISBSG Telecom
S5	20.00	_	77.70	Nasa93	S39	19.00	_	83.00	Kemerer
S5	38.40	_	46.67	Telecom	S39	59.00	_	42.00	Desharnais
S6	20.10	19.50	62.00	ISBSG	S39	26.00	_	72.00	ISBSG
S6	26.14	12.00	72.70	Desharnais	S40	16.00[b]	_	81.82[b]	ISBSG
S6	21.70	21.90	60.00	COCOMO	S40	14.00[b]	_	83.33[b]	Kemerer
S6	36.50	26.10	46.70	Kemerer	S40	45.00[b]	_	33.33[b]	Leung02
S6	32.30	19.90	58.30	Albrecht	S41	45.00	_	46.00	Albrecht
S6	69.80	18.60	56.50	Maxwell	S41	71.00	_	32.00	Desharnais
S6	34.90	10.90	67.10	China	S41	61.00	_	29.00	Maxwell
S7	33.10[b]	_	_	Albrecht	S42	36.00[b]	33.00[b]	40.00[b]	Desharnais
S7	10.10[b]	_	_	China	S42	28.00[b]	19.00[b]	67.00[b]	Maxwell
S7	58.50[b]	_	_	COCOMO	S43	30.00[b]	27.00[b]	63.00[b]	Albrecht
S7	35.80[b]	_	_	Desharnais	S43	32.00[b]	29.00[b]	44.00[b]	Desharnais
S7	30.80[b]	_	_	Kemerer	S44	41.00[b]	25.00[b]	36.00[b]	Albrecht
S7	46.20[b]	_	_	Maxwell	S44	52.00[b]	32.00[b]	36.00[b]	Desharnais
S7	36.70[b]	_	_	Telecom	S44	80.00[b]	45.00[b]	35.00[b]	Maxwell
S8	13.55[b]	_	84.00[b]	ISBSG	S44	74.00[b]	42.00[b]	30.00[b]	ISBSG
S9	28.70[b]	21.80[b]	54.70[b]	ISBSG	S49	21.40[a,b]	_	71.28[a,b]	Tukutuku
S9	38.50[b]	31.70[b]	42.40[b]	Desharnais	S51	40.67[b]	36.80[b]	38.80[b]	Desharnais
S10	11.30	7.20	91.10	Desharnais	S51	23.00	23.40	59.40	ISBSG
S10	19.90	13.90	70.00	COCOMO	S52	19.71	9.09	71.43	Abran-Robillard
S11	33.30	22.00	55.20	ISBSG	S52	40.17	34.00	43.14	ISBSG
S11	30.60	17.50	64.70	Desharnais	S53	177.79	57.98	22.73	ISBSG
S11	23.20	14.80	66.70	COCOMO	S53	54.36	25.98	47.31	NASA93
S11	36.20	33.20	52.90	Kemerer	S54	36.59[b]	14.23[b]	71.43[b]	Abran
S11	51.10	48.00	28.60	Albrecht	S54	68.50[b]	48.77[b]	28.57[b]	Finnish
S12	28.55	17.80	59.80	ISBSG	S54	49.38[b]	29.58[b]	42.86[b]	COCOMO
S12	33.37	20.36	62.33	COCOMO	S55	52.79	_	_	Kemerer
S12	26.89	19.32	64.94	Desharnais	S56	136.00	51.00	_	COTS
S12	50.08	30.75	50.00	Albrecht	S59	62.00	_	33.00	Albrecht
S12	55.65	24.24	53.33	Kemerer	S59	39.00	_	38.00	Atkinson
S13	_	_	61.00[b]	COCOMO	S59	64.00	_	36.00	Desharnais
S14	43.00[b]	20.00[b]	61.00[b]	Albrecht	S59	41.00	_	39.00	Finnish
S14	52.00[b]	36.00[b]	43.00[b]	Abran-Robillard	S59	62.00	_	40.00	Kemerer
S15	69.00[b]	53.00[b]	30.00[b]	ISBSG	S59	78.00	_	21.00	Mermaid
S15	32.00[b]	24.00[b]	70.00[b]	Albrecht	S59	74.00	_	23.00	Real-time1
S18	18.38[a]	_	89.41[a]	COCOMO	S59	39.00	_	44.00	Telecom1
S20	58.60	_	84.91	Tukutuku	S59	37.00	_	51.00	Telecom2
S22	31.00	26.00	_	Jeffery & Stathis	S61	23.84[b]	_	70.37[b]	ISBSG
S22	39.00	31.00	_	Jørgensen97	S63	119.10	54.00	_	ISBSG
S23	47.60	_	_	Desharnais	S63	84.40	36.30	_	Kitchenham
S24	45.07[a]	_	44.43[a]	Desharnais	S63	48.60	31.10	_	Desharnais
S25	100.60	_	_	Albrecht	S64	55.00	_	24.00	Australian
S25	60.30	_	_	Telecom	S65	151.00	_	21.00	COCOMO
S25	68.10	_	_	Kemerer	S65	62.00	_	43.00	Desharnais
S26	66.60	_	_	Desharnais	S65	26.00	_	67.00	NASA
S27	33.67[a]	_	49.50[a]	Desharnais					

