Supplier Selection for Global Service Providers: a Decision Support System

Bruno Petrato Bruck[a] and Manuel Iori[b] and Carlo Alberto Magni[c] and Daniele Pretolani[b] and Dario Vezzali[b,d]

[a]Centro de Informática, UFPB, 58058-600 João Pessoa, Brazil; [b]Department of Sciences and Methods for Engineering, UNIMORE, Via Amendola 2, 42122, Reggio Emilia, Italy; [c]Department of Economics "Marco Biagi", UNIMORE, Viale Berengario 51, 41121, Modena, Italy; [d] "Marco Biagi" Foundation, UNIMORE, Largo Marco Biagi 10, 41121, Modena, Italy

ARTICLE HISTORY
Compiled May 27, 2024

Abstract
In this paper, we develop a decision support system (DSS) aimed at solving a real-world supplier selection problem (SSP) for a global service provider (GSP) operating in the facility management (FM) industry. The GSP provides its customers with FM services, which are subcontracted to external suppliers selected on the basis of multiple criteria, like economic soundness, quality of service, capacity, and closeness. The SSP is formulated as a multi-objective generalized assignment problem, where the quality and the closeness of the selected suppliers are maximized, whereas a penalty produced by overcapacity assignments is minimized. The quality of each supplier is computed by applying a weighted sum method, resulting from a multi-criteria decision analysis in which the criteria weights are determined through an Analytic Hierarchy Process. The DSS is developed using a modular architecture with a relational database, a supplier evaluator, and a simulator, as well as an additional user-friendly interface. The simulator relies on a rolling horizon algorithm and three alternative configurations to assign contracts to suppliers. The effectiveness of the DSS is assessed by means of extensive computational experiments on historical data. The results show a significant average improvement of 25% compared to the solution adopted by the company.

KEYWORDS
Decision Support System, Supplier Selection Problem, Global Service Providers, Integrated Multi-Criteria Decision Analysis

1. Introduction

The supplier selection problem (SSP) is an important strategic issue in supply chain management. Carefully selecting suppliers may indeed lead to strong advantages over the competitors (Goffin et al. 1997), while adopting a multi-criteria evaluation that considers several characteristics of suppliers may help select the best ones. As described in the survey by Ho et al. (2010), quality, delivery, and cost are the most common criteria in the SSP, but other criteria may be equally important depending on the field of application (see, e.g., Weber et al. 1991 and Aguezzoul 2014). However, grouping and weighing multiple criteria is not trivial, and an accurate analysis is required to achieve the best results in terms of economic and operational performance.

CONTACT Dario Vezzali. Email: dario.vezzali@unimore.it

Carefully selecting suppliers is also critical in the Facility Management (FM) industry, where Global Service Providers (GSPs) compete with each other to provide their customers (e.g., banks, hotels, offices) with FM services (e.g., air-conditioning and heating systems, cleaning services, electrical systems, elevator maintenance, fire protection systems) by subcontracting their execution to qualified external suppliers. Establishing a comprehensive multi-criteria evaluation is key to supporting GSPs in selecting the most appropriate business partners.

Multi-criteria decision analysis (MCDA) is a well-known field of research that focuses on decision-making problems, such as ranking, choice, and sorting, in which multiple criteria are considered in the decision-making process (Pomerol and Barba-Romero 2000; Ishizaka and Nemery 2013; Greco et al. 2016). For this reason, the application of MCDA to the SSP is of remarkable interest. In addition, integrated approaches combining MCDA and other methods (e.g., multi-objective optimization, simulation) are somewhat diffused in the SSP literature (Ghodsypour and O'Brien 2001; Çebi and Bayraktar 2003; Ho et al. 2010).

This paper presents a DSS aimed at solving a real-world multi-period multi-objective SSP arising in the FM industry. Specifically, the DSS was designed to help *H2H Facility Solutions SpA*, an Italian GSP based in Bologna, in the selection of the best suppliers for some FM contracts. *H2H Facility Solutions SpA* provides its customers with a number of services, classified as planned preventive maintenance, corrective maintenance, or extraordinary maintenance. The categories of service provided range from air-conditioning and heating systems to electrical systems, elevator systems, fire protection systems, cleaning services, alarm systems and security, and so forth. *H2H Facility Solution SpA* faces a decision problem any time an FM contract for a category of service must be subcontracted. To help the company solve this problem, we developed a DSS, which relies on a particular multi-objective formulation of the generalized assignment problem whose objective is to determine the optimal assignment of contracts to suppliers by (i) maximizing the suppliers' quality score, (ii) minimizing the suppliers' distance score, and (iii) minimizing the suppliers' penalty score induced by overcapacity assignments.

The quality score was determined from a hierarchical tree of criteria determined together with the company. To compute the weights of the identified criteria, we implemented a simplified Analytic Hierarchy Process (AHP) and a revised Simos' procedure. Following a thorough data preparation process, the quality score of suppliers for each category of service was then obtained. The distance score is, instead, intended to take account of the distance between customers' facilities and the appointed supplier's location. Indeed, the closeness between customers and suppliers is desirable as it guarantees better compliance with service level agreements. The penalty score aims at penalizing the assignment of contracts that exceed the suppliers' service capacity.

It is important to note that, in the SSP application addressed in this work, all suppliers that intend to collaborate with the GSP sign a long-term framework agreement, which includes a detailed determination of the cost of each category of service provided. As a consequence, we do not seek to minimize such a cost and we intend to find the best supplier only in terms of quality score, distance score and penalty score.

To formally describe and solve the decision problem, a multi-objective mixed integer linear programming (MILP) model and a heuristic algorithm were developed. They were both implemented and integrated in a rolling horizon algorithm to perform the assignment of FM contracts to suppliers. Furthermore, a web application with a user-friendly interface was developed and tested with potential users on historical data provided by the company as well as additional data collected with an online survey sent to a restricted list of suppliers. In addition, several computational experiments were performed to assess the effectiveness of the proposed methods and gain some practical insights.

Apart from presenting a real-world case study involving a large number of suppliers, our work is of broad interest as it provides both a strategic and operational tool for dynamically solving a multi-period multi-objective SSP by combining MCDA and multi-objective optimization. In this sense, it represents a valuable contribution to the FM literature, where little attention has been paid to the adoption of DSSs based on MCDA for multi-objective SSPs. Additionally, we are among the first to consider the dynamic aspect of periodically updating the score of suppliers when solving multi-period SSPs, which is a noteworthy contribution to the SSP literature.

The remainder of the paper is organized as follows. Section 2 reports a review of the literature on integrated approaches for supplier selection and a brief overview of the FM literature. The SSP in the context of GSPs is formally defined in Section 3. The MCDA, the computation of weights, and the computation of quality scores are discussed in Section 4. Section 5 provides some details on the DSS implementation. Section 6 contains the outcome of extensive computational results. Finally, in Section 7, we draw some conclusions and outline possible future research directions. A preliminary version of this work, reporting a limited set of experiments obtained only by using a simplified greedy algorithm was presented as Bruck et al. (2021).

2. Literature Review

Integrated approaches combining MCDA and mathematical programming for supplier selection have been widely studied over the past thirty years. We refer the interested reader to Ghodsypour and O'Brien (1998, 2001) for seminal works, and to Ho and Ma (2018), Chai and Ngai (2020), and Saputro et al. (2022) for recent surveys. Furthermore, in recent years, the issue of sustainability has been attracting more and more attention in supply chain management due to environmental concerns. For an overview of the green SSP we refer to the survey by Govindan et al. (2015), while for relevant real-world cases, we refer to Fallahpour et al. (2017) and Gupta et al. (2019).

The AHP is a well-known multi-criteria decision-making method, whose purpose is to break down a decision (e.g., a selection or ranking problem) into factors arranged according to a hierarchical structure (Saaty 1990). Because of its simplicity, ease of use, and flexibility, the AHP can be applied both individually or in combination with other methods. Among the multi-criteria decision-making approaches for supplier evaluation and selection surveyed by Ho et al. (2010), AHP-based integrated approaches have proven to be the most widely used. In addition, Ho and Ma (2018) mention that AHP-based integrated approaches are frequently applied in the areas of manufacturing and logistics, where the most commonly studied problem is supplier evaluation and selection.

The integrated approach that most interests our work is AHP-mathematical programming (Ghodsypour and O'Brien 1998). Among other techniques used in conjunction with AHP for supplier evaluation and selection, one finds Fuzzy Set Theory (Chan and Kumar 2007), Lexicographic Goal Programming (Çebi and Bayraktar 2003), Goal Programming (Kull and Talluri 2008), Preemptive Goal Programming (Wang et al. 2004), and Dynamic Programming (Mafakheri et al. 2011). See also Ho and Ma (2018) for a review of these papers.

