Abstract—Context: Heuristic optimization has been of strong focus in the recent modeling of the Resource Constrained Project Scheduling Problem (RCPSP), but lack of evidence exists in systematic assessments. New solution methods arise from random evaluation of existing studies. Objective: The current work conducts a secondary study, aiming to systemize existing primary studies in heuristic optimization techniques applied to solving classes of RCPSPs. Method: The systemizing framework consists of performing a systematic mapping study (SM), following a 3-staged protocol. Results: 371 primary studies have been depicted from the multi-stage search and filtering process, to which inclusion and exclusion criteria have been applied. Results have been visually mapped in several distributions. Conclusions: Specific RCPSP classes have been grounded and therefore a rigorous classification is required before performing a systematic mapping. Focusing on recent developments of the RCPSP (2010-2015, a strong interest has been acknowledged on solution methods incorporating AI techniques in meta- and hyper-heuristic algorithms.

I. INTRODUCTION

Dealing with resource allocation in a constrained project scheduling space has risen to a solid multidisciplinary topic under the (project) scheduling theory, addressed since late 60s, as the Resource(–)Constrained Project Scheduling Problem (RCPSp). Empirically, the goal of the RCPSP is to determine a time and resource-feasible schedule, with given precedence or resource constraints, following an objective function. A comprehensive characterization of the RCPSP positioning in the research literature has been given by Mohring et. al (2003), where RCPSP “one of the most intractable problems in Operations Research”, has become “a popular playground for the latest optimization techniques, including virtually all local search paradigms”.

A. Problem context and motivation

According to the (computational) complexity theory, the RCPSP has been widely characterized as belonging to the class of NP-hard (in the strong sense) problems, with NP-complete (in the strong sense) decision version, therefore combinatorial optimization techniques do apply. Several dimensions, namely research themes, fundament a topology construction for the Project Scheduling Problem (PSP)/RCPSP space, mainly consisting of: project type, problem approaches, solution methods. Two main implications, can be therefore considered:

1) Proposing new solution methods require a solid mapping of recent and past advancements in solution techniques with an extensive literature search on multiple RCPSP dimensions. In such a context, a literature review process carried out to propose a new solution, involves a cross-domain scanning of a large amount of existing evidence in connected fields (Operations Research, Artificial Intelligence, Applied Mathematics), where computational comparison and benchmarking is of use in performance assessment.

2) Proposing new RCPSP solution methods follows trends of converging research areas, lacking in a systematic assessment in the RCPSP problem space to identify trends and gaps. Current solution methods move forward with connected fields, and include Intelligent Systems and AI techniques, without a stand-alone specific process to identify gaps, based on existing trends. Even though RCPSP research themes have been identified and fully characterized, current secondary studies in RCPSP have the form of reviews, surveys, classifications, and are to be included in the primary studies’ class, as they do not follow a rigorous systemic approach in synthesizing literature review, but provide more of narrative descriptions.

B. Proposed approach and outline

The current work applies the principles of systematic mapping studies (SM) to the RCPSP, with the aim of tracking existing solution methods and identifying trends and gaps based on several studies’ distributions with a focus stands in classic heuristic algorithms and its extended classes (meta-heuristics, hyper-heuristics). Section 2 provides an overview of existing guidelines for the systematic mapping and implications in Operations Research and Scheduling. Section 3 describes the research protocol applied and the studies’ frequency resulted in each stage. Section 4 provides an analysis of several studies’ mappings according to time-based and problem space-based distributions. Suggestions for further improvement are described in Section 5.
II. BACKGROUND AND EVIDENCE