[a] Mean of accuracy values.
[b] Accuracy of the optimal configuration.

prediction accuracy of ASEE techniques and to overcome the limitations of existing ASEE approaches.

What are the techniques reportedly used most frequently in combination with analogy? Statistical methods and fuzzy logic are the techniques most frequently used in combination with analogy, followed by genetic algorithms.

Have the various steps of the analogy procedure received the same amount of attention from researchers? Feature and case subset selection (FCSS) is the step that has been investigated the most, followed by adaptation, and, finally, similarity evaluation.

What is the overall estimation accuracy of ASEE techniques? In general, ASEE methods tend to yield acceptable estimates. Specifically, the mean of the prediction accuracy values is 49.8% for MMRE, 29.37% for MdMRE, and 51.23% for Pred(25).

Do ASEE techniques perform better than the other estimation models (both ML and non ML)? The overall picture suggests that ASEE techniques outperform the other prediction models. This conclusion is supported by most of the selected papers.

What are favorable estimation contexts for ASEE techniques? Several studies suggest that ASEE techniques can model the complex relationships between effort and software attributes. Furthermore, they can be applied at an early stage of a software project and can mitigate problems with outliers. In contrast, classical ASEE techniques cannot handle categorical attributes or missing values.

Table D.24
Comparison of MMRE, MdMRE, and Pred(25) using ASEE techniques and non-ML models ("+" indicates that an ASEE model outperforms a non-ML model, "−" indicates that a non-ML model outperforms an ASEE technique, the number between brackets indicates the difference between an ASEE technique and a non-ML model, using MMRE, MdMRE, or Pred(25)).

Criterion	Regression	COCOMO	Expert	FP
MMRE+	S1(+6) Albrecht, S11(+9.3) Desharnais, S11(+107) COCOMO, S11(+18.1) Kemerer, S11(+8.2) Albrecht, S12(+20.2) ISBSG, S12(+63.23) COCOMO, S12(+7.71) Desharnais, S12(+11.16) Albrecht, S12(+106.08) Kemerer, S14(+29) Albrecht, S14(+36) Abran, S15(+121) ISBSG, S15(+38) Albrecht, S41(+46) Albrecht, S41(+16) Desharnais, S41(+22) Maxwell, S42(+26) Desharnais, S42(+6) Maxwell, S44(+53) Albrecht, S44(+21) Desharnais, S44(+29) Maxwell, S44(+8) ISBSG, S52(+12.26) Abran, S52(+9.22) ISBSG, S53(+146.46) ISBSG, S53(+13.52) NASA93, S59(+28) Albrecht, S59(+6) Atkinson, S59(+2) Desharnais, S59(60) Finnish, S59(+45) Kemerer, S59(+174) Mermaid, S59(+47) Telecom1, S59(+105) Telecom2, S64(+13) Australian	S19(+25.54) COCOMO, S55(+566.2) Kemerer	S56(+107) COTS	S55(+49.95) Kemerer
MMRE−	S1(−18) Abran, S11(−0.1) ISBSG, S42(−9) Maxwell, S56(−9) COTS	S40(−12.1) Leung02	S56(−22.07) Kemerer	N
MdMRE+	S11(+4.5) ISBSG, S11(+20.7) Desharnais, S11(+44.1) COCOMO, S11(+6.5) Kemerer, S11(+9.1) Albrecht, S12(+20.49) ISBSG, S12(+62.04) COCOMO, S12(9.28) Desharnais, S12(+1.55) Albrecht, S12(+50.64) Kemerer, S14(+21) Albrecht, S14(+5) Abran, S15(+36) ISBSG, S15(+21) Albrecht, S44(+30) Albrecht, S44(+2) Desharnais, S44(+31) Maxwell, S44(+18) ISBSG, S52(+11.61) Abran, S52(+8.62) ISBSG, S53(+28.78) ISBSG, S53(+10.8) NASA93	N	S56(+8) COTS	N
MdMRE−	S56(−16) COTS	N	N	N
Pred+	S1(+8) Albrecht, S11(+6.6) ISBSG, S11(+22.7) Desharnais, S11(+41.7) COCOMO, S11(+6.2) Kemerer, S11(+27.8) Albrecht, S12(+23) ISBSG, S12(+39.23) COCOMO, S12(+19.44) Desharnais, S12(+12.5) Albrecht, S12(+46.63) Kemerer, S14(+28) Albrecht, S14(+10) Abran, S15(+18) ISBSG, S15(+46) Albrecht, S41(+17) Albrecht, S41(+10) Desharnais, S41(+6) Maxwell, S44(+19) Albrecht, S44(+1) Desharnais, S44(+12) Maxwell, S44(+11) ISBSG, S52(+14.29) Abran, S52(+11.77) ISBSG, S53(+18.18) ISBSG, S53(+15.05) NASA93, S59(+18) Finnish, S59(+27) Kemerer, S59(+7) Mermaid, S59(+24) Telecom2, S64(+8) Australian	S18(+39.32) COCOMO	N	N
Pred−	S1(−9.4) Abran, S59(−5) Atkinson, S59(−6) Desharnais	N	N	N