The field of FM is relatively new with seminal ideas presented by Kincaid (1994), Amaratunga et al. (2000), and Amaratunga and Baldry (2003). Over the years, it has garnered attention leading many authors to define indicators for measuring facility performance (see, e.g., Lavy et al. 2010 for an overview). Recently, the focus has shifted towards the integration of digital technologies into FM (Wong et al. 2018), specifically the adoption of building information modeling (see, e.g., Chen et al. 2018, Gao and Pishdad-Bozorgi 2019, Matarneh et al. 2019, and Patacas et al. 2020 for in-depth reviews). Despite these advancements, there remains a noticeable gap in the literature concerning the definition of relevant indicators for measuring the performance of suppliers in this specific field and MCDA-based decision-supporting technologies for supplier selection. The only contribution in this specific area comes from Pun et al. (2018). Their work presents a cloud-based DSS using Fuzzy-AHP as a multi-criteria decision-making method. While the study selects the best supplier on the basis of five main criteria, it did not conduct extensive tests using real-world data.

Interesting real-world applications of multi-criteria supplier selection may well be found in other fields where, in the last few years, several authors have succeeded in developing MCDA-based DSSs for the SSP, (partially) overcoming those barriers that prevented an early adoption of these systems (Derek 1993). A noticeable work that is comparable with ours is Dweiri et al. (2016), where the authors developed an integrated AHP-based DSS to solve an SSP arising in the automotive industry. In their real-world application, the AHP is applied to rank automotive suppliers in Pakistan, identifying four main criteria broken down into a number of subcriteria. The weights of criteria and subcriteria are computed using an AHP. The resulting DSS is then tested on a simplified case study involving three suppliers.

Another work that is worth considering is Bruno et al. (2016). Motivated by a collaboration with a prominent Italian leading company in the railway and transportation industry, the authors combined AHP and Fuzzy Set Theory (FST) to evaluate suppliers based on a detailed hierarchy of criteria drawing from both the literature and interviews with the company's management. In particular, AHP is used to compute the weights of criteria, while FST is used to measure the performance of suppliers. The proposed approach is tested on a case study involving four suppliers of a component, yielding valuable managerial insights on the transferability of MCDA-based approaches from academic theory to business practice.

Real-world applications of multi-criteria supplier selection incorporating both MCDA and multi-objective optimization are exemplified in the works of Hamdan and Cheaitou (2017) and Kellner et al. (2019). In Hamdan and Cheaitou (2017), the authors address a multi-period SSP using Fuzzy TOPSIS to rank suppliers. They employ AHP to define the relative importance of both traditional and green sets of criteria. A bi-objective formulation is then utilized to assign orders to suppliers. Similarly, Kellner et al. (2019) integrate Analytic Network Process and a multi-objective portfolio model to solve an SSP from a German automotive manufacturer. However, it is worth noting that both of these works, while providing valuable insights, are based on small numerical examples and do not consider the dynamic aspect of the problem.

Recently, MCDA-based approaches for multi-criteria supplier selection have also been proposed in less conventional applications, such as project scheduling (Nemati-Lafmejani and Davari-Ardakani 2021) and cyber security (Zhang et al. 2023), further confirming the wide applicability of these methods.

With respect to the reviewed literature, our work makes a number of contributions, namely:

- The specific SSP of *H2H Facility Solutions SpA* is formally defined as a multi-objective MILP model and solved both exactly and heuristically.
- Classical evaluation criteria, widely used in the SSP literature, and FM-specific

criteria were selected with the company to define a complex, multilevel tree of criteria for evaluating suppliers in the context of GSPs.

- AHP is used as a multi-criteria decision-making method for computing the weights of criteria, and the results are compared with those obtained using a revised Simos' procedure (Figueira and Roy 2002). To simplify the surveying process that precedes the definition of the comparison matrices, our pairwise comparisons are based on a simplified 1-3 scale, which makes the proposed methodology easily applicable in practice.
- Real-world data were collected and stored in a large database of 158 suppliers and 12,412 contracts, which makes the problem particularly significant in terms of size compared to other real-world applications found in the SSP literature.
- A simulator based on a rolling horizon algorithm was developed to dynamically solve the problem over multiple periods.
- A web application with a user-friendly interface and interactive visual tools is proposed to favor the evaluation of suppliers and support the user in the decision-making process.
- All these elements were embedded into a modular DSS, specifically developed to solve the SSP of *H2H Facility Solutions SpA*, but designed to be easily replicated for other real-world SSPs.
- Extensive computational experiments were performed to demonstrate the effectiveness of the proposed methodology in solving large-scale real-world SSPs. In particular, three alternative configurations implemented in the simulator were tested: a *company* configuration that recreates and evaluates the choices made by the company, a *greedy* configuration based on a weighted utility function, and a *MILP-based* configuration that has at its core the aforementioned multi-objective MILP model. All configurations use a rolling horizon algorithm to perform the assignments of contracts to suppliers, so as to catch the dynamic nature of the problem.
- The proposed methodology is intended to inspire innovative decision-supporting modules to be integrated into enterprise information systems.

To the best of our knowledge, dynamic aspects are still little considered in the SSP literature, nor exists a similar strategic and operational tool. In our opinion, such a tool may be well-suited for addressing multi-period multi-objective SSPs, where a large number of suppliers must be dynamically evaluated based on a multitude of criteria, and alternative configurations or scenarios must be compared. Furthermore, given the emerging role of GSPs in many industries, our work represents a valuable real-world application involving MCDA and multi-objective optimization.

3. Problem Definition

An FM contract is service- and facility-related. Whenever the GSP formalizes a new contract with a customer, the contract is subcontracted to a qualified external supplier able to provide the required service within a predefined service-level agreement. Being fixed a priori, cost is independent of the solution of the SSP and is thus not considered an objective in the problem definition nor a criterion in the multi-criteria evaluation.

Formally, given a set C of contracts and a set F of suppliers, the SSP in the context of GSPs is to subcontract contracts to suppliers with the multiple objectives of (i) maximizing the total quality score of selected suppliers, (ii) maximizing the total distance score, and (iii) minimizing the total penalty score due to the assignments of contracts that exceed a supplier's capacity.

For each supplier, we define a normalized quality score $S_f \in [0, 100]$ obtained by scaling a quality score s_f derived from the MCDA. Then, we define $D_{cf} \in [0, 100]$ as a normalized distance score derived from the geographical distance d_{cf} between the facility of customer to which contract c relates and the branch of supplier f providing the FM service. In Sections 4 and 5, we describe in detail how S_f and D_{cf} are computed. Further, we define q_f as the capacity of supplier f. This value represents an ideal number of contracts that the supplier can serve at the same time. Higher numbers are allowed but penalized. Additional details on how q_f is computed are provided in Section 5.2.

Let x_{cf} be a binary variable that takes the value 1 if contract c is subcontracted to supplier f and 0 otherwise, and let $y_f = \max\{\sum_{c\in C} x_{cf} - q_f, 0\}$ be a continuous variable reporting the number of contracts assigned over the capacity of supplier f, if any. The SSP can then be modeled as follows:

$$\text{(SSP)} \qquad \max z_{(\text{SSP})} = \left(\sum_{c\in C} \sum_{f\in F} S_f x_{cf}; \sum_{c\in C} \sum_{f\in F} D_{cf} x_{cf}; -\sum_{f\in F} y_f \right) \tag{1}$$

subject to

$$\sum_{f\in F} x_{cf} = 1 \qquad c \in C \tag{2}$$

$$\sum_{c\in C} x_{cf} \leq q_f + y_f \qquad f \in F \tag{3}$$

$$x_{cf} \in \{0, 1\} \qquad c \in C, f \in F \tag{4}$$

$$y_f \geq 0 \qquad f \in F \tag{5}$$

The objective function (1) maximizes the total quality score of selected suppliers and the total distance score, and minimizes the total penalty score for contracts assigned over suppliers' capacity. Constraints (2) impose that each contract c must be assigned to exactly one supplier, while constraints (3) are soft capacity constraints that link variables x_{cf} and y_f. Finally, constraints (4) and (5) define the domain of the variables. Note that an independent SSP is solved for each category of service, as capacity q_f varies depending on the category of service. This aspect could have been highlighted using an additional index in the mathematical model, but we decided to omit it for better readability.

The minimization of overcapacity is particularly important to guarantee the assignment of contracts to several suppliers, instead of using always the same ones. In addition, it is worth mentioning that the rolling horizon algorithm proposed in Section 5 adds a dynamic aspect to the problem, which is the daily update of quality score s_f (and, consequently, S_f) due to the assignment of new contracts to suppliers. This dynamic evaluation is intended to avoid the issue of assigning most of the contracts to the same few suppliers, with the negative effects of saturating their capacity and gradually reducing their performance in the long-term.

4. Multiple Criteria Evaluation

In this section, we describe how the quality score s_f is evaluated for each supplier $f \in F$. We obtain this value as the solution of an underlying *multiple criteria group decision* problem, which involves a plurality of decision makers as well as a hierarchy of criteria. Due to the large number of criteria and alternatives, this problem bears a strong resemblance to the computation of a *composite index*. Composite indexes are a powerful and widespread tool for obtaining a numerical synthesis of multiple assessments from different perspectives. The European Commission created the Competence Centre on Composite Indicators and Scoreboards (COIN) (European Commission 2002) to provide guidelines and tools for building robust composite indexes. Similarly, the United Nations Environment Programme developed the Sustainability Assessment of Technologies (SAT) Methodology (United Nations Environment Programme 2012) to support the assessment process in the context of sustainable development.