A. Systematic mapping and implications in scheduling

Although large-scale reviews are conducted under the Systematic Literature Review (SLR) umbrella, most of the large-scale SLRs in Software Engineering (SE) and IT are conducted using the Systematic Mapping guidelines [1]. Derived from Kitchenham’s guidelines [2], Petersen and Budgen [1], have proposed an extension of guidelines for constructing systematic maps in software engineering. Additionally, a more detailed presentation of the included studies in Kitchenham’s [3] tertiary review, has been offered with respect to the SLRs characteristics. Of relevance, is the comparison between the number of potentially relevant studies and the effective number of relevant studies included in each aforementioned SLRs. According to Petersen [4], Dyba et al. performed a SLR based on 5453 potentially relevant studies, from which 78 have been extracted. The same number of primary studies was used by Hannay et al., Kampenes et al. and Sjøberg et al. with a result of 24, 78 and 103 finally selected studies. MacDonell & Shepperd used the least number of potentially studies (185), to achieve a relevant set of 10 studies. Most of the aforementioned SLRs have included between 1-5% of the primary studies (Dyba et al. – 1.43%, Grimstad et al. – 2.49%, Hannay et al. – 0.44%, Kampenes et al. – 1.43%, Kitchenham et al. – 0.74%, Sjøberg et al. – 1.88%, MacDonell & Shepperd – 5.45%, Davis et al. - 4.6%), with an exception of Mendes (49%). Correlated to Kitchenham’s [3] contribution in aggregating SLRs under a tertiary study, Bugen [5] has applied Kitchenham’s guidelines [2] to SMs, and conducted informal tertiary studies.

Latest SM literature, focuses on refining guidelines and protocols, as well as providing improvement strategies. While emphasizing the benefits of using SM in educational environments, as means of building students’ transferable research skills, 3 challenges are positively assessed by Kitchenham [6]: 1) no matter the level of studies (post-, undergraduate), students have the required skills of conducting SMs 2) SMs are valuable means of organizing evidence of current state of literature 3) SMs can be treated as solid projects as preparation for research careers. Several further directions, referred as improvements [7], still need to be covered in the future: search and selection process improvement; quality assessment grounding, studies aggregation process improvement. Based on these improvement statements, Petersen [8] offers a valuable update to the existing SM framework applied to software engineering, by mapping evidence regarding 4 research questions: frequency of publications in the SE field, covered topics, revenues of publications, and review process. A need of conducting SMs on existing or extracted topic-specific classifications has been pointed out, observing that the majority of SM studies deliver new classification schemes themselves, rather than building on existing ones.

In terms of extending its field applicability [8], the SM process can be alternatively used as means of evidence gathering in operational fields, for growing disciplines with multiple facets. Applicability of systematic studies to operational fields has been recently acknowledged in the manufacturing [9] and industrial software processes [10]. With respect to the scheduling field, a SLR has been conducted for staffing and scheduling problems in software projects, where 52 papers have been extracted and analyzed with the purpose of identifying main issues in adopting resource scheduling features in industry. Specific to software scheduling, a SLR has been performed [11] addressing uncertainty assessment in software projects, where 165 studies have been analyzed in 5 distributions.

B. Comparative studies in heuristic optimization for solving the RCPSP problem

The Project Scheduling field, specifically the RCPSP, can therefore be introduced as a successful candidate for extending MSs to new fields. In the fields of complexity, discrete and combinatorial mathematics [19], [20], the RCPSP’s main characterization is the generalization of the job-shop scheduling problem [21], although the job shop scheduling problem is sometimes treated as a “special case” of the RCPSP [22]. Basically, the optimization problem is to define a start time for each project activity, based on precedence and resource constraints, while following a uni or multi-objective (project makespan minimization, time/cost/resource trade-offs etc.).

Large amount of effort has been put during the past decade in building RCPSP classification schemes. Two main dimensions have been extensively tackled: general problem description [12][13] and solution methods[14]–[16], while [17] extends the attributes of the classification scheme to: type of constraints, type of precedence relations, type of resources, source, type of activity splitting, number of execution modes, number of objectives, type of objective function, level of information, distribution of information. Heuristic methods have been considered as state-of-art solution techniques and a comprehensive description with computational analysis has been presented since late 90 [14]. Based on the Artificial Intelligent and Mathematic Optimization advancements, in the field of Computer Science (CS), classic heuristics applied to RCPSP (exact methods) have evolved into modern heuristics: meta-heuristics (MH) (including evolutionary MH, memetic algorithms, nature-inspired MH), hyper-heuristics, simheuristics and hybrid-heuristics. An attempt of mapping the solution space of the RCPSP, in terms of solution methods, has been presented by [18], where 174 studies (1971-2012) have been classified based on a defined taxonomy. No systematic assessment has been carried, though. Increased interest in the use of meta-heuristics (both derived and applied or just applied) has been shown during the past 5 years: a computational comparison based on the PSPLIB J30, J60 and J120 problem sets has been presented by [19], [20]. MHs applied to the multi-mode, multi-skill and multi-objective RCPSP have been compared by [21]–[24]. Analysis and comparison of combined meta-heuristics (hybrid-MH) applied to the RCPSP and dynamic RCPSP has been summarized in [21]–[26].
III. RESEARCH METHODOLOGY