Table D.25
Comparison of MMRE, MdMRE, and Pred(25) using ASEE techniques and ML models ("+" indicates that an ASEE technique outperforms an ML model, "−" indicates that an ML model outperforms an ASEE technique, the number between brackets indicates the difference between an ASEE technique and an ML model, using MMRE, MdMRE, or Pred(25)).

Criterion	ANN	DT	SVR	RBF	BN	GP	AR
MMRE+	S11(+36.2) ISBSG, S11(+30.6) Desharnais, S11(+32.3) COCOMO, S11(+11.7) Kemerer, S11(+28.5) Albrecht, S14(+47) Albrecht, S14(+18) Abran, S15(+101) ISBSG, S15(+72) Albrecht, S42(+11) Desharnais, S43(+19) Albrecht, S43(+25) Desharnais, S44(+44) Albrecht, S44(+15) Desharnais, S44(+52) Maxwell, S44(+22) ISBSG	S14(+34) Albrecht, S14(+37) Abran, S15(+120) ISBSG, S15(+35) Albrecht, S42(+54) Desharnais, S43(+140) Albrecht, S43(+20) Desharnais, S44(+103) Albrecht, S44(+19) Desharnais, S44(+72) Maxwell, S44(+33) ISBSG	S43(+15) Albrecht, S43(+8) Desharnais	S43(+19) Albrecht, S43(+10) Desharnais	N	N	N
MMRE−	N	N	N	N	N	N	N
MdMRE+	S11(+7.5) ISBSG, S11(+24.6) Desharnais, S11(+27.4) COCOMO, S11(+4.4) Kemerer, S11(+14.6) Albrecht, S14(+41) Albrecht, S15(+41) ISBSG, S15(+27) Albrecht, S43(+24) Albrecht, S43(+14) Desharnais, S44(+14) Albrecht, S44(+6) Desharnais, S44(+17) Maxwell, S44(+18) ISBSG	S14(+30) Albrecht, S14(+7) Abran, S15(+16) Albrecht, S43(+62) Albrecht, S43(+6) Desharnais, S44(+41) Albrecht, S44(+12) Desharnais, S44(+20) Maxwell, S44(+19) ISBSG	S43(+16) Albrecht, S43(+8) Desharnais	S43(+12) Albrecht	N	N	N
MdMRE−	N	S15(−1) ISBSG	N	N	N	N	N
Pred+	S11(+10.3) ISBSG, S11(+20.7) Desharnais, S11(+16.7) COCOMO, S11(+2.9) Kemerer, S11(+23.6) Albrecht, S14(+39) Albrecht, S14(+33) Abran, S15(+18) ISBSG, S15(+53) Albrecht, S43(+38) Albrecht, S43(+22) Desharnais, S44(+3) Albrecht, S44(+5) Desharnais, S44(+22) Maxwell, S44(+5) ISBSG	S14(+35) Albrecht, S14(+14) Abran, S15(+9) ISBSG, S15(+39) Albrecht, S43(+50) Albrecht, S43(+14) Desharnais, S44(+19) Albrecht, S44(+11) Desharnais, S44(+9) Maxwell, S44(+12) ISBSG	S43(+38) Albrecht, S43(+7) Desharnais	S43(+38) Albrecht, S43(+7) Desharnais	N	N	N
Pred−	N	N	N	N	N	N	N

Several techniques extending the traditional ASEE technique have been proposed to overcome these limitations.

What is the impact on estimation accuracy of combining analogy with another technique? The overall results suggest that estimation accuracy is improved when analogy is used in combination with another technique to generate estimates. Fuzzy logic, genetic algorithms, the model tree, and the collaborative filtering are the techniques that improve the performance of ASEE techniques the most.