Taking into consideration the features mentioned above, we solved the multiple group decision problem by partitioning it into two distinct subproblems. In the first one, we aggregate the evaluations issued by several experts to assign a weight to all the indicators in the hierarchy of criteria. In the second subproblem, we exploit the criteria weights to compute a quality score for each alternative; in this phase, we apply a *weighted sum* method (see, e.g., Pomerol and Barba-Romero 2000, Ch. 4.1) where we preprocess the suppliers' evaluations according to the guidelines recommended for composite indexes.

In the following subsections, we show how to derive the quality score s_f. In particular, we (i) build the criteria hierarchy, (ii) assess the criteria weights, and (iii) compute the suppliers' quality scores.

4.1. Definition of the Criteria Hierarchy

The multiple criteria setting on which the supplier evaluation is based is the result of an analysis performed together with the company. This analysis was conducted through several rounds of interviews, which led to the definition of the multi-level tree of criteria reported in Figure 1.

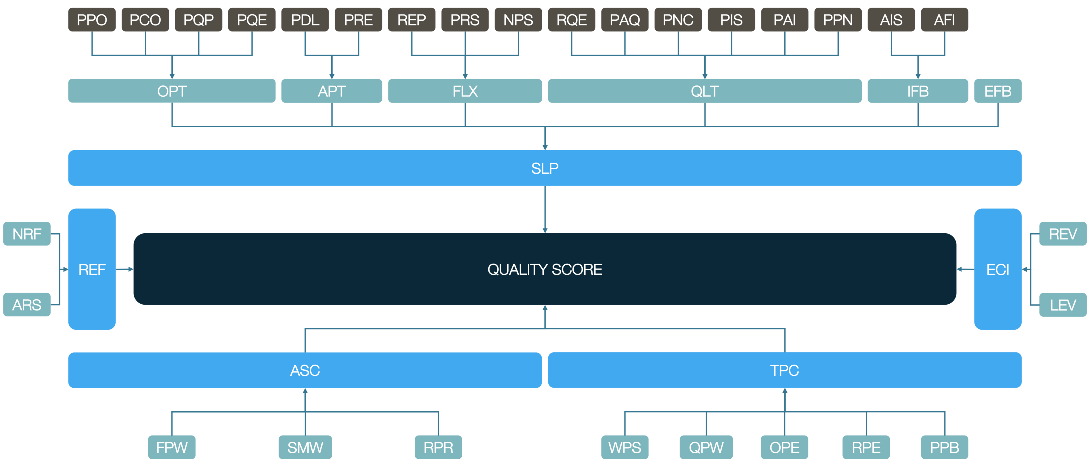


Figure 1. Tree of criteria

The tree contains three levels of criteria. The first level is composed by *macro* criteria, which directly contribute to define the quality score s_f for each supplier f. The first level is broken down into a second level of *micro* criteria, which, in a few cases, are further split into a third level of *nano* criteria. We selected five macro criteria that describe the main dimensions of supplier evaluation in the context of GSPs:

- economic indicators (ECI)

- technical and professional capability (TPC)
- additional saturation capacity (ASC)
- service level performance (SLP)
- references (REF).

We here describe only the five macro criteria, whereas we refer to the on-line Appendix A.1 for the micro and nano criteria. The ECI gives an evaluation of the economic soundness of suppliers, based on the last year's financial statements. The TPC assesses the organizational structure, the competencies and the outreach of suppliers along the territory. The ASC gives the residual capacity of suppliers; namely, this macro criterion indicates whether a supplier can accept new contracts. The SLP carefully evaluates suppliers across several performance indicators, on the basis of historical data. The REF is particularly important for qualifying suppliers, since it is based on references from customers with whom suppliers have already worked.

4.2. Assessing Criteria Weights

The weight of each item in the multi-level tree of criteria is computed as follows: at the *first* level, the weights of macro criteria are determined; at the *second* level, the weights of the related micro criteria are determined for each macro criterion; at the *third* level, the weights of the related nano criteria (if defined) are determined for each micro criterion.

Note that the sum of macro criteria weights must be equal to one, and the same holds for the sum of micro (respectively, nano) criteria weights for each macro (respectively, micro) criterion. The weight assessment was performed following a rather conservative approach: we employed a simplified AHP procedure and validated the set of weights by calculating a distinct set of weights with the revised Simos' procedure. As a matter of fact, the results obtained in the two cases turned out to be remarkably similar, although some small fluctuations in the values were detected.

Overall, the weights obtained with the AHP procedure, which we used in our computational experiments, can be considered sufficiently robust. In this section, we describe in detail their computation, whereas we refer to Appendix A.2 for the weights obtained with the revised Simos' procedure.

Criteria weights are the result of a group decision procedure, where the answers from 20 decision makers at the GSP were collected through an online survey. For each participant, a simplified AHP was performed, based on the three levels of the *reduced scale* reported in Table 1. The rationale behind the use of a reduced scale, instead of the fundamental scale by Saaty, is to simplify the collection of pairwise judgments, possibly reducing inconsistencies[1].

Table 1. Reduced scale

Relative Importance	Strongly less	Moderately less	Equal	Moderately more	Strongly more
Comparison Value	1/5	1/3	1	3	5

The participants were asked to use the reduced scale to answer standard questions like "*What is the relative importance of criterion A compared to criterion B?*". For each

[1]Note that the use of small size evaluation scales is a rather common practice in the computation of composite indexes. For example, a three-level scale was adopted (within a weighting procedure simpler than ours) for the 2016 European Digital City Index (Bannerjee et al. 2016).

participant and for each level of the criteria hierarchy, pairwise comparison judgments were converted into numerical values and recorded in a reciprocal comparison matrix A. Let n be the number of criteria. Each entry a_{ij} of A gives the comparison value of criterion i with respect to criterion j for all $i, j = 1, ..., n$. In addition, $a_{ji} = 1/a_{ij}$ for all i, j, and $a_{ii} = 1$ for all i. Given the comparison matrix, the corresponding vector of weights p was derived by applying the so-called "mean of row" method (see, e.g., Ishizaka and Labib 2011), which uses three steps:

(1) Sum the elements of each column j: $S_j = \sum_{i=1}^{n} a_{ij} \; \forall \; j$.
(2) Divide each element a_{ij} by the relative column sum: $a'_{ij} = a_{ij}/S_j \; \forall \; i, j$.
(3) Compute the mean of each row i: $p_i = \sum_{j=1}^{n} a'_{ij}/n \; \forall \; i$.

Finally, for each node in the criteria tree the *aggregated* weight was obtained by computing the geometric mean of the weights assigned by all the participants (the geometric mean is the standard aggregation method in group decision making contexts like ours; see, e.g., Aczél and Saaty 1983). The resulting weights, denoted by $\overline{p}_i$ for each macro criteria i, are reported in Table 2.

Table 2. Aggregated weights for macro criteria using the AHP

i	ECI	TPC	ASC	SLP	REF
$\overline{p}_i$	0.1527	0.2672	0.1794	0.2394	0.1614

4.3. Computing the Scores

Once determined the weight of each criterion, we applied a weighted sum method to compute the score of each supplier. The method consists of three steps: statistical treatment of outliers (*winsorization*); normalization; and aggregation.

Let us denote by I the set of nodes in the criteria tree (i.e., macro, micro, and nano criteria) and by $L \subset I$ the set of leaves of the criteria tree. The set L contains the first-order criteria, that is, the nano criteria for macro criterion SLP, the micro criterion EFB of SLP, and the micro criteria for the other macro criteria (see again Figure 1). For each supplier $f \in F$ and each leaf $i \in L$ we are given an evaluation e_{if}, expressed on a criterion-specific cardinal scale. In what follows, we describe the above three steps separately and finally discuss some theoretical properties of the resulting weighted sum procedure.

Outliers detection is quite relevant in our context, where a small number of suppliers may be characterized by uncommon features. Outliers are detected applying a rather simple *box plot* method. Given a criterion $i \in L$, we find values Q_1 and Q_3 of the first and third quartiles of the evaluations e_{if}, respectively; we compute the Inter Quantile Range IQR $= Q_3 - Q_1$; then, we define the lower threshold $T^l = Q_1 - 3 \cdot \text{IQR}$ and the upper threshold $T^u = Q_3 + 3 \cdot \text{IQR}$. A value e_{if} larger than T^u or smaller than T^l is identified as an outlier for criterion i. Note that, in many cases, we have $T^l < 0$, while the evaluations are restricted to non-negative values. Outliers are then treated by applying the following winsorization process:

- each evaluation $e_{if} > T^u$ is replaced by the value $e_i^{\max} = \max\{e_{if} : f \in F, \, e_{if} \leq T^u\}$
- each evaluation $e_{if} < T^l$ is replaced by zero.

This rather simple treatment of outliers is sufficient in our context, but clearly more sophisticated methods exist. For example, an iterative process based on higher moments is suggested by COIN (European Commission 2002; see, in particular, Nardo et al. 2020).