To assess the heuristic techniques applied to solving scheduling problems, with focus on the RCPSP formulations and recent developments, a systematic mapping has been conducted following guidelines of Kitchenham [27] as well as updates provided in [28], [4], [8].

A. Review question

On the basis of systemizing evidence regarding heuristic optimization applied to the RCPSP, the following research question has been considered:

RQ: Which is the evolution of the heuristic algorithms used as solution methods for the RCPSP and to which RCPSPs have they been applied?

B. Search strategy

The search strategy represents a combination between a set of search strings and databases on which to be applied. 4 databases have been selected: IEEE Xplore (http://ieeexplore.ieee.org/Xplore), Elsevier Science Direct (http://www.sciencedirect.com), SpringerLink (http://link.springer.com/), SCOPUS (http://www.scopus.com/) (Table 1). As the RQ of the current study addresses solution methods in forms of techniques and algorithms, IEEE Xplore has been considered relevant. On the other hand, the RCPSP is an OR specific problem, therefore, 2 multidisciplinary databases have been chosen: Springer Link, Science Direct.

<table>
<thead>
<tr>
<th>Database</th>
<th>Rationale and limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEEE Xplore</td>
<td>multi-format export (includes .csv, .bibtex key), all metadata export (includes abstract); number of entries/export limited at 100</td>
</tr>
<tr>
<td>Science Direct</td>
<td>multi-format export (no .csv export, includes .bibtex key), metadata does not include abstract; number of entries/export limited at 100</td>
</tr>
<tr>
<td>Springer Link</td>
<td>multi-format export (includes .csv, .bibtex key), metadata does not include abstract; number of entries/export limited at 100</td>
</tr>
<tr>
<td>SCOPUS</td>
<td>aggregates literature from multiple databases; multi-format export (includes .csv, .bibtex key), metadata does include abstract; number of entries/export limited at 100</td>
</tr>
</tbody>
</table>

TABLE 1. RATIONALE FOR DATABASE CHOICE

Six search strings have been constructed based on the PICO strategy [8] (Population, Intervention, Comparison and Outcomes) and submitted for inquiry to each database: S1: project scheduling *heuristic*; S2: "project scheduling" *heuristic*; S3: resource constrained project scheduling *heuristic*; S4: resource-constrained project scheduling *heuristic*; S5: "resource constrained project scheduling" *heuristic*; S6: RCPSP* *heuristic*. The string generation strategy follows the principle of the selection process, where, firstly, a list of the literature-proposed heuristic algorithms was extracted, followed by a list of possible applicable algorithms and a mapping of solution methods on RCPS problem types. No manual search or snowballing has been performed. Special characters have been used to compress search strings, where applicable.

C. Study selection and frequency distribution

The process of study selection as been divided in 5 major steps.

The first phase focuses on preliminary study selection and assessment (Fig.1). Constructed search strings have been applied on the 4 databases individually, as an evidence of the studies frequency, in each case, was of interest. For each search string a list of studies has been exported automatically from each database, using built-in exporting options. Lists have been aggregated for each database (IN1) and duplicates have been removed automatically (IN2). Filtering based on title keywords has been applied ("project" AND "schedule"; RCPS), following a word-form derivation (i.e. "scheduling"; "schedule (s)"), to extract studies relevant for the project scheduling area (IN3A), specific, RCPSP (IN3B). Results have been merged and duplicates removed (IN4). Studies referring to other scheduling scenarios (i.e. machine scheduling, network scheduling, etc.) have been excluded, being of no interest for the current work. A second filtering based on title keywords extracted studies that only address solution techniques. General keywords have been considered (i.e. “heuristic” or “algorithm”) (IN5A). Solution-specific algorithms (i.e. metaheuristics optimization algorithm: Tabu Search, Simulated Annealing, Ant Colony Optimization etc.) do not contain any intuitive common keywords, therefore a specific filtering has been applied based on metaheuristic optimization taxonomies [29]–[32](IN5B).