What are the ASEE tools most frequently used to generate estimates? ANGEL, developed by Shepperd et al., is the tool most frequently used to predict effort based on ASEE techniques.

Appendix A. Description of classification criteria

See Tables A.18 and A.19.

Appendix B. List of selected studies

See Tables B.20 and B.21.

Appendix C. Classification results

See Table C.22.

Table D.26
Accuracy improvement in terms of MMRE, MdMRE, and Pred(25), using each technique in combination with analogy.

Techniques used in combination with ASEE methods	Paper Id	Dataset	MMRE improvement (%)	MdMRE improvement (%)	Pred(25) improvement (%)
ANN	S44	Albrecht	52.87	41.86	9.09
		Desharnais	13.33	23.81	5.88
		Maxwell	23.08	27.42	66.67
		ISBSG	24.49	28.81	36.36
BA	S5	Albrecht	27.21	N	85.62
		Kemerer	28.09	N	16.75
		Desharnais	28.95	N	41.67
		COCOMO	63.72	N	219.69
		Nasa93	75.37	N	133.33
		Telecom	36.00	N	40.02
CF	S40	Kem87	77.42	N	108.33
		Leung02	−35.54	N	N
CF + RSA	S39	Kem83	69.35	N	107.50
		Desharnais	7.81	N	16.67
		ISBSG	51.85	N	60.00
EJ	S56	COTS	11.69	1.92	N
FL	S8	ISBSG	77.19	N	89.19
	S9	ISBSG	2.38	−5.72	0.00
		Desharnais	23.00	12.19	12.47
	S18	COCOMO	N	N	181.61
	S12	ISBSG	45.43	41.12	40.01
		COCOMO	29.45	39.76	78.09
		Desharnais	29.61	37.27	51.38
		Albrecht	21.13	20.95	50.15
		Kemerer	12.77	27.27	33.33
FL + GRA	S11	ISBSG	37.17	38.89	34.31
		Desharnais	19.90	43.18	50.82
		COCOMO	20.00	40.80	29.09
		Kemerer	39.26	18.83	32.25
		Albrecht	20.16	−23.39	−14.11
	S10	Desharnais	70.42	76.16	112.55
		COCOMO	31.38	44.40	35.40
GA	S14	Albrecht	27.12	45.95	56.41
		Abran-Robillard	58.40	37.93	126.32
	S15	ISBSG	36.11	19.70	400.00
		Albrecht	30.43	25.00	84.21
	S43	Albrecht	38.78	44.90	384.62
		Desharnais	48.39	42.00	100.00
	S51	Desharnais	44.30	24.12	56.45
LSR	S52	Abran-Robillard	44.59	59.74	36.37
		ISBSG	11.81	15.13	57.16
	S53	ISBSG	65.87	17.47	38.94
		NASA93	37.47	39.38	33.34
MT	S6	ISBSG	72.95	53.01	74.16
		Desharnais	60.33	73.74	210.68
		COCOMO	67.32	56.02	89.27
		Kemerer	32.78	−3.98	0.00
		Albrecht	59.42	67.75	249.10
		Maxwell	47.79	75.42	409.01
		China	41.44	72.14	129.01
SM (Mantel correlation)	S27	Desharnais	6.03	N	15.38
SM (Principal Components Analysis + Pearson correlation coefficients)	S65	COCOMO	27.05	N	23.53
		Desharnais	4.62	N	10.26
		NASA	35.00	N	19.64
SM (Regression toward the mean)	S22	Jeffery & Stathis	20.51	27.78	N
		Jørgensen97	11.36	8.82	N

Appendix D. Review results

See Tables D.23–D.26.

References

[1] J. Wen, S. Li, Z. Lin, Y. Huc, C. Huang, Systematic literature review of machine learning based software development effort estimation models, Inf. Softw. Technol. 54 (1) (2012) 41–59.

[2] M. Jørgensen, M. Shepperd, A systematic review of software development cost estimation studies, IEEE Trans. Softw. Eng. 33 (1) (2007) 33–53.

[3] M. Shepperd, C. Schofield, Estimating software project effort using analogies, IEEE Trans. Softw. Eng. 23 (11) (1997) 736–743.

[4] M. Azzeh, D. Neagu, P. Cowling, Software effort estimation based on weighted fuzzy grey relational analysis, in: Proceedings of the 5th International Conference on Predictor Models in Software Engineering, Vancouver, British Columbia, Canada, 2009, pp. 1–10.

[5] Y.F. Li, M. Xie, T.N. Goh, A study of project selection and feature weighting for analogy-based software cost estimation, J. Syst. Softw. 82 (2) (2009) 241–252.