In this phase, each evaluation e_{if} for $i \in L$ is mapped onto a *normalized evaluation* $E_{if} \in [0, 1]$. For every criterion i, we distinguish between *direct* and *reverse* normalization:

- For a maximization criterion i (i.e., the better f, the greater e_{if}) direct normalization gives $E_{if} = e_{if}/e_i^{\max}$, where, after winsorization, we have $e_i^{\max} = \max\{e_{if} : f \in F\}$. Note that outliers previously falling over the upper threshold T^u take value $E_{if} = 1$.
- For a minimization criterion i (i.e., the better f, the smaller e_{if}) reverse normalization gives $E_{if} = 1 - e_{if}/e_i^{\max}$; outliers previously falling below T^l take value $E_{if} = 1$.

As a result, the better f, the greater E_{if}.

Since $e_{if} \geq 0$ in our context, we have $e_i^{\max} = \|e_{i\cdot}\|_\infty$, where $e_{i\cdot} \in \mathbb{R}^{|F|}$ is the vector of evaluations for criterion i. Normalization based on the infinity norm has been often advocated in MCDA, together with other norms such as $\|\cdot\|_1$ and $\|\cdot\|_2$; here $\|\cdot\|_\infty$ was chosen also because it is not sensitive to the number of outliers. Note that E_{if} does not necessarily attain the extremes of the interval $[0, 1]$ for each criterion, because it may be $\min_f E_{if} > 0$ for direct normalization and $\max_f E_{if} < 1$ for reverse normalization. This fact is acceptable in our context, even though it could be prevented by a more complex normalization step (see, e.g., Pomerol and Barba-Romero 2000, Ch. 4.1).

The aggregation phase can be seen as a three-step bottom-up recursive process. At the first step, for each micro criterion i of SLP (the only macro criterion divided up to nano criteria), except EFB, we compute

$$E_{if} = \sum_{j \in S_i} \overline{p}_j E_{jf} \qquad \forall f \in F \tag{6}$$

where S_i is the set of nano criteria for i, and $\overline{p}_j$ is the weight of nano criterion j. Since we have $\sum_{j \in S_i} \overline{p}_j = 1$, it follows that each value E_{if} is normalized between 0 and 1. Thus, at the end of this step, we have a normalized evaluation E_{if} for each $f \in F$ and each micro criterion i.

In the second step, we define for each macro criterion i the values E_{if} as in (6), where, in this case, S_i is the set of micro criteria for i and $\overline{p}_j$ is the weight of micro criterion j. Again, at the end of this step, we have a normalized evaluation E_{if} for each $f \in F$ and each macro criterion i.

In the last step, we obtain the quality score s_f for each $f \in F$ as

$$s_f = \sum_{i \in M} \overline{p}_i E_{if} \qquad \forall f \in F$$

where $M = \{\mathrm{ECI}, \mathrm{TPC}, \mathrm{ASC}, \mathrm{SLP}, \mathrm{REF}\}$ denotes the set of macro criteria and $\overline{p}_i$ is the weight of $i \in M$.

Due to the winsorization and normalization steps, weighted sum lacks the *independence* property, that is, a quality score s_f is not uniquely determined by the evaluations e_{if}, but depends on the evaluations of the whole set of suppliers F. This implies, in particular, that our scores are exposed to *rank reversal*: given two suppliers $f, g \in F$, their relative ranking as determined by s_f and s_g may be reversed if another supplier is added to or removed from F, or if its evaluations change. The rank reversal phenomenon is almost ubiquitous (and often debated) in MCDA methods; see, for example, Wang and Luo (2009), Garcia-Cascales and Lamata (2012) for discussion, explicative examples, and further references.

Observe that our scores have been conceived to be computed repeatedly throughout a wide time horizon, during which the set F and the suppliers' evaluations are assumed to evolve. Thus, we may question the *stability* of our scores over time: a similar issue has been discussed in Pretolani (2020) for the SDEWES Index (SDEWES Centre 2016), a composite index that resembles our scores in many aspects. As shown in Pretolani (2020), the combination of winsorization and normalization may occasionally lead to rather unexpected outcomes. However, stability is not a very significant issue in our context, for at least two reasons. First of all, the actual occurrence of rank reversals is rather unlikely, also due to winsorization. Most importantly, our scores should not be considered as an absolute measure of the "quality" of a supplier, but, rather, as a relative measure of attractiveness with respect to a particular time instant. This aspect is further clarified in the next section, which presents our DSS for supplier selection.

5. DSS Implementation

The DSS consists of three main modules: a relational database, a supplier evaluator, and a simulator. The overall DSS architecture is depicted in Figure 2. In the following, a description of the first and the third modules is provided. We omit to describe the second module, which was coded in C++ and simply contains the methodology presented in Section 4 to evaluate the quality score of the suppliers. An additional user-friendly interface, for the sake of conciseness reported in Appendix A.3, allows the decision maker to easily interact with the system and make use of simple visual tools.

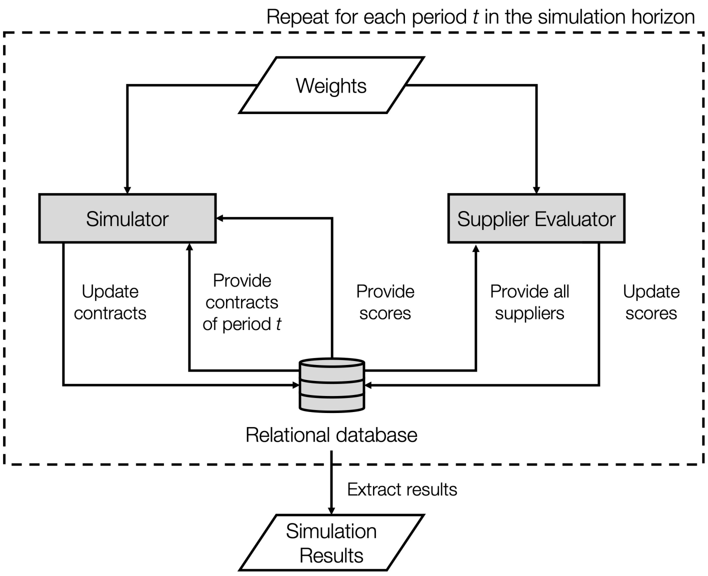


Figure 2. DSS architecture

5.1. Database Module

The first module is a MySQL relational database that stores data regarding all suppliers available to the company and all the necessary information about contracts.

To fill the database, we first identified a sample of suppliers together with the company. An online survey was sent to this sample to collect several data on their organizational structure (e.g., headquarters, branches, number of workers, number of office workers), technical capabilities (e.g., categories of service offered to customers, type of qualifications, number of qualifications per type), and economic soundness (e.g., revenue from last year's financial statement). In this way, we collected data for 158 suppliers.

Then, we obtained from the company historical data on 12,412 contracts assigned over multiple years. Each contract is associated with a single category of service required by a customer for a specific facility, and it has a planned duration and a fixed cost. In addition, we obtained from the company detailed performance data registered on all contracts. After performing a preliminary analysis on the database, we identified 20 categories of service: air-conditioning and heating systems, alarm systems and security, automatisms, cleaning services, construction, consulting and support services, deratting and disinfestation, electrical systems, elevator maintenance, elevator systems, facility management, fire protection systems, furniture and equipment, green services, mechanical systems, porter services, reception desk services, special systems, technological presidium, and water supply systems.

5.2. Simulator Module

The third module is a simulator that performs the assignment of contracts to suppliers. In particular, three alternative configurations were implemented in this module: a *company* configuration, a *greedy* configuration and a *MILP-based* configuration. All these configurations can be selected in a rolling horizon algorithm. The simulator was coded in C++ and CPLEX 12.9 was used as MILP solver in the *MILP-based* configuration.

The *company* configuration recreates and evaluates the choices made by the company during each period; such a configuration is used exclusively to set a benchmark for the other two configurations. The *greedy* configuration performs the assignments of contracts to suppliers based on a scalar objective function derived from the objective function of the multi-objective MILP model defined in Section 3. In the *MILP-based* configuration, the decision on the assignments of contracts to suppliers is guided by the same scalar objective function subject to constraints (2)-(5). The scalar objective function is

$$\max \zeta_{(\mathrm{SSP})} = \alpha \sum_{c\in C}\sum_{f\in F} S_f x_{cf} + (1-\alpha)\sum_{c\in C}\sum_{f\in F} D_{cf} x_{cf} - \beta \sum_{f\in F} y_f \tag{7}$$

where the normalized quality score of supplier f is expressed by $S_f = 100\cdot(s_f/s_{\max})$, given the quality score s_f and the maximum quality score $s_{\max}$, while $D_{cf} = 100\cdot(1-(d_{cf}/d_{\max}))$ identifies the normalized distance score, given the geographical distance d_{cf} and the maximum distance $d_{\max}$. Both S_f and D_{cf} are thus scaled in the interval $[0, 100]$.