![Fig. 1 MS preliminary study selection and assessment workflow](image-url)
Based on the given references, Table II includes the criteria based on which filtering has been performed to identify specific modern heuristics.

**Table II.**

**STUDIED FREQUENCY IN PRELIMINARY SELECTION AND ASSESSMENT**

<table>
<thead>
<tr>
<th>General solution method</th>
<th>Algorithm-specific keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>General solution method</td>
<td>“heuristic” OR “algorithm”</td>
</tr>
<tr>
<td>PS/RCPS</td>
<td>“local search” OR “grasp” OR “variable neighbourhood” OR “guided local” OR “iterated local” OR “basic local” OR “simulated annealing” OR “hill climbing” OR “tabu search” OR “random optimization” OR “evolution” OR “genetic” OR “memetic” OR “swarm” OR “stochastic scatter” OR “ant colony” OR “particle” OR “bee colony” OR “immune” OR “neural” OR “hybrid”</td>
</tr>
<tr>
<td>Modern Heuristic</td>
<td>“electromagnetism” OR “frog” OR “multi-pass” OR “filter and fan”</td>
</tr>
</tbody>
</table>

Resulted lists from each database for all search strings (S1, S2...S6) have been merged in a single list and duplicates have been automatically removed (IN6). Studies distribution for each step are presented in Table III.

**Table III.**

**STUDIED FREQUENCY IN PRELIMINARY SELECTION AND ASSESSMENT**

<table>
<thead>
<tr>
<th>Database</th>
<th>IN1</th>
<th>IN2</th>
<th>IN3</th>
<th>IN4</th>
<th>IN5</th>
<th>IN6</th>
</tr>
</thead>
<tbody>
<tr>
<td>IEEE Xplore</td>
<td>817</td>
<td>289</td>
<td>127</td>
<td>133</td>
<td>63</td>
<td>51</td>
</tr>
<tr>
<td>Science Direct</td>
<td>7633</td>
<td>4768</td>
<td>330</td>
<td>342</td>
<td>102</td>
<td>70</td>
</tr>
<tr>
<td>Springer Link</td>
<td>11129</td>
<td>6679</td>
<td>343</td>
<td>373</td>
<td>94</td>
<td>77</td>
</tr>
<tr>
<td>SCOPUS</td>
<td>3864</td>
<td>1239</td>
<td>624</td>
<td>659</td>
<td>324</td>
<td>185</td>
</tr>
</tbody>
</table>

Relevancy-based study selection and assessment has been conducted in the second phase (Fig.2). Resulted lists from each database have been merged and duplicates have been removed automatically based on title. A casual manual check has been undertaken to remove duplicates. Filtering based on title keywords has been applied, to extract studies that addressed resource-oriented project scheduling (i.e. “resource”; RCPSP). Both keywords have been separately applied as the studies frequency in each case was of interest. Another relevancy-based criterion tested the relevancy of the solution method. Studies referencing the “algorithm” keyword in their titles have been manually checked for relevancy to the current work. All the 188 entries, have been though found relevant, therefore included in the further process. A total number of 359 entries have been obtained and submitted as input for the next phase.

Language and content type exclusion (EC) and inclusion criteria (IC) have been applied in the 3rd step. The language criterion has been considered eliminatory. From entries that had language specified in metadata, studies in other languages than English, such as Spanish, Chinese have been excluded, (EC). From the resulted entries, only English studies were extracted (IC). In case of filled in language attribute, that did not respect the EC, the language EC was updated and entries re-checked to pass the IC. Entries with blank language field, have been checked manually based on the publication language. The language field has been updated for all the remained items. All studies with updated language field have been English-written, therefore successfully passed EC/IC check and included in the final list. A list of 339 studies have been obtained by applying the language EC/IC.