[6] M. Azzeh, A replicated assessment and comparison of adaptation techniques for analogy-based effort estimation, Empir. Softw. Eng. 17 (1–2) (2012) 90–127.

[7] J. Li, G. Ruhe, A. Al-Emran, M. Richter, A flexible method for software effort estimation by analogy, Empir. Softw. Eng. 12 (1) (2007) 65–106.

[8] M. Azzeh, D. Neagu, P. Cowling, Analogy-based software effort estimation using Fuzzy numbers, J. Syst. Softw. 84 (2) (2011) 270–284.

[9] Y.F. Li, M. Xie, T.N. Goh, A study of the non-linear adjustment for analogy based software cost estimation, Empir. Softw. Eng. 14 (6) (2009) 603–643.

[10] N.-H. Chiu, S.-J. Huang, The adjusted analogy-based software effort estimation based on similarity distances, J. Syst. Softw. 80 (4) (2007) 628–640.

[11] L.C. Briand, T. Langley, I. Wieczorek, A replicated assessment and comparison of common software cost modeling techniques, in: Proceedings of the 22nd

International Conference on Software Engineering, Limerick, Ireland, 2000, pp. 377–386.

[12] I. Myrtveit, E. Stensrud, A controlled experiment to assess the benefits of estimating with analogy and regression models, IEEE Trans. Softw. Eng. 25 (4) (1999) 510–525.

[13] B. Kitchenham, S. Charters, Guidelines for Performing Systematic Literature Reviews in Software Engineering, Tech. Rep. EBSE-2007-01, Keele University and University of Durham, 2007.

[14] B. Kitchenham, D. Budgen, O.P. Brereton, The value of mapping studies – a participant–observer case study, in: Proceedings of the 14th International Conference on Evaluation and Assessment in Software Engineering, Keele University, UK, 2010, pp. 1–9.

[15] B. Kitchenham, E. Mendes, G.H. Travassos, A systematic review of cross vs. within company cost estimation studies, in: Proceedings of the Empirical Assessment in Software Engineering (EASE) Conference, 2006, pp. 89–98.

[16] J.P. Higgins, S. Green, Cochrane Handbook for Systematic Reviews of Interventions, Version 5.0.2, The Cochrane Collaboration, 2009. <www.cochrane-handbook.org> (updated September 2009).

[17] Computer Science Conference Rankings CORE, 2011. <http://lamp.infosys.deakin.edu.au/era/?page=cforse110,2011>.

[18] A. Fernandez, E. Insfran, S. Abrahão, Usability evaluation methods for the Web: a systematic mapping study, Inf. Softw. Technol. 53 (8) (2011) 789–817.

[19] R.J. Light, D.B. Pillemer, Summing Up: The Science of Reviewing Research, Harvard University Press, Cambridge, MA, USA, 1984.

[20] J.W. Keung, Empirical evaluation of analogy-X for software cost estimation, in: Proceedings of the Second ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, Kaiserslautern, Germany, 2008, pp. 294–296.

[21] J.W. Keung, B. Kitchenham, Optimising project feature weights for analogy-based software cost estimation using the Mantel correlation, in: Proceedings of the 14th Asia–Pacific Software Engineering Conference, Aichi, Japan, 2007, pp. 222–229.

[22] J.W. Keung, B. Kitchenham, Experiments with analogy-X for Software cost estimation, in: Proceedings of the 19th Australian Conference on Software Engineering, 2008, pp. 229–238.

[23] J.W. Keung, B. Kitchenham, D.R. Jeffery, Analogy-X: providing statistical inference to analogy-based software cost estimation, IEEE Trans. Softw. Eng. 34 (4) (2008) 471–484.

[24] L. Angelis, I. Stamelos, A simulation tool for efficient analogy based cost estimation, Empir. Softw. Eng. 5 (1) (2000) 35–68.

[25] N. Mittas, M. Athanasiades, L. Angelis, Improving analogy-based software cost estimation by a resampling method, Inf. Softw. Technol. 50 (3) (2008) 221–230.

[26] I. Stamelos, L. Angelis, Managing uncertainty in project portfolio cost estimation, Inf. Softw. Technol. 43 (13) (2001) 759–768.

[27] I. Stamelos, L. Angelis, M. Morisio, E. Sakellarisc, G.L. Bleris, Estimating the development cost of custom software, Inform. Manage. 40 (8) (2003) 729–741.

[28] M. Azzeh, D. Neagu, P. Cowling, Software project similarity measurement based on fuzzy C-means, in: Proceedings of the International Conference on Software Process, Leipzig, Germany, 2008, pp. 123–134.