The quality score and the distance score are multiplied by coefficients α and $(1-\alpha)$, where α, which lies in $[0, 1]$, controls the relative importance of these two terms of the objective function and can be customized by the decision maker. The third term is instead multiplied by coefficient β, which corresponds to the penalty generated by each contract assigned over capacity. In Section 6, we test several combinations of these coefficients.

Note that, once α and β have been fixed, the optimization problem underlying the *MILP-based* configuration becomes trivial, as it corresponds to a minimum cost flow problem with a single objective. Further, we stress that the *greedy* configuration solves the problem to optimality when $\beta = 0$, since the capacity of suppliers becomes irrelevant. Indeed, without penalty for contracts assigned over capacity, the soft capacity constraints of the MILP model are relaxed and both the *greedy* configuration and the *MILP-based* configuration solve the same problem of finding a minimum cost assignment for each contract. This is confirmed by the results reported in Section 6.3.

The three configurations were implemented and integrated in a rolling horizon algorithm, which was chosen to reproduce the dynamic nature of the problem. Indeed, the quality score of each supplier must be updated periodically based on newly assigned contracts, which partially or totally consume its capacity and thus affect the attractiveness of receiving new contracts. In particular, the rolling horizon algorithm decomposes the whole problem into narrower periods so that, for each period, the DSS is able to retrieve updated information from the database, recompute the quality score of each supplier, perform the assignments of contracts to suppliers, and store the results in the database. Note that, with this structure, we are able to run simulations for any period of time based on real-data from the company.

At the beginning of the simulation horizon, we compute the median of contracts per worker for each category of service. Then, for each category of service and for each supplier providing that particular category of service, the corresponding median of contracts per worker is multiplied by the number of workers of the supplier and the result is rounded up to the nearest integer number to obtain the initial capacity q_f for each supplier f. Then, this capacity is also dynamically updated during the simulation horizon due to the assignment of new contracts to suppliers.

In Algorithm 1, we report the pseudo-code of the proposed rolling horizon algorithm, where $t \in T$ is a period of λ days in the simulation horizon T.
For each period, the results are stored in the database, so that the scores are recomputed and updated accordingly whenever new contracts must be assigned. Because the evaluation of suppliers depends on the (past and present) information stored in the database, the score of a certain supplier may change over time. If new contracts are assigned to the same supplier, the supplier may become saturated, potentially reducing its score for future contracts. This dynamic aspect is particularly important because it reproduces a typical characteristic of the real problem. However, it is worth mentioning that these simulations are intended to evaluate the effectiveness of the proposed system and validate the results together with the company. In practice, the DSS is designed to process a large volume of data providing the decision-makers with the piece of information they need to make a knowledgeable decision.

6. Computational Evaluation

In this section, we present the results of the computational experiments performed to test the rolling horizon algorithm presented in Section 5.2. The experiments were run on a PC equipped with an Intel Core i5 dual-core CPU processor @ 2.70 GHz and 8 GB of RAM. We remind that the rolling horizon algorithm was coded in C++ and CPLEX 12.9 was used as MILP solver with the default configuration.

Algorithm 1 Rolling Horizon Algorithm

Require: $\alpha \in [0, 1]$, $\beta \geq 0$ (integer), λ, $|T|$ ▷ Set parameters

$Config \leftarrow$ `Select`$\{company, greedy, MILP\text{-}based\}$ ▷ Select a configuration
for $t \leftarrow 1$ **to** $|T|$ **do**
  Get subset $\overline{C} \subseteq C$ of contracts to assign during period t
  if $\overline{C} \neq \emptyset$ **then**
    for $f \leftarrow 1$ **to** $|F|$ **do**
      Update ASC macro criterion
      Recompute supplier quality score s_f
    end for
    Scale quality score S_f in $[0, 100]$
    for $c \leftarrow 1$ **to** $|\overline{C}|$ **do**
      for $f \leftarrow 1$ **to** $|F|$ **do**
        Evaluate the branch of supplier f having the shortest distance d_{cf}
      end for
    end for
    Scale distance score D_{cf} in $[0, 100]$
    if $Config = company$ **then** ▷ If the *company* configuration was selected
      Recreate and evaluate the choices made by the company during period t
    else if $Config = greedy$ **then** ▷ If the *greedy* configuration was selected
      Solve the SSP for subset $\overline{C}$ of contracts by maximizing (7)
    else ▷ If the *MILP-based* configuration was selected
      Solve the SSP for subset $\overline{C}$ of contracts by maximizing (7) subject to (2)-(5)
    end if
    Update the simulation statistics
  end if
end for

6.1. Experimental Data

To perform the computational experiments, we utilized all the data stored in the first module of the DSS. In particular, the company provided detailed information on customers (e.g., full name), their facilities (e.g., type, geographical location), and 12,412 contracts assigned over multiple years (e.g., facility to which the contract refers, supplier to whom the contract was assigned, category of service, start date, end date). These data were integrated with those collected from an online survey distributed to a sample of suppliers identified by the company. In detail, 158 suppliers replied to the survey providing detailed information on their headquarters (e.g., geographical location, number of workers, number of office workers), branches (e.g., geographical locations, number of workers, number of office workers), geographical areas in which they operate, categories of service offered to customers, qualifications, and revenues from last year's financial statement. Additionally, the company provided supplementary data on the historical performance of these 158 suppliers. Due to commercial restrictions, full details of these data cannot be publicly disclosed.

6.2. Parameter Setting and Dedicated Analysis on Coefficient β

We list below the parameters that must be given as an input to the rolling horizon algorithm:

- Coefficient $\alpha \in [0, 1]$, which controls the relative importance of quality score S_f. As a consequence of defining α, coefficient $(1 - \alpha)$, which controls the relative importance of distance score D_{cf}, is also automatically defined.

- Coefficient $\beta \geq 0$ (integer), which corresponds to the penalty generated by each contract assigned over capacity. A dedicated analysis to fine-tune this coefficient is reported below.
- Length λ of each period t in the simulation horizon, expressed in number of days.
- Number of periods $|T|$ in the simulation horizon, where $|T| = 3,287$ corresponds to the minimum number of periods (i.e., days in this case) to query all contracts from the database.

Using the rolling horizon algorithm with the *MILP-based* configuration and parameters $\alpha = 0.3$, $\lambda = 1$ and $|T| = 3,287$, we performed a dedicated analysis to evaluate the number of contracts assigned over capacity for different values of coefficient β. From the results reported in Figure 3, we see that the number of contracts over capacity reaches an asymptotic value when parameter β is between 70 and 75. For this reason, in the computational experiments, we decided to choose $\beta \in \{0, 10, 100\}$ to represent different scenarios in terms of penalty for over capacity.

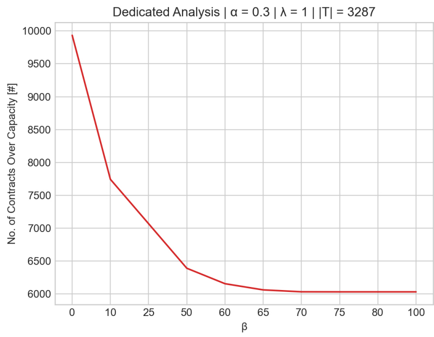


Figure 3. Dedicated analysis on coefficient β

6.3. Experimental Results

In this section, we illustrate the experimental results obtained while testing the three configurations (i.e., *company*, *greedy*, and *MILP-based*) of the rolling horizon algorithm. We remind that the *company* configuration, which recreates and evaluates the decisions made by the company over the simulation horizon, serves exclusively as a benchmark for the other two configurations.

For each value $\beta \in \{0, 10, 100\}$ we are left with a bi-objective problem involving the quality and distance scores. Computing the entire Pareto front for this problem would be computationally cumbersome and not really relevant to our analysis. Instead, we generated an approximation of the Pareto set, obtained by setting $\alpha \in \{0.0, 0.1, 0.2, \ldots, 1.0\}$.

Regarding the other parameters, we set $\lambda = 1$ and $|T| = 3,287$. In other words, we solved a *daily* SSP for all the days in the simulation horizon. Additional experiments with $\lambda = 7$ (i.e., solving a *weekly* SSP) and $\lambda = 30$ (i.e., solving a *monthly* SSP) were performed but no significant differences were noticed. This may be due to the particular structure of the real data used for the experiments. Also, note that the distances between the facilities and the suppliers were evaluated using the classical *haversine formula*.

Next, we report the experimental results of three alternative scenarios: a first scenario without penalty for contracts assigned over capacity ($\beta = 0$), a second scenario with a limited penalty for contracts assigned over capacity ($\beta = 10$), and a third scenario with a high penalty for contracts assigned over capacity ($\beta = 100$).

The results that were obtained for the experiments with $\beta = 0$ are reported in Table 3, where columns "α" and "$(1 - \alpha)$" give the relative weights of the quality score and the distance score, respectively; column "Obj." gives for each configuration of the rolling horizon algorithm the objective function value; column "Sec." gives for the *greedy* configuration and the *MILP-based* configuration the total computing time for running the algorithm, without considering the time spent performing database operations, which is independent of the algorithm; and columns "$\%\text{gap}_{company-greedy}$" and "$\%\text{gap}_{greedy-MILP\text{-}based}$" give the percentage gap between the *company* configuration and the *greedy* configuration and the percentage gap between the *greedy* configuration and the *MILP-based* configuration, respectively. In this scenario, the *greedy* configuration and the *MILP-based* configuration obtained the same results, because, as noticed in Section 5, they solve the same problem when $\beta = 0$.

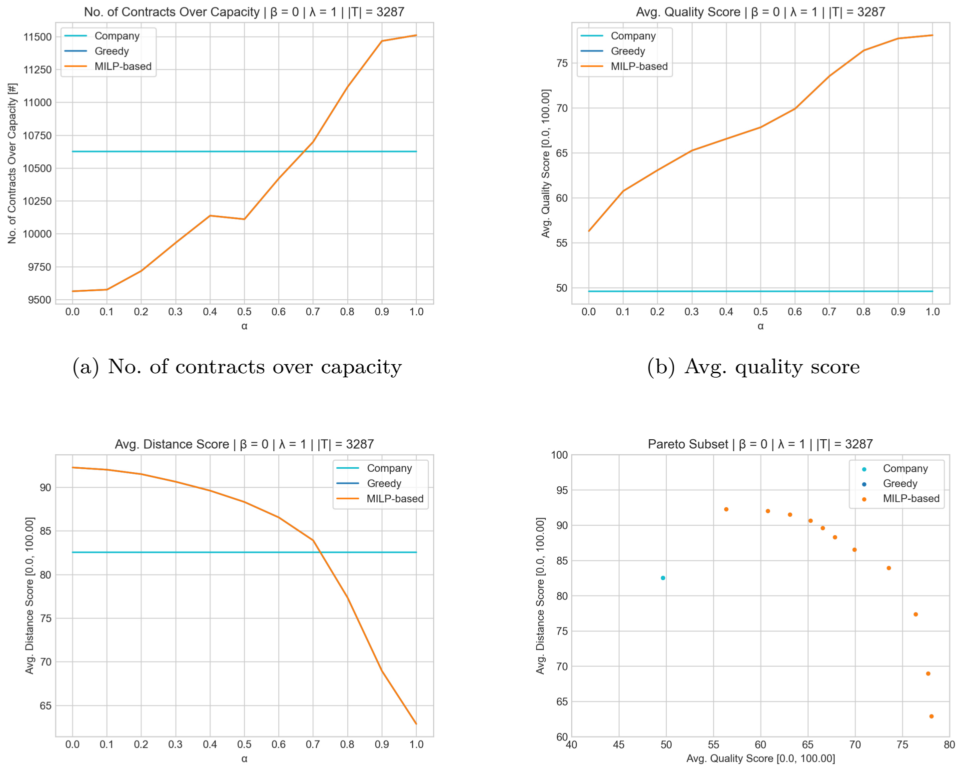
From the results reported in Table 3, we see that, on average, the *greedy* configuration and the *MILP-based* configuration improved the result of the *company* configuration by 25%. This behavior is more evident for higher values of α, suggesting that the *company* configuration tends to favor proximity towards quality. It is worth highlighting that for the *company* configuration, we obtained always the same solution for all α, having a total quality score of 616,191 and a total distance score of 1,024,615. Indeed, the values reported in Table 3 are obtained as a weighted sum of these two values. As for the total computing time, we see that, on average, the *MILP-based* configuration takes slightly longer than the *greedy* configuration. This is understandable since the assignments of contracts to suppliers is driven by the MILP model. In any case, the overall difference in terms of computing time is not significant and both configurations are very fast in solving each single period.

Table 3. Objective function values and computing times for $\beta = 0$. Best values in **boldface**

α	$(1-\alpha)$	*company* Obj.	*greedy* Obj.	*greedy* Sec.	*MILP-based* Obj.	*MILP-based* Sec.	$\%\text{gap}_{company-greedy}$	$\%\text{gap}_{greedy-MILP\text{-}based}$
0.0	1.0	1,024,615	**1,145,267**	3.27	**1,145,267**	16.26	11.8	0.0
0.1	0.9	983,772	**1,103,407**	3.75	**1,103,407**	14.58	12.2	0.0
0.2	0.8	942,930	**1,065,234**	3.05	**1,065,234**	16.38	13.0	0.0
0.3	0.7	902,088	**1,030,535**	3.47	**1,030,535**	16.46	14.2	0.0
0.4	0.6	861,245	**997,898**	3.37	**997,898**	14.46	15.9	0.0
0.5	0.5	820,403	**969,193**	2.93	**969,193**	14.49	18.1	0.0
0.6	0.4	779,560	**950,321**	3.23	**950,321**	14.46	21.9	0.0
0.7	0.3	738,718	**951,462**	3.03	**951,462**	14.32	28.8	0.0
0.8	0.2	697,876	**950,596**	2.74	**950,596**	13.00	36.2	0.0
0.9	0.1	657,033	**953,763**	3.01	**953,763**	14.20	45.2	0.0
1.0	0.0	616,191	**969,082**	3.33	**969,082**	13.71	57.3	0.0
avg		820,403	**1,007,887**	3.20	**1,007,887**	14.76	25.0	0.0

Additional results are reported in Figure 4. In Figure 4-(a), we plot the number of contracts over capacity. Here, we observe that, for $\alpha < 0.7$, the *greedy* configuration and the *MILP-based* configuration assigned fewer contracts over capacity than the *company* configuration. The opposite holds for $\alpha \geq 0.7$. In Figure 4-(b), we plot the average quality score values. On average, the *greedy* configuration and the *MILP-based* configuration outperformed the *company* configuration by 38.4%. In Figure 4-(c), we plot the average distance score values. Here, we observe that the *greedy* configuration and the *MILP-based* configuration obtained higher average distance score values than the *company* configuration for $\alpha < 0.8$. Finally, the graphical representation of the

Pareto subsets is reported in Figure 4-(d). Here, on the x-axis we report the average quality score, while on the y-axis we report the average distance score. In this scenario, we notice that the solution obtained by the *company* configuration is dominated for $\alpha < 0.8$, since both the *greedy* configuration and the *MILP-based* configuration show higher average quality score and average distance score. This does not hold for $\alpha \geq 0.8$, since the *company* configuration obtained a higher average distance score. Additionally, the points corresponding to the solution values found by the *greedy* configuration and the *MILP-based* configuration overlap.



(a) No. of contracts over capacity

(b) Avg. quality score

(c) Avg. distance score

(d) Pareto subset

Figure 4. Experimental results for $\beta = 0$

The same experiments were repeated for $\beta = 10$ and the results that were obtained are reported in Table 4. We notice that, in line with the results found in the previous scenario, the objective function value of the *greedy* configuration is constantly higher than the objective function value of the *company* configuration, while the objective function value of the *MILP-based* configuration is higher than the objective function value of the *greedy* configuration by only 0.2%. This indicates, on one hand, that with a limited penalty on over capacity there is still a remarkable difference between the *greedy* configuration and the *company* configuration, and, on the other hand, that the *greedy* configuration and the *MILP-based* configuration perform similarly for $\beta = 10$. It is worth noting that the *greedy* configuration slightly outperforms the *MILP-based* configuration when $\alpha = 0.9$. This is not surprising and is due to the dynamic nature of the problem, in which the optimal assignments of contracts in one period may produce worse results in the future. The total computing times are in line with those obtained in the previous scenario.

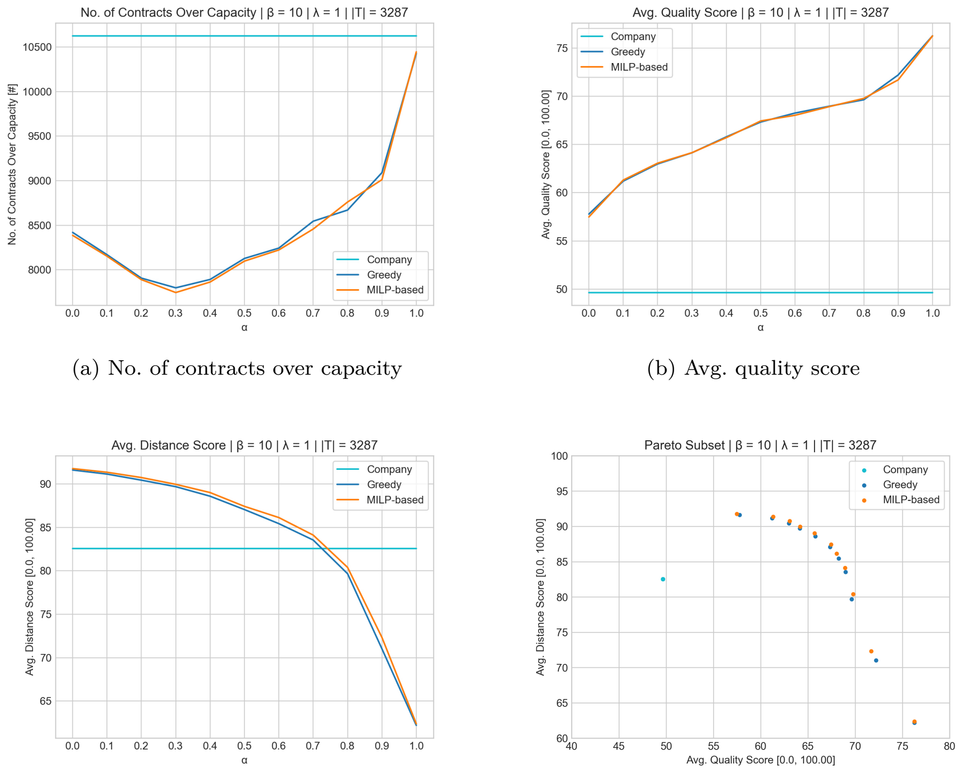
Table 4. Objective function values and computing times for $\beta = 10$. Best values in **boldface**

α	$(1-\alpha)$	*company* Obj.	*greedy* Obj.	Sec.	*MILP-based* Obj.	Sec.	%gap$_{company-greedy}$	%gap$_{greedy-MILP\text{-}based}$
0.0	1.0	918,355	1,052,938	2.95	**1,055,364**	15.40	14.7	0.2
0.1	0.9	877,512	1,012,479	3.20	**1,015,138**	15.21	15.4	0.3
0.2	0.8	836,670	975,412	3.22	**978,717**	15.58	16.6	0.3
0.3	0.7	795,828	940,162	3.23	**943,141**	15.22	18.1	0.3
0.4	0.6	754,985	907,508	2.97	**910,560**	14.81	20.2	0.3
0.5	0.5	714,143	876,919	2.90	**880,117**	15.10	22.8	0.4
0.6	0.4	673,300	850,154	3.22	**852,057**	15.68	26.3	0.2
0.7	0.3	632,458	824,887	3.03	**827,507**	15.31	30.4	0.3
0.8	0.2	591,616	802,538	3.62	**804,898**	14.92	35.7	0.3
0.9	0.1	550,773	**803,928**	3.63	800,594	15.21	46.0	−0.4
1.0	0.0	509,931	842,442	3.73	**842,058**	15.74	65.2	0.0
avg		714,143	899,033	3.25	**900,923**	15.29	28.3	0.2

Additional results are reported in Figure 5. In Figure 5-(a), where we plot the number of contracts over capacity, we observe that the *greedy* configuration obtained constantly better results than the *company* configuration in terms of number of contracts over capacity (with an average reduction of 20.2%). Then, the *MILP-based* configuration obtained a further reduction of 0.3% if compared with the *greedy* configuration. This indicates that both the *greedy* configuration and the *MILP-based* configuration show a similar potential in managing the available capacity of suppliers. Indeed, the two curves almost overlap. In Figure 5-(b), where we plot the average quality score values, we see that the results are in line with those found in the previous scenario, thus indicating that the average quality score obtained by the *greedy* configuration and the *MILP-based* configuration remained steady, despite the number of contracts assigned over capacity was significantly decreased. In addition, it is worth noting that the *greedy* configuration performed slightly better than the *MILP-based* configuration on this indicator. In Figure 5-(c), we plot the average distance score values. As before, the *company* configuration obtained a higher distance score for higher values of α and the *MILP-based* configuration performed constantly better than the *greedy* configuration. Finally, the graphical representation of the Pareto subsets is reported in Figure 5-(d). In this scenario, the Pareto subsets of the *MILP-based* configuration are slightly above the Pareto subsets of the *greedy* configuration.

The results of the experiments performed for $\beta = 100$ are reported in Table 5. We observe that the objective function values for the *company* configuration are negative; this is due to the predominant effect of the penalty score. As for the gap between the *greedy* configuration and the *MILP-based* configuration, on average, the latter outperformed the former by 4.7%, which means that with a high penalty on overcapacity, the advantage of using the *MILP-based* configuration becomes more evident. The computing times are comparable to those obtained in the previous scenarios, with a slight increase for the *MILP-based* configuration due to the effect of the higher penalty for overcapacity assignments.

Additional results are reported in Figure 6. In Figure 6-(a), the *greedy* configuration significantly improved the results of the previous scenario. Indeed, the average reduction in the number of contracts over capacity increased to 43.1%, if compared with the *company* configuration. Again, the *MILP-based* configuration shows an additional average reduction of 0.8%. This confirms the potential in managing the available capacity of suppliers shown by both the *greedy* configuration and the *MILP-based* configuration, especially



(a) No. of contracts over capacity

(b) Avg. quality score

(c) Avg. distance score

(d) Pareto subset

Figure 5. Experimental results for $\beta = 10$

in those contexts in which the assignment of contracts over capacity is particularly penalizing. In Figure 6-(b), the results on the average quality score values got worse, if compared with the previous scenario. So, we may conclude that the great reduction in terms of number of contracts over capacity is accompanied by a decrease in terms of average quality score. Again, the *greedy* configuration performed slightly better than the *MILP-based* configuration on this indicator. In Figure 6-(c), the *company* configuration outperformed both the *greedy* configuration and the *MILP-based* configuration in terms of average distance score. This means that, in this scenario, the noteworthy improvement in the number of contracts over capacity was obtained at the expense of quality and proximity. Finally, the graphical representation of the Pareto subsets is reported in Figure 6-(d). Differently from the previous scenarios, here it is worth noticing that the solution obtained by the *company* configuration is never dominated by the solutions obtained by the *greedy* configuration.

6.4. Additional Analysis in Case of Unexpected or Disruptive Events

The computational experiments performed in the previous sections were all deterministic in the sense that we used historical data and did not account for unexpected or disruptive events that could have changed the outcome. However, uncertainty is an intrinsic feature of real-world SSPs. In particular, most suppliers often collaborate with other companies and get additional contracts from them. This may lead to unexpected events, as their

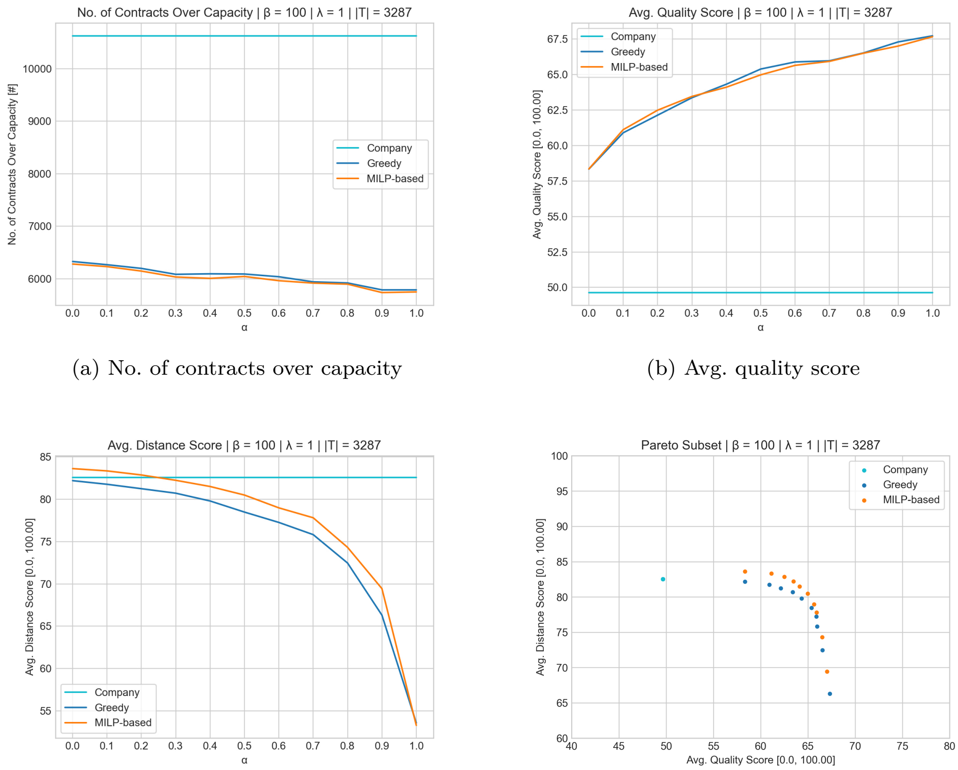
Table 5. Objective function values and computing times for $\beta = 100$. Best values in **boldface**

α	$(1-\alpha)$	*company* Obj.	*greedy* Obj.	*greedy* Sec.	*MILP-based* Obj.	*MILP-based* Sec.	%gap$_{company-greedy}$	%gap$_{greedy-MILP\text{-}based}$
0.0	1.0	−37,985	387,276	3.63	**409,646**	22.60	n/a	5.8
0.1	0.9	−78,828	362,090	3.85	**383,584**	22.28	n/a	5.9
0.2	0.8	−119,670	341,156	3.74	**363,213**	20.82	n/a	6.5
0.3	0.7	−160,512	328,754	3.02	**347,393**	21.07	n/a	5.7
0.4	0.6	−201,355	304,095	3.21	**324,492**	20.64	n/a	6.7
0.5	0.5	−242,197	283,767	3.66	**298,414**	21.60	n/a	5.2
0.6	0.4	−283,040	270,371	3.88	**284,656**	26.86	n/a	5.3
0.7	0.3	−323,882	261,322	3.36	**270,801**	26.64	n/a	3.6
0.8	0.2	−364,724	248,492	3.27	**255,195**	16.50	n/a	2.7
0.9	0.1	−405,567	255,339	3.60	**261,031**	19.24	n/a	2.2
1.0	0.0	−446,409	260,632	3.35	**265,015**	19.29	n/a	1.7
avg		−242,197	300,299	3.51	**314,858**	21.59	n/a	4.7

capacity may suddenly be reduced to the point where they are unavailable for some periods. Disruptive events, such as the unavailability of a supplier for long periods due to external causes (e.g., floods, earthquakes), are also possible, although less frequent. Accurately simulating unexpected or disruptive events is complex since the company does not have access to always up-to-date and detailed information on the activities of all suppliers. A regular update of this information could be done, but besides the additional effort and cost, this would most likely not solve the problem as new contracts are continuously assigned.