![Fig. 2 MS relevancy study selection and assessment workflow](image-url)
The next phase, consisted of applying a content type-based language EC/IC, in a similar way to the language-based EC/IC. Studies that have been categorized as technical papers or reports, reviews, workshop papers or book chapters have been excluded (EC). Papers published in peer-reviews publications (journals) or conference proceedings have been included (IC). Papers not categorized have been submitted for manual checking based on the “Publication title” attribute. 22 studies had the content type updated. Based on the IC criteria, 18 studies have been selected and 4 removed. A total number of 295 have been therefore obtained. Both phases of EC/IC (language and content type) are illustrated in Fig.3.

A manual refinement has been undertaken as a last step of the selection process, based on abstracts’ screening. 4 studies had the full content published in Chinese and have been excluded based on language (4 studies). Studies that were proposing new solution approaches were of interest, therefore, (computational or experimental) evaluation or comparison papers, as well as classification papers, have been excluded (15 studies). Where in doubt, full text has been screened. Generalized studies that have not addressed a particular RCPS problem or solution method (i.e. generalized algorithms not addressing any heuristic/meta-heuristic solution method), as well as studies that could not be positioned in a classification scheme (either irrelevant or not accessible) have also been excluded (10 studies). A final list of 252 studies has been furtherly considered for mapping and analysis.

IV. STUDIES ANALYSIS AND MAPPING

To answer the proposed research question (RQ), several types of mappings have been extracted: studies identification on RCPS problem types and solution algorithms classes, frequency of publications for classes of RCPSP solution methods and algorithms, as well as algorithm mapping on the specific RCPS problem types. Statistics have been extracted based on title, abstract or brief content screening (when required), without a full evaluation, as in case of SLRs [4].

In terms of visual representations, several assumptions have been made in the SM/SLR literature. With respect to SLRs, Kitchenham [27] emphasizes on the use of forest plots as being the most common mechanism for presenting quantitative results, while vulnerability to publication bias is likely to be assessed by using funnel plots. On the other hand, Petersen [8], in the study of SM guidelines updating, identifies 6 approaches to visualize SM mappings. Based on the analysis, bar plots and bubble plots proved to be the most common representations, while heatmaps, although rare, were considered interesting ways to directly visualize the relative amount of publications in different categories [8]. Based on the above observations, the current study proposes bar charts and bubble plots as visual representations. Three categories of mapping have been considered relevant to identify the evolution of heuristic optimization applied to the RCPSP: a) distribution of heuristic classes on years b) distribution of heuristic algorithms on years c) heuristic algorithms solving classes of RCPSP problems.
Several year-based scales have been considered to be relevant for the studies mapping: unique discrete year values have been considered for the 2010 – 2016 (first decade) interval; 4 intervals have been considered between 1990 – 2009; one category for studies published before 1990. Acronyms have been used for RCPS problems definition or heuristic optimization classes and algorithms, following existing taxonomies. A bubble chart representation has been preferred, according to updated guidelines [8], where bubble plots are recommended ways of representing frequency distributions [4]. Bubbles have been positioned in the intersection of two x-y scatter plots. The size of the bubble indicates the number of studies for the given intersection point.

The first mapping, answers the given RQ by presenting a time-distribution of classes of heuristic optimization algorithms (Fig.4). The mapping process was conducted based on title, and when in doubt, the abstract was scanned. Studies addressing more than one solution method have been mapped to multiple categories. Several classes of algorithms have been identified: 2 classes referencing more generalized studies (H: general heuristics, MH: general metaheuristics); 2 classes of algorithms addressing extended heuristics (SH: simheuristics, HH: hyper-heuristics); xH: other heuristics; N/A: not assigned heuristics. All the other classes (H-MH: hybrid MH; T-MH:trajectory-based MH; E-MH:evolutionary MH; SI-MH:swarm intelligence-based MH; NI-MH:nature-inspired MH excluding SI-MH; ML-MH: machine learning-inspired MH; P-MH: probability-based MH; D-MH: deterministic MH; EM-H: exact methods H; Pb-H: priority-based H; I-H: improvement H). Studies have been included in a separate category, when more than one algorithm was addressed as “improvement” of one class, and properties of H-MH; HH, MA (memetic algorithms) were applied (screening of abstracts). 252 studies have been extended to 288, some entries presenting more than one algorithm.

2 metaheuristic categories proved to be dominant: E-MH (29.1%) and SI-MH (16.3%), followed by the class of T-MH (11.4%) and H-MH (10.4%). In both cases, an increased interest is shown during 2013-2016: E-MH (25 studies), SI-MH (16 studies) compared to the period of 2005-2009, referenced in the literature: E-MH (20 studies), SI-MH (16 studies). From the state-of-the art heuristics, as advertised, Pb-H are the most popular algorithms, including: scheduling schemes and priority rules (7.6%).

Both E-MH and SI-MH classes have been decomposed on specific optimization algorithms (Fig.4.b and Fig.4.c). Studies belonging to the E-MH classes included solution methods such as: EA (evolutionary algorithms and strategies), GA (genetic algorithms), DE (differential evolution); MA (memetic algorithms), SFLA (shuffled frog leaping algorithm). Studies categorized as belonging to the class of SI-MH, addressed algorithms such as: ACO (ant colony optimization), BCO (bee colony optimization); CS (cuckoo search), CSO (cat swarm optimization), FA (firefly algorithm), AIS (artificial immune system).
According to the year-based distributions, preferred algorithms are GA (52 studies) and PSO (18 studies), followed by EA (13 studies) and ACO (12 studies). Fewer attempts for developing RCPSP solutions included: derivations of CSO, CS, FA. Compared to other classes of algorithms, the SI-MH appear to be a recent class of solving algorithms, having a debut in the 2000-2004 period. Several independent studies present directions for future interest: a) despite the specific identified algorithms, 2 studies have been depicted addressing solving approaches based on learning strategies b) 1 agent-based solution method.

According to the RCPSP solution space, a solution framework consists of both solution method and the problem definition addressed. Fig.4. maps the distribution of general modern metaheuristics (HH, MH, H-MH), the specific algorithms belonging to the E-MH, SI-MH classes, as well as T-MH class (LS: local search, GRASP: greedy randomized adaptive search, VNS: variable search neighborhood, SA: simulated annealing, TS: tabu search) on RCPSP problems. 19 specific problem classes have been identified in the extracted list of entries, based on existing classification schemes [17]. Additionally, 91 studies addressed a general class of the RCPSP, while 62 studies did not address the specific RCPSP. 4 classes of algorithms, where found of not addressing any RCPSP problem (xH, HH, Ni-MH, xH). The frequency distribution presents the MMRCPSP (multi-mode RCPSP) (14.2%) and the RCMPSP (multi-project RCPSP) (6.2%) as the most addressed problems. Solution methods for the MMRSPSP are popular within E-MH (17 studies) and SI-MH (7 studies). In terms of general heuristic optimization classes, H-MH (14 studies) algorithms prove an increased popularity in addressing specific RCPS problems (F-RCPSP, MMRCPS, MS-RCPSP), comparable to the T-MH class.

V. CONCLUSIONS

While current studies that seek to provide a representation for the RCPS problem space are more oriented on taxonomy construction, the current work adopts the structured process of SM studies to provide an overview and decomposition on the evolution of solution methods and their applicability to specific RCPS problems. Several contributions are to be claimed: 1) providing a framework under which the SMs can be extended to other fields, specifically validation of SM applicability to the RCPSP 2) validation of the solution methods and problems classification based on an extensive database of studies 3) trend identification and decomposition based on existing solution evolution. To reduce bias in assessment we recommend a further evaluation of quality metrics, an optional step in conducting SMs (opposite to SLRs, where quality assessment is mandatory to validate study selection and filtering). A second proposed improvement addresses construction of initial search strings based on words synonymy or derived attributes (i.e. “strategies”, “methods” when referring to “algorithms”). For a complete validation and for the purpose of transforming a SM in a work of reference for the RCPSP, computational comparison is required to evaluate algorithms performance, based on specific setups. Several existing RCPSP libraries (PSPLIB, Peterson set) provide project descriptions (number of activities, types of activities and resources) on which proposed algorithms are tested for computational parameters (average/min/std. deviation, computational time etc), in different scenarios (number of iterations). Therefore, the current study provides the input for a further SLR, being a successful candidate for a meta-analysis, that would provide a more solid background in algorithmic benchmarking.

VI. REFERENCES