[29] M. Azzeh, D. Neagu, P. Cowling, Improving analogy software effort estimation using fuzzy feature subset selection algorithm, in: Proceedings of the 4th International Workshop on Predictor Models in Software Engineering, Leipzig, Germany, 2008, pp. 71–78.

[30] M. Azzeh, D. Neagu, P. Cowling, Fuzzy grey relational analysis for software effort estimation, Empir. Softw. Eng. 15 (1) (2010) 60–90.

[31] A. Idri, A. Abran, A fuzzy logic based set of measures for software project similarity validation and possible improvement, in: Proceedings of the 7th International Symposium on Software Metrics, London, UK, 2001, pp. 85–96.

[32] A. Idri, A. Abran, Evaluating software project similarity by using linguistic quantifier guided aggregations, in: Proceedings of the Joint 9th IFSA World Congress and 20th NAFIPS International Conference, Vancouver, British Columbia, Canada, 2001, pp. 470–475.

[33] A. Idri, A. Abran, T.M. Khoshgoftaar, Estimating software project effort by analogy based on linguistic values, in: Proceedings of the Eighth IEEE Symposium on Software Metrics, 2002, pp. 21–30.

[34] A. Idri, T.M. Khosgoftaar, A. Abran, Investigating soft computing in case-based reasoning for software cost estimation, Eng. Intell. Syst. 10 (3) (2002) 147–157.

[35] A. Idri, A. Zahi, E. Mendes, A. Abran, Software cost estimation by fuzzy analogy for Web hypermedia applications, in: Proceedings of the International Conference on Software Process and Product Measurement, Cadiz, Spain, 2006, pp. 53–62.

[36] A. Idri, A. Zakrani, A. Abran, Functional equivalence between radial basis function neural networks and fuzzy analogy in software cost estimation, in: Proceedings of the 3rd IEEE international Conference in Information and Communication Technologies: From Theory to Application, Damas, Syria, 2008, pp. 1–5.

[37] R. Premraj, M. Shepperd, M. Cartwright, Meta-data to guide retrieval in CBR for software cost prediction, in: Proceedings of the 8th UK Workshop on Case-based Reasoning, 2003, pp. 26–37.

[38] S.-J. Huang, N.-H. Chiu, Optimization of analogy weights by genetic algorithm for software effort estimation, Inf. Softw. Technol. 48 (11) (2006) 1034–1045.

[39] D. Milios, I. Stamelos, C. Chatzibagias, Global optimization of analogy-based software cost estimation with genetic algorithms, in: Proceedings of EANN/AIAI (2), 2011, pp. 350–359.

[40] J.M. Desharnais, Analyse statistique de la productivité des projets de développement en informatique à partir de la technique des points de fonction, Master's Thesis, University of Montreal, 1989.

[41] International Software Benchmarking Standards Group (ISBSG). <http://www.isbsg.org>.

[42] A.J. Albrecht, J.E. Gaffney, Software function, source lines of code, and development effort prediction: a software science validation, IEEE Trans. Softw. Eng. 9 (6) (1983) 639–648.

[43] B.W. Boehm, Software Engineering Economics, Prentice Hall PTR, New Jersey, 1981.

[44] C.F. Kemerer, An empirical validation of software cost estimation models, Commun. ACM 30 (5) (1987) 416–429.

[45] K.D. Maxwell, Applied Statistics for Software Managers, Prentice-Hall, Upper Saddle River, 2002.

[46] A. Abran, P.N. Robillard, Function point analysis: an empirical study of its measurement processes, IEEE Trans. Softw. Eng. 22 (12) (1996) 895–910.

[47] G. Boetticher, T. Menzies, T. Ostrand, PROMISE Repository of Empirical Software Engineering Data Repository, West Virginia University, Department of Computer Science. <http://promisedata.org/>.

[48] L.C. Briand, K. El Emam, D. Surmann, I. Wieczorek, K.D. Maxwell, An assessment and comparison of common software cost estimation modeling techniques, in: Proceedings of the 21st International Conference on Software Engineering, Los Angeles, California, 1999, pp. 313–323.

[49] M. Shepperd, G. Kadoda, Comparing software prediction techniques using simulation, IEEE Trans. Softw. Eng. 27 (11) (2001) 1014–1022.

[50] S.G. MacDonell, M.J. Shepperd, Combining techniques to optimize effort predictions in software project management, J. Syst. Softw. 66 (2) (2003) 91–98.

[51] M. Jørgensen, Forecasting of software development work effort: evidence on expert judgment and formal models, Int. J. Forecast. 23 (3) (2007) 449–462.

[52] M. Shepperd, C. Schofield, B. Kitchenham, Effort estimation using analogy, in: Proceedings of the 18th International Conference on Software Engineering, Berlin, 1996, pp. 170–178.

[53] T. Mukhopadhyay, S. Steven, M. Vicinanza, J. Prietula, Examining the feasibility of a case-based reasoning model for software effort estimation, MIS Quart. 16 (2) (1992) 155–171.

[54] S. Schulz, CBR-Works, in: Proceedings of the 7th German Workshop on Case-based Reasoning, Heidelberg, German, 1999, pp. 3–5.

[55] I. Stamelos, L. Angelis, E. Sakellaris, BRACE: bootstrap-based analogy cost estimation, in: Proceedings of the 12th European Software Control Metrics, 2001, pp. 17–23.

[56] M. Auer, S. Biffl, Increasing the accuracy and reliability of analogy-based cost estimation with extensive project feature dimension weighting, in: Proceedings of the 2004 International Symposium on Empirical Software Engineering, 2004, pp. 147–155.

[57] E. Kocaguneli, T. Menzies, A. Bener, J.W. Keung, Exploiting the essential assumptions of analogy-based effort estimation, IEEE Trans. Softw. Eng. 38 (2) (2012) 425–438.

[58] R. Bisio, F. Malabocchia, Cost estimation of software projects through case-based reasoning, in: Proceedings of the First International Conference on Case-Based Reasoning Research and Development, 1995, pp. 11–22.

[59] F. Walkerden, R. Jeffery, An empirical study of analogy-based software effort estimation, Empir. Softw. Eng. 4 (2) (1999) 135–158.

[60] J.W. Keung, Theoretical maximum prediction accuracy for analogy-based software cost estimation, in: Proceedings of the 15th Asia–Pacific Software Engineering Conference, Beijing, China, 2008, pp. 495–502.

[61] T. Foss, E. Stensrud, B. Kitchenham, I. Myrtveit, A simulation study of the model evaluation criterion MMRE, IEEE Trans. Softw. Eng. 29 (11) (2003) 985–995.

[62] H. Al-Sakran, Software cost estimation model based on integration of multi-agent and case-based reasoning, J. Comput. Sci. 2 (3) (2006) 276–282.

[63] M. Auer, A. Trendowicz, B. Graser, E. Haunschmid, S. Biffl, Optimal project feature weights in analogy-based cost estimation: improvement and limitations, IEEE Trans. Softw. Eng. 32 (2) (2006) 83–92.

[64] M. Azzeh, Adjusted case-based software effort estimation using bees optimization algorithm, in: Proceedings of the 15th International Conference on Knowledge-based and Intelligent Information and Engineering Systems, 2011, pp. 315–324.

[65] M. Azzeh, Model tree based adaption strategy for software effort estimation by analogy, in: Proceedings of the 2011 IEEE 11th International Conference on Computer and Information Technology, 2011, pp. 328–335.

[66] M. Jørgensen, U. Indahl, D. Sjøb, Software effort estimation by analogy and "regression toward the mean", J. Syst. Softw. 68 (3) (2003) 253–262.

[67] G. Kadoda, M. Cartwright, L. Chen, M. Shepperd, Experiences using case-based reasoning to predict software project effort, in: Proceedings of the Conference on Evaluation and Assessment in Software Engineering, Keele University, UK, 2000, pp. 23–28.

[68] Y. Kamei, J.W. Keung, A. Monden, K.-I. Matsumoto, An over-sampling method for analogy-based software effort estimation, in: Proceedings of the Second ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, Kaiserslautern, Germany, 2008, pp. 312–314.

[69] C. Kirsopp, E. Mendes, R. Premraj, M. Shepperd, An empirical analysis of linear adaptation techniques for case-based prediction, in: Proceedings of the 5th International Conference on Case-based Reasoning: Research and Development, 2003, pp. 231–245.

[70] C. Kirsopp, M. Shepperd, J. Hart, Search heuristics, case-based reasoning and software project effort prediction, in: Proceedings of the Genetic and Evolutionary Computation Conference, 2002, pp. 1367–1374.

[71] E. Kocaguneli, T. Menzies, How to find relevant data for effort estimation, in: Proceedings of the 2011 International Symposium on Empirical Software Engineering and Measurement, Banff, Canada, 2011, pp. 255–264.

[72] M.V. Kosti, N. Mittas, L. Angelis, DD-EbA: an algorithm for determining the number of neighbors in cost estimation by analogy using distance distributions, in: Proceedings of the 3D Artificial Intelligence Techniques in Software Engineering Workshop, Larnaca, Cyprus, 2010.

[73] T.K. Le-Do, K.-A. Yoon, Y.-S. Seo, D.-H. Bae, Filtering of inconsistent software project data for analogy-based effort estimation, in: Proceedings of the 2010 IEEE 34th Annual Computer Software and Applications Conference, Seoul, Korea, 2010, pp. 503–508.

[74] S. Letchmunan, M. Roper, M. Wood, Investigating effort prediction of web-based applications using CBR on the ISBSG dataset, in: Proceedings of the 14th international conference on Evaluation and Assessment in Software Engineering, 2010, pp. 15–24.

[75] J. Li, A. Al-Emran, G. Ruhe, Impact analysis of missing values on the prediction accuracy of analogy-based software effort estimation method AQUA, in: Proceedings of the First International Symposium on Empirical Software Engineering and Measurement, Madrid, Spain, 2007, pp. 126–135.

[76] J. Li, G. Ruhe, Decision support analysis for software effort estimation by analogy, in: Proceedings of the 3d International Workshop on Predictor Models in Software Engineering, 2007.

[77] J. Li, G. Ruhe, Analysis of attribute weighting heuristics for analogy-based software effort estimation method AQUA+, Empir. Softw. Eng. 13 (1) (2008) 63–96.

[78] Y.F. Li, M. Xie, T.N. Goh, A study of analogy-based sampling for interval based cost estimation for software project management, in: Proceedings of the 4th IEEE International Conference on Management of Innovation and Technology, Bangkok, Thailand, 2008, pp. 281–286.

[79] Y.F. Li, M. Xie, T.N. Goh, A study of mutual information-based feature selection for case-based reasoning in software cost estimation, Expert Syst. Appl. 36 (3) (2009) 5921–5931.

[80] C. Mair, M. Shepperd, The consistency of empirical comparisons of regression and analogy-based software project cost prediction, in: Proceedings of the 4th International Symposium on Empirical Software Engineering, Noosa Heads, Australia, 2005, pp. 509–518.

[81] E. Mendes, S. Counsell, N. Mosley, Towards the prediction of development effort for hypermedia applications, in: Proceedings of the 12th ACM conference on Hypertext and Hypermedia, 2000, pp. 249–258.

[82] E. Mendes, S. Counsell, N. Mosley, Measurement and effort prediction for Web applications, in: Proceeding of Web Engineering, Software Engineering and Web Application Development, 2001, pp. 295–310.

[83] E. Mendes, N. Mosley, Further investigation into the use of CBR and stepwise regression to predict development effort for web hypermedia applications, in: Proceedings of the 2002 International Symposium on Empirical Software Engineering, 2002, pp. 79–90.

[84] E. Mendes, N. Mosley, S. Counsell, A replicated assessment of the use of adaptation rules to improve Web cost estimation, in: Proceedings of the 2003 International Symposium on Empirical Software Engineering, 2003, pp. 100–109.

[85] E. Mendes, I. Watson, C. Triggs, N. Mosley, S. Counsell, A comparative study of cost estimation models for Web hypermedia applications, Empir. Softw. Eng. 8 (2) (2003) 163–196.

[86] N. Mittas, L. Angelis, Combining regression and estimation by analogy in a semi-parametric model for software cost estimation, in: Proceedings of the First Proceedings of the Second ACM-IEEE International Symposium on Empirical Software Engineering and Measurement, Kaiserslautern, Germany, 2008, pp. 70–79.

[87] N. Mittas, L. Angelis, LSEbA: least squares regression and estimation by analogy in a semi-parametric model for software cost estimation, Empir. Softw. Eng. 15 (5) (2010) 523–555.

[88] N. Ohsugi, A. Monden, N. Kikuchi, M.D. Barker, Is this cost estimate reliable? – The relationship between homogeneity of analogues and estimation reliability, in: Proceedings of the First International Symposium on Empirical Software Engineering and Measurement, Madrid, Spain, 2007, pp. 384–392.

[89] M. Shepperd, C. Schofield, Estimating software project effort using analogies, IEEE Trans. Software Eng. 23 (12) (1997) 736–743.

[90] A. Tosun, B. Turhan, A.B. Bener, Feature weighting heuristics for analogy-based effort estimation models, Expert Syst. Appl. 36 (7) (2009) 10325–10333.

[91] M. Tsunoda, A. Monden, T. Kakimoto, K. Matsumoto, An empirical evaluation of outlier deletion methods for analogy-based cost estimation, Proceedings of the 7th International Conference on Predictive Models in Software Engineering, Banff, Canada, 2011.

[92] J. Wen, S. Li, L. Tang, Improve analogy-based software effort estimation using principal components analysis and correlation weighting, in: Proceedings of the 16th Asia–Pacific Software Engineering Conference, Penang, Malaysia, 2009, pp. 179–186.