To test the robustness of the rolling horizon algorithm, we performed two additional experiments in which, for each period of the simulation horizon, each supplier has a probability γ of becoming unavailable. The duration of the unavailability may last one, three, or six months, with relative probabilities of 60, 30 and 10%, respectively. These experiments were performed using the rolling horizon algorithm with the *greedy* configuration and the *MILP-based* configuration and parameters $\alpha \in \{0.4, 0.6\}$, $\beta = 10$, $\lambda = 1$ and $|T| = 3,287$. We set $\gamma \in \{0.01, 0.05, 0.10\}$ and, for each value of γ, the rolling horizon algorithm was run 30 times. In this way, the objective is to evaluate both solutions in which more importance is given to the distance score (i.e., when $\alpha = 0.4$) and the penalty for contracts assigned over capacity is limited (which might be possible in practice), and solutions in which less importance is given to the distance score (i.e., when $\alpha = 0.6$) and the penalty for contracts assigned over capacity is equally limited. The results obtained are reported in Tables 6 and 7, where columns "γ" and "#Disrupt. Events" give the disruption probability and the average number of disruptive events that occur over the simulation horizon, respectively; column "Obj." gives for the *greedy* configuration and the *MILP-based* configuration the average objective function value; column "%gap" gives for the *greedy* configuration and the *MILP-based* configuration the average percentage gap computed as $100 \cdot (z - \overline{z})/z$, where z is the objective function value obtained in the deterministic case (i.e., when $\gamma = 0.00$) and $\overline{z}$ is the average objective function value obtained over the 30 runs of the algorithm; columns "Qual. Score" and "Dist. Score" specify for each configuration the average quality score and the average distance score, respectively.

Interestingly, the results confirm the robustness of the rolling horizon algorithm, that in the worst case obtains a 0.55% gap on the objective function. Not surprisingly, the average objective function values decrease as the disruption probability increases. Furthermore, it is worth highlighting that the results obtained by the *MILP-based*



(a) No. of contracts over capacity

(b) Avg. quality score

(c) Avg. distance score

(d) Pareto subset

Figure 6. Experimental results for $\beta = 100$

configuration are more robust than those obtained by the *greedy* configuration. Indeed, the objective function value and the average percentage gap to the deterministic case (i.e., without disruptive events) are always better for the former. Also, the results show that the *MILP-based* configuration is more sensitive than the *greedy* configuration to the relative weights set by parameter α. Specifically, the *MILP-based* configuration obtains significantly smaller gaps for $\alpha = 0.6$ than for $\alpha = 0.4$; on the contrary, the *greedy* configuration obtains essentially the same gaps in the two cases. This seems to suggest that the *MILP-based* configuration obtains better results when the relative importance of the quality score is higher. This does not mean, however, that the quality score should be prioritized when disruptive events are likely to occur, also because the resulting improvement is expected to be rather limited. Overall, our experiments suggest that the possibility of disruptive events should not have an impact on the choice of a strategic parameter like α.

Table 6. Additional analysis for $\alpha = 0.4$ and $\beta = 10$

		greedy				*MILP-based*			
γ	#Disrupt. Events	Obj.	%gap	Qual. Score	Dist. Score	Obj.	%gap	Qual. Score	Dist. Score
0.00	0	907,507.65	0.00	65.79	88.60	910,560.42	0.00	65.71	89.02
0.01	20.70	906,561.75	0.10	65.74	88.49	909,840.69	0.08	65.71	88.91
0.05	103.17	904,571.52	0.32	65.59	88.30	908,088.00	0.27	65.56	88.74
0.10	205.10	902,507.58	0.55	65.40	88.09	905,900.80	0.51	65.33	88.53

Table 7. Additional analysis for $\alpha = 0.6$ and $\beta = 10$

		greedy				*MILP-based*			
γ	#Disrupt. Events	Obj.	%gap	Qual. Score	Dist. Score	Obj.	%gap	Qual. Score	Dist. Score
0.00	0	850,153.51	0.00	68.26	85.44	852,056.85	0.00	68.03	86.14
0.01	20.70	849,231.60	0.11	68.24	85.32	852,013.67	0.01	68.11	85.99
0.05	103.17	847,473.15	0.32	68.05	85.18	850,725.74	0.16	68.02	85.78
0.10	205.10	846,250.66	0.46	67.93	85.00	849,395.33	0.31	67.92	85.56

7. Conclusions and Future Research

In this paper, we addressed a supplier selection problem (SSP) arising at *H2H Facility Solutions SpA*, an Italian global service provider (GSP) competing in the facility management industry, and we developed a decision support system (DSS) to help the decision makers at the company in the supplier evaluation and selection process. The SSP was formulated as a multi-objective generalized assignment problem and the evaluation of suppliers was based on a multi-criteria decision analysis performed with the company, for which the five macro criteria that directly contribute to defining the quality score for each supplier are the following: economic indicators, technical and professional capability, additional saturation capacity, service level performance, and references.

The DSS was implemented using a modular architecture. The first module is a MySQL relational database that stores information on contracts and suppliers. The second module is a supplier evaluator whose purpose is to compute the quality score of each supplier. The third module is a simulator that simulates the assignment of contracts to suppliers based on a rolling horizon algorithm with three alternative configurations (*company*, *greedy*, and *MILP-based*). A user-friendly interface that gives quick access to simple visual tools was also developed.

The effectiveness of the proposed system was tested by means of extensive computational experiments on real-data. The results proved the advantage of using the DSS, especially in those contexts in which we have a considerable number of contracts that must be assigned to a multitude of suppliers, for several categories of service, and where the assignments of contracts over the supplier's capacity is highly penalized. Furthermore, given the alternative solutions composing the approximate Pareto front, a decision maker can then choose the most suitable one according to his/her own experience and to the company's strategy. To further test the robustness of the rolling horizon algorithm, an additional analysis in case of unexpected or disruptive events was also performed.

As a concluding remark, we emphasize the flexibility and reproducibility of the proposed methodology which can be fully replicated with some adjustments for other real-world SSPs, in different contexts as well, by slightly reconsidering the proposed criteria. As future work, given the fast-growing relevance of GSPs, the high transferability of the proposed methodology, and the increasing adoption of intelligent systems to support decision-making processes, we intend to address other real-world applications and implement analogous DSSs, possibly embedding enhanced heuristics. Further future research directions may be represented by the introduction of a weight for each contract, which would make the problem NP-HARD to solve, and the execution of additional experiments to find, for example, a good estimation of the minimum number of suppliers that are needed to efficiently cover each geographical area or category of service.

Acknowledgements

We thank *H2H Facility Solutions SpA* for the support given to this project. We also thank the reviewers for their comments, which helped improve the paper.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

This work was funded under the National Recovery and Resilience Plan (NRRP), Mission 04 Component 2 Investment 1.5–NextGenerationEU, Call for tender n. 3277 dated 30/12/2021, Award Number: 0001052 dated 23/06/2022, and by the Brazilian research agency Conselho Nacional de Desenvolvimento Científico e Tecnológico, Grant: 432014/2018-7.

Data Availability Statement

Data is not available due to commercial restrictions.

References

Aczél, J. and Saaty, T. (1983). Procedures for synthesizing ratio judgements. *Journal of Mathematical Psychology*, 27(1):93–102.

Aguezzoul, A. (2014). Third-party logistics selection problem: A literature review on criteria and methods. *Omega*, 49:69–78.

Amaratunga, D. and Baldry, D. (2003). A conceptual framework to measure facilities management performance. *Property Management*, 21(2):171–189.

Amaratunga, D., Baldry, D., and Sarshar, M. (2000). Assessment of facilities management performance–what next? *Facilities*, 18(1/2):66–75.

Bannerjee, S., Bone, J., Finger, Y., and Haley, C. (2016). European digital city index methodology report. Available at `https://www.nesta.org.uk/blog/edci-2016-updating-the-european-digital-city-index/`. [Online; Accessed 31 January 2024].

Bruck, B., Vezzali, D., Iori, M., Magni, C., and Pretolani, D. (2021). A Decision Support System to Evaluate Suppliers in the Context of Global Service Providers. In *Proceedings of the 23rd International Conference on Enterprise Information Systