This is the accepted version of a paper presented at *International Conference on Global Software Engineering*.

Citation for the original published paper:


N.B. When citing this work, cite the original published paper.

Permanent link to this version: http://urn.kb.se/resolve?urn=urn:nbn:se:bth-11144
Effort Estimation in Global Software Development: A Systematic Literature Review

Ricardo Britto  
Software Engineering  
Department  
Blekinge Institute of Technology  
Sweden  
ricardo.britto@bth.se

Vitor Freitas  
Software Engineering  
Federal University of Juiz de Fora  
Brazil  
vfreitas@ice.ufjf.br

Emilia Mendes  
Software Engineering  
Department  
Blekinge Institute of Technology  
Sweden  
emilia.mendes@bth.se

Muhammad Usman  
Software Engineering  
Department  
Blekinge Institute of Technology  
Sweden  
muhammad.usman@bth.se

Abstract—Nowadays, software systems are a key factor in the success of many organizations as in most cases they play a central role helping them attain a competitive advantage. However, despite their importance, software systems may be quite costly to develop, so substantially decreasing companies’ profits. In order to tackle this challenge, many organizations look for ways to decrease costs and increase profits by applying new software development approaches, like Global Software Development (GSD). Some aspects of the software project like communication, cooperation and coordination are more challenging in globally distributed than in co-located projects, since language, cultural and time zone differences are factors which can increase the required effort to globally perform a software project. Communication, cooperation and coordination aspects affect directly the effort estimation of a project, which is one of the critical tasks related to the management of a software development project. There are many studies related to effort estimation methods/techniques for co-located projects. However, there are evidences that the co-located approaches do not fit the GSD. So, this paper presents the results of a systematic literature review of effort estimation in the context of GSD, which aimed at help both researchers and practitioners to have a holistic view about the current state of the art regarding effort estimation in the context of GSD. The results suggest that there is room to improve the current state of the art on effort estimation in GSD.

I. INTRODUCTION

Many software development companies have tried to increase profits by improving the time-to-market of their products, reducing costs by hiring people from countries with cheaper work-hours and defying the “clock” by running the projects during 24 hours. As a result, we can see a great number of software development projects performed globally, distributed in many different sites and normally located in different countries. This distribute setting of managing a software project is called Global Software Development (GSD).

The advantages related to GSD can only be achieved if projects are well managed. There are many challenges related to managing GSD projects. We could remark the following key challenges [1], [2]:

- **Communication** Different languages used by each site of a GSD project hinder the execution of the work.
- **Trust** Cultural variation complicates the process of trusting between each site.

- **Coordination over distance** - Different time zones makes harder the process of hand off between sites.

Those challenges affect directly the effort estimation of a project, which is one of the critical tasks related to the management of a software development project [3].

There are many studies related to effort estimation methods/techniques for co-located projects [4]–[6]. Nevertheless, existing evidence suggests that effort estimation methods/techniques and cost drivers used in the context of co-located projects may not be readily applicable to global software development projects [7]–[10].

There are many systematic literature reviews (SLR) regarding GSD. Kroll et al. [11] identified areas within Software Engineering have been investigated regarding GSD. Schneider et al. [12] performed a SLR to categorize solutions in the GSD context by using process areas. Jalali et al. [13] and Hossain et al. [14] performed systematic literature reviews looking at the use of agile methods in GSD. Da Silva et al. performed two SLRs on project management in the context of GSD [15], [16]. Prikladnitski et al. [17] identified papers that describe process models for GSD in the context of overseas outsourcing. Mishra et al. developed a SLR to find the current research and practice of quality management in GSD projects [18]. Smite et al. [19] investigated empirical evidence in GSD-related research literature. Nurdiani et al. [20] discuss risks and strategies to mitigate those risks in the context of GSD projects. Monasor et al. [21] present a SLR regarding methodologies for training and teaching students and developers to the needs of GSD projects. Nidhra et al. [22] present the state of the art and the state of the practice regarding knowledge transfer in GSD. Matusse et al. [23] investigated evidences about the existence of metrics and indicators that are specific to GSD. Many of these previous SLRs focused on the challenges and solutions regarding communication, coordination and cooperation problems in GSD projects [24]–[27].

Nevertheless, to the best of our knowledge, there is no SLR regarding effort estimation in the context of GSD. Since nowadays GSD plays an important role to the software industry, a SLR on effort estimation in the context of GSD is deemed quite important as its results on the state of the art in that
field can inform both researchers and practitioners. Therefore, the goal and main contribution of this paper is to present a SLR of effort estimation in the context of GSD. In doing so, research gaps that can be the focus of future work were also identified.

The remainder of this paper is organized as follows: Section 2 details the SLR, followed by the presentation and discussion of its results in Sections 3, 4 and 5 respectively. Finally, Section 6 presents the conclusions and view on future work.

II. SYSTEMATIC LITERATURE REVIEW

This section details the SLR, which includes the research questions that guided this SLR as well as the procedures employed to perform the entire SLR process. The SLR detailed herein followed the guidelines defined by Kitceham and Charters [28]. The protocol was developed jointly by the first, third and forth authors.

A. Research questions

The research questions were formulated by using the PICOC criteria [29]. Note that we did not include the comparison attribute because our focus is not to compare interventions. The PICOC is as follows:

- **Population** - Global software development projects.
- **Intervention** - Effort estimation methods/techniques/size metrics/cost drivers.
- **Outcomes** - The accuracy of the effort estimation methods/techniques.
- **Context** - Any possible study, as long as it is an empirical study within the context of GSD will be considered.

Therefore, the addressed research questions are:

- **Question 1** - What methods/techniques have been used to estimate effort in GSD?
  - 1a - What metrics have been used to measure the accuracy of effort estimation methods/techniques in GSD projects?
  - 1b - What are the accuracy levels for the observed estimation methods?
- **Question 2** - What effort predictors (cost drivers/size metrics) have been used to estimate effort in GSD?
- **Question 3** - What are the characteristics of the datasets used for effort estimation in GSD?
  - 3a - What are the domains represented in the dataset (academia/industry projects)?
  - 3b - What are the types represented in the dataset (single-company/cross-company)?
  - 3c - What are the application types represented in the dataset (web-based/traditional)?
- **Question 4** - What are the used sourcing strategies (offshore outsourcing/offshore insourcing)?
  - 4a - What are the countries involved?
  - 4b - How many sites are involved?
  - 4c - How many sites per country are involved?
  - 4d - Which topologies (centralized/distributed) have been used regarding the effort estimation process?
- **Question 5** - Which activities were considered in the effort estimation process?

B. Search strategy

After defining the research questions, we set up a strategy to define a search string and, consequently, identifying the primary studies. To avoid researcher bias, we used the following procedure to define the search string used in this paper:

1. Analyze the questions and identifying the main words in terms of population, intervention and outcome;
2. Analyze collected relevant papers and checking the keywords;
3. Point out alternative spellings and synonyms for major terms, if existing;
4. Connecting alternative spellings and synonyms by using the Boolean OR;
5. Linking the main terms from population, intervention and outcome by using the Boolean AND;
6. Evaluate the resultant search string and perform adjustments by using the quasi-gold standard [30].

The steps one to five were performed by the first author and the last one was performed jointly by the first and forth authors.

As a result, we obtained the following search string:

```
  (effort OR cost OR resource OR size OR metric OR measuring) AND (estimation OR estimating OR estimate OR prediction OR predicting OR predict OR assessment OR forecasting OR forecast OR calculation OR calculate OR calculating OR calculation OR sizing OR measure OR measuring) AND ("global software development" OR "globally distributed development" OR "globally distributed work" OR "distributed software development" OR "distributed software engineering" OR "distributed development" OR "multi-site development" OR "multi-site development" OR "multi-site development" OR "multi-site development" OR "geographically distributed software" OR "collaborative software engineering" OR "collaborative software development" OR "dispersed software development" OR "global software teams" OR "distributed teams" OR "spread teams" OR "dispersed teams" OR "global teams" OR "virtual teams" OR "offshore outsourcing" OR "offshore software development")
```

C. Search process

Once the search string was defined, the search process took place and was divided in two phases: initial search phase and secondary search phase.

1) The initial search phase: The initial phase involved identifying primary sources and searching primary studies through those sources by using the search string defined in Subsection II-B.

The used primary sources are listed in Table I, as well as the number of search results and number of relevant papers. We selected those primary sources based on venues where
the literature related to GSD and effort estimation is mostly published.

The search process was performed using titles and abstracts. In addition, results were limited to peer-reviewed conference papers and journal articles, published between 2001 and 2013. The starting year was set to 2001 because that was the year when the term Global Software Development was coined [31].

In the second stage, we read the full-text of the 24 papers selected in the previous stage. The secondary search phase (II-C) was performed concurrently. The final list of the first three authors were compared and discussed until consensus was reached.

As a result of the secondary search phase, the first author found an additional study. After a discussion with the other authors, that study was included, thus making it a total of 25 studies to be screened. This 25th study was identified as D1, to be differentiated from the other 24 studies retrieved in the initial search phase.

As a result of the second stage of the study selection, the authors consensually decided to exclude 17 studies, due to the following reasons:

- Nine studies were excluded because they did not focus on effort estimation models, methods, size metrics or cost drivers in the GSD context (Exclusion criterion 1).
- Eight studies were excluded because they did not provide empirical evidences to support their findings (Exclusion criterion 2).

Thus, by the end of the study selection phase, 8 papers passed our inclusion criteria and were assessed in terms of quality (Subsection II-E).

### E. Study quality assessment

In order to evaluate the quality of the selected studies, we developed a questionnaire with 13 questions. The quality score for a paper could range from 0 to 13. The higher the score, the better the quality of the evidence provided by the paper. The questionnaire was developed by using the guidelines defined by Kitchinham and Charters [28].

Each question could be answered using YES (1.0), NO (0.0) or PARTIALLY (0.5). If a question was answered with PARTIALLY, it means that the work addresses the question in a correct way, but the answer could not be confirmed by using the content of the paper. Any study which scored 3.25 or below (first quartile) was excluded.

The questionnaire had the following questions:

- Are the research aims clearly specified?
- Was the study designed to achieve these aims?
- Are the prediction techniques used clearly described and their selection justified?
- Are the variables considered by the study suitably measured?
- Are the data collection methods adequately detailed?
- Is the data collected adequately described?
- Is the purpose of the data analysis clear?
- Are the statistical techniques used to analyze the data adequately described and their use justified?
- Are the results discussed, rather than only presented?
Do the researchers discuss any problems with the validity/reliability of their results?
Are all research questions answered adequately?
How clear are the links between data, interpretation and conclusions?
Are the findings based on multiple projects?

The first and the second authors of this work have performed the quality assessment for the 8 selected primary studies. The third author evaluated one of the 8 selected studies. The results were discussed until consensus was reached. The final scores can be seen in Table II-E. Three studies were excluded due to their low quality score.

**TABLE II**
**FINAL SCORES FOR THE EVALUATED WORKS**

<table>
<thead>
<tr>
<th>Study ID</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>S7</td>
<td>10.5</td>
</tr>
<tr>
<td>S10</td>
<td>7</td>
</tr>
<tr>
<td>S11</td>
<td>10.5</td>
</tr>
<tr>
<td>S13</td>
<td>3</td>
</tr>
<tr>
<td>S15</td>
<td>13</td>
</tr>
<tr>
<td>S20</td>
<td>5</td>
</tr>
<tr>
<td>S24</td>
<td>3</td>
</tr>
<tr>
<td>D1</td>
<td>3</td>
</tr>
</tbody>
</table>

The lists with the included and excluded studies after the study selection and quality assessment are shown in Table II-E.

**TABLE III**
**INCLUDED AND EXCLUDED WORKS AFTER THE STUDY SELECTION AND QUALITY ASSESSMENT**

<table>
<thead>
<tr>
<th>Included works</th>
<th>Excluded works</th>
</tr>
</thead>
<tbody>
<tr>
<td>S7 [9], S10 [32], S11 [38], S15 [7], S20 [8]</td>
<td>S1 [33], S2 [34], S3 [35], S4 [36], S5 [37], S6 [39], S8 [10], S9 [40], S12 [41], S13 [42], S14 [43], S16 [44], S17 [45], S18 [46], S19 [47], S21 [48], S22 [49], S23 [50], S24 [51], D1 [52]</td>
</tr>
</tbody>
</table>

F. Data extraction

The first and the second authors of this paper have performed the data extraction for all the five final selected studies. The third author performed the extraction procedure for one of those five studies, thus to compare the quality of the extractions.

All the extracted values were compared and discussed until consensus was reached. The performed comparison was very important since some values were noticed just by one of the authors.

III. RESULTS

After the data extraction, the extracted data were consolidated in 12 tables so to make it easier to understand the relationship between data and the research questions.

The percentage column, which is part of all tables of this section (except for Table VI), indicates the percentage of studies out of the total number of studies (five).

A. Question 1

Data extraction for Question 1 looked at what methods which have been used for estimating effort in GSD, the accuracy metrics to evaluate the methods and also the calculated accuracy values. Those data were synthesized in Tables IV, V and VI.

Table IV shows the identified effort estimation methods/techniques, representing three different approaches: i) expert based approaches (Delphi [53], planning poker [54], expert judgment [55] and expert judgment based on ISBSG database [56]); ii) algorithmic based approaches (COCOMO II [55] and SLIM [57]); and iii) an artificial intelligence approach (Case-based reasoning [55]).

One of the primary studies did not investigate directly any effort estimation approach, but rather focused on the cost drivers. Another primary study used function points and use case points as surrogates for effort estimation techniques. However, both are size metrics. The primary study S20 presented an approach based on linear regression but that approach was not empirically evaluated. So, from S20 we only extracted cost drivers as they were empirically validated.

**TABLE IV**
**IDENTIFIED EFFORT ESTIMATION METHODS**

<table>
<thead>
<tr>
<th>Estimation method</th>
<th>Approach type</th>
<th>Study ID</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>COCOMO II</td>
<td>Algorithmic</td>
<td>S11</td>
<td>20</td>
</tr>
<tr>
<td>SLIM</td>
<td>Algorithmic</td>
<td>S11</td>
<td>20</td>
</tr>
<tr>
<td>Case-based</td>
<td>Artificial</td>
<td>S15</td>
<td>20</td>
</tr>
<tr>
<td>reasoning</td>
<td>Intelligence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expert judgment</td>
<td>Expert</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>ISBSG based expert judgment</td>
<td>Expert</td>
<td>S11</td>
<td>20</td>
</tr>
<tr>
<td>Delphi</td>
<td>Expert</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>Planning poker</td>
<td>Expert</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>Function point count</td>
<td>-</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>Use case point count</td>
<td>-</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>No estimation</td>
<td>-</td>
<td>S10</td>
<td>20</td>
</tr>
</tbody>
</table>

Table V shows the accuracy metrics used in the primary studies. As we can see, just one of the primary studies presented the accuracy of the proposed effort estimation approach based on the most used accuracy metrics MMRE, MdMRE and Pred(25) [55].

The MRE (Magnitude of Relative Error) is calculated by considering the absolute difference between the actual and estimated effort, relative to the actual effort. The MMRE and MdMRE are respectively the mean and median MRE. Pred(25) is the the percentage of estimates with a MRE of 25% or less.

The primary study S11 used just the difference between the actual and the estimated efforts (deviation). The primary study S10 did not use any accuracy metric, since that work just focused on the cost drivers. The studies S7 and S20 did not evaluate the considered effort estimation approaches in terms of accuracy metrics.

Table VI shows the measured accuracy approaches in terms of accuracy metrics.
Pred(25) to evaluate the accuracy of the presented effort estimation approach, thus making it difficult to assess how good were the estimation techniques put forward in studies S7 and S11.

In S11, three projects were used to evaluate three different approaches. The deviation between the actual and the estimated effort was calculated for each approach in the context of each project.

The ISBSG database [56], used in S11, provides three different estimation values: the lower time to perform a project; the time in which a project is more likely to be performed; and the maximum time to finish a project. So, in S11, for each project three deviations were calculated using the actual effort and the three above-mentioned values: Lower (L), Expected (E) and Upper (M).

When applying the thresholds suggested by Conte et al. [58] (good MMRE if calculated value ≤ 25%, good MdMRE if calculated value ≤ 25% and good Pred(25) if calculated value ≥ 75%) to assess the effort prediction accuracy detailed in S15, results suggest that the prediction technique proposed in S15 provided good estimates.

It is impossible to compare the estimation accuracy between the effort estimation approaches identified in our SLR, since most of the primary studies did not use any prediction accuracy metric. In addition, the two studies that employed some accuracy measure did not use the same one.

### B. Question 2

Data collection for Question 2 looked for data regarding the used cost drivers and size metrics. The extracted data is presented in Tables VII and VIII, where all the predictors that do not represent size measures were grouped as cost drivers. Note that whenever the same cost factor presented different names in primary studies, we chose the name we believed to be the most suitable and added it to Tables VII and VIII.

Table VII shows that there is a wide range of cost drivers in the selected primary studies, where only a single study (S7) did not show any cost driver. Most of the cost drivers that seem to impact GSD projects are related to the specific factors of globally performed software projects, like differences in culture, language and time zone. The cost drivers that were used the most were "Time zone" (80% of the primary studies), "Language and cultural differences" (60% of the primary studies), "Communication" (40% of the primary studies) and "Process model" (40% of the primary studies).
Table VIII lists the size metrics employed in the primary studies, showing that function points and lines of code are the most used size metrics in the context of effort estimation for GSD. One of the studies also employed use case points and user story points as possible size metrics. The primary study S10 did not show any size metrics because it focused solely on cost drivers.

The use of lines of code as size metric seemed somewhat surprising to us given that this metric is really very difficult to forecast early in the development life cycle [59]. Such difficulties were one of the main motivations as to why function points were proposed so many years ago.

It is interesting to notice that just length metrics (lines of code) and functionality metrics (function points, use case points and story points [60]) were identified in the selected primary studies, despite other possible types of size metrics (e.g. complexity metrics, like cyclomatic complexity [61]).

**TABLE VIII**

<table>
<thead>
<tr>
<th>Size metric</th>
<th>Study ID</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function points</td>
<td>S7, S11, S20</td>
<td>60</td>
</tr>
<tr>
<td>Lines of code</td>
<td>S11, S15, S20</td>
<td>60</td>
</tr>
<tr>
<td>Use case points</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>Story points</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>No size metric used</td>
<td>S10</td>
<td>20</td>
</tr>
</tbody>
</table>

**C. Question 3**

Data collection for Question 3 looked at domains and types regarding the data used to evaluate the proposals in each primary study. That question also looked at the evaluated application types. The data were arranged in Tables IX, X and XI.

Table IX presents the domain relating to the type of dataset used in the primary studies’ investigation. It shows that all the selected primary studies employed industrial data to evaluate and/or making conclusions about the effort estimation approaches and/or predictors (size metrics and cost drivers).

**TABLE IX**

<table>
<thead>
<tr>
<th>Domain</th>
<th>Study ID</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry</td>
<td>S7, S10, S11, S15, S20</td>
<td>100</td>
</tr>
<tr>
<td>Academia</td>
<td>none</td>
<td>0</td>
</tr>
</tbody>
</table>

Table X presents the identified dataset types in the primary studies. Two of the primary studies (S10 and S15) used single-company data (data that comes from a single company). Study S11 used cross-company data, since the ISBSG database was used to estimate effort in that study. Studies S7 and S20 did not state the type of the dataset used.

Table XI presents the identified application types used in the primary studies. We considered only two types of applications during the data extraction - traditional and Web-based applications. In this work, a traditional application means an application which does not require Web infrastructure. Only one study (S11) stated the types of the application used in the study, which included both traditional and web-based applications.

**D. Question 4**

Data collection for Question 4 looked at specific aspects of Global Software Development. The extracted data is presented in Tables XII, XIII, XIV and XV.

Table XII presents the identified sourcing strategies in the primary studies. During the data extraction, we considered three options to address the question about sourcing strategies:

- **Offshore outsourcing** - A company moves the software development to an external third party abroad [62].
- **Offshore insourcing** - A company moves the software development to a branch established abroad [62].
- A combination of both above-mentioned strategies.

Table XII shows that just the offshore insourcing strategy was considered by the selected primary studies (S7, S10, S11, S15). One of the works did not inform which kind of sourcing strategy was used (S20).

**TABLE XII**

<table>
<thead>
<tr>
<th>Sourcing strategy</th>
<th>Study ID</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Offshore insourcing</td>
<td>S7, S10, S11, S15</td>
<td>80</td>
</tr>
<tr>
<td>Offshore outsourcing</td>
<td>none</td>
<td>0</td>
</tr>
<tr>
<td>Both</td>
<td>none</td>
<td>0</td>
</tr>
<tr>
<td>Not stated</td>
<td>S20</td>
<td>20</td>
</tr>
</tbody>
</table>

Table XIII presents the identified number of involved countries in the collected works. In the selected primary studies, at least three countries were identified as involved in the context of a GSD project.

Table XIV presents the name of the identified countries, thus showing a wide range of countries. The USA, China, India and UK are the four countries mostly involved in GSD projects within the context of this SLR.

All the selected studies that presented this information had at least a site in a developed country and a site in developing...
country. Since one of the main reasons for using GSD is the cost-reduction, it is not surprising that these countries were identified. USA and UK have higher costs, specially regarding human resources, compared to countries like India [63].

### Table XIII

<table>
<thead>
<tr>
<th>Number</th>
<th>Study ID</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>S11</td>
<td>20</td>
</tr>
<tr>
<td>7</td>
<td>S10</td>
<td>20</td>
</tr>
<tr>
<td>10</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>Not stated</td>
<td>S15, S20</td>
<td>40</td>
</tr>
</tbody>
</table>

Table XV presents the identified number of sites in the primary studies. A primary study should have at least two participating sites, each placed in a different country, to be characterized as within the context of GSD.

We observed a maximum number of four countries in a GSD project. Study S7 did not state the exact number of sites, and studies S15 and S20 did not report the number of sites involved.

### Table XIV

<table>
<thead>
<tr>
<th>Name</th>
<th>Study ID</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>S7, S10, S11</td>
<td>60</td>
</tr>
<tr>
<td>UK</td>
<td>S7, S11</td>
<td>40</td>
</tr>
<tr>
<td>India</td>
<td>S7, S10</td>
<td>40</td>
</tr>
<tr>
<td>China</td>
<td>S7, S10</td>
<td>40</td>
</tr>
<tr>
<td>Malaysia</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>Japan</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>Taiwan</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>Ireland</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>Brazil</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>Slovak Republic</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>Finland</td>
<td>S10</td>
<td>20</td>
</tr>
<tr>
<td>Germany</td>
<td>S10</td>
<td>20</td>
</tr>
<tr>
<td>Norway</td>
<td>S10</td>
<td>20</td>
</tr>
<tr>
<td>Sweden</td>
<td>S10</td>
<td>20</td>
</tr>
<tr>
<td>Pakistan</td>
<td>S11</td>
<td>20</td>
</tr>
<tr>
<td>Not stated</td>
<td>S15, S20</td>
<td>40</td>
</tr>
</tbody>
</table>

Table XV presents the identified number of sites in the primary studies. A primary study should have at least two participating sites, each placed in a different country, to be characterized as within the context of GSD.

We observed a maximum number of four countries in a GSD project. Study S7 did not state the exact number of sites, and studies S15 and S20 did not report the number of sites involved.

### Table XV

<table>
<thead>
<tr>
<th>Number of sites</th>
<th>Study ID</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 more them 2</td>
<td>S10, S11</td>
<td>40</td>
</tr>
<tr>
<td>3</td>
<td>S7</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>S10</td>
<td>20</td>
</tr>
<tr>
<td>Not stated</td>
<td>S15, S20</td>
<td>40</td>
</tr>
</tbody>
</table>

Question 4d deals with the effort estimation process topologies. In this work, the topology of the effort estimation process is the way the work regarding the estimation process is performed by the involved sites. There are three possible topologies:

- **Centralized** - All the inherent effort estimation calculations are centralized in just one site and the remaining sites just supply knowledge and/or data to the central site.
- **Distributed** - Each site of a GSD project performs their own effort estimation process independently. The estimates only relate to what the site is doing as part of the project.
- **Hybrid** - All the effort estimation is performed as in a centralized topology, but late effort estimates can be obtained in a distributed fashion. We consider early effort estimation the one performed just after the requirement analysis. The late effort estimation process is the one performed after an implementation iteration.

Question 4e deals with the roles for sites in the effort estimation process. As above-mentioned, a site can be involved in the effort estimation process as a knowledge/data supplier and/or as the responsible for performing the inherent calculations.

We identified that none of the selected primary studies documented data about the aspects topology and the roles for a site in the effort estimation process. For this reason, we did not address Questions 4d and 4e.

### E. Question 5

Data extraction for Question 5 looked at the activities of a GSD project considered in the effort estimation process. Herein we considered generic activities embraced by most of the well-known life cycles, e.g. Requirements and Testing.

The main goal in extracting this data was to identify the life cycle activities considered the most relevant as far as effort estimation is concerned, within the context of GSD projects. However, it was not possible to address this question, since none of the selected primary studies documented this aspect either.

### IV. Discussion

In order to facilitate the reading of our observations regarding the extracted data, we divided our comments according to the SLR research questions.

Regarding Question 1 we have identified that:

- The used effort estimation approaches are well-known in the context of co-located projects. Those approaches have been used in the context of globally distributed projects without any kind of adaptation in the mechanisms of each approach.
- There is no standard effort estimation approach, since all the primary studies presented different solutions to the observed problem. However, it seems that expert-based approaches are the ones used the most by the practitioners.
- Software engineering works with imprecise and uncertain knowledge [64]. However, none of the primary studies treated explicitly the uncertainty regarding the data usually used to perform the effort estimation process.
- Since just one of the selected studies employed commonly used accuracy metrics, it was impossible to compare the primary studies in terms of prediction techniques’ estimation accuracy.

Regarding Question 2 we have observed that:
• The primary studies showed a wide range of cost drivers where most of them were related with the GSD specific issues, like cultural, language and time zone differences.
• None of the studies discussed deeply about the listed cost drivers.
• The observed primary studies show a trend in applying original co-located effort estimation approaches in GSD projects with cost drivers which reflect the specific problems of globally distributed projects.
• Solely length and functionality size metrics were identified in the studies, where the metrics used the most are lines of code and function points.

Regarding Question 3 we have identified that:
• The extracted data showed that only industrial data has been considered to evaluate the proposed effort estimation approaches.
• Single-company data appears to be the most used type.
• Most of the primary studies did not explicitly specify the type of the estimated application.

Regarding Question 4 we have observed that:
• None of the primary studies considered the particularities of each sourcing strategy.
• We did not identify any study considering offshore outsourcing as sourcing strategy. Therefore, there is a lack of research about effort estimation in offshore outsourced projects. Consequently, all the findings of this SLR are related to offshore insourced projects.
• The primary studies showed a trend in which there was always a site located in a country where salaries for software developers are quite high, and another site located in a country where salaries for developers are very low.
• The selected primary studies did not document the topology and the role of each site in the effort estimation process.

Finally, regarding Question 5, none of the selected studies explained how the data used to estimate the effort were extracted during the evaluated projects.

V. Threats to Validity

In terms of threats to the validity of this SLR, the major issue is whether we have failed to find all the relevant primary studies. In order to mitigate that threat, we performed a very deep search strategy. We used keywords found in the most relevant works in effort estimation and Global Software Development.

We cannot claim that we retrieved all the available literature. However, we can state that we retrieved as many studies as possible, considering the restrictions that we have applied to our search string to reduce the number of irrelevant papers.

We would like to remark that one important task performed during the search process, the secondary search phase, was performed solely to the primary studies selected in the initial search phase, it was just a small portion of the complete set of studies retrieved in the search process. However, since the inclusion and exclusion criteria of this SLR are very straight, we believe that most relevant works regarding the main topic of this work were evaluated.

It is important to note that only some of the initially retrieved studies complied with our inclusion criteria. Seven studies were excluded because they did not show any empirical evidence to support their proposals. Another three studies were excluded because of their quality. So, since we had just few studies to extract data and drawing conclusions about the state of the art of effort estimation in the GSD context, it is very hard to generalize the findings.

VI. Conclusions

This paper presented the results of a systematic literature review of effort estimation in the context of Global Software Development. The main goal of this paper was to present the state of the art regarding effort estimation in the context of GSD, in order to inform both research and practice.

The initial search phase returned 379 unique results and just 24 were selected. Out of those 24, 17 were excluded after a further investigation performed by reading the full text of those primary studies. The remaining 7 studies were then used in the secondary search phase, from which 1 more study was retrieved, bringing the total of selected studies to 8. Finally, after the achievement of the quality assessment process, we had a final list of 5 studies.

Only few studies complied with this SLR’s inclusion criteria. Many studies were excluded because either they lack empirical evidence (eight) or have low quality (three).

It is important to note that none of the selected primary studies considered offshore outsourcing as sourcing strategy, it means that all the findings of this SLR are applicable to projects which apply offshore insourcing as sourcing strategy.

The identified effort estimation approaches are well-known in the context of co-located projects. The main difference appears to be in the used cost drivers. There is no standard effort estimation approach and the observed effort estimation approaches did not consider the uncertainty that is inherent to such domain. It was not possible to evaluate the accuracy of the observed approaches because the studies did not use the same accuracy metrics.

This SLR identified a wide range of cost drivers, where most of the primary studies presented cost drivers regarding cultural, language and time zone differences; these are factors directly related to a globally performed software project and they make the fulfillment of those projects more challenging.

The selected primary studies showed that the researchers are using just industrial single-company data; some studies did not document anything regarding the used data, as well as the type of the estimated applications.

Most of the questions regarding specific aspects of GSD projects were not addressed. Studies did not document the effects of the used sourcing strategy and the topology of the effort estimation process. On the other hand, we identified that the primary studies showed a trend in which there were always a site located in a developed country, where salaries...
for software developers are quite high, and a site located in a developing country, where salaries for developers are very low. This was expected, since one of the biggest reasons to perform a software project globally is to reduce costs.

Finally, the selected primary studies did not state which activities of the development process were considered to calculate the effort estimation.

We believe that further investigation regarding effort estimation approaches used in co-located context is needed given the scarcity of primary studies identified in this SLR. We also put forward that the adaptation of those approaches considering the specific aspects of GSD and also the inherent uncertainty of the data [5] may provide more accurate effort estimates.

It seems that the GSD sourcing strategy and the effort estimation process topology can have a significant influence upon the effort estimates. For example, the transition of software or part of software developed by using the offshore outsourcing strategy seems to require more effort compared to the development carried out by using the offshore insourcing strategy. An effort estimation process performed in a distributed way seems to be more flexible and adequate to companies which apply agile methods.

However, since we did not find any evidence to support the above-mentioned assumptions, we believe that more investigation must be performed to confirm so. A single effort estimation technique may not be equally applicable to offshore outsourcing and offshore insourcing projects, thus warranting separate investigations that can later be combined. The same can be applied to the effort estimation process topologies.

Since we did not identify any study considering effort estimation in the context of offshore outsourced projects, we believe researches should explore this existent gap in order to provide solutions which could help practitioners in that kind of scenario.

Regarding the cost drivers, we believe that there is room for assessing their impacts on the effort and developing much more formal ways to measure them, which are directions for our future work.

VII. ACKNOWLEDGMENTS

We would like to thank FAPEMIG, CNPq, UFPI and INES, for partially supporting this work. We also would like to thank Professor Claes Wohlin for his feedback on a previous version of this paper.

REFERENCES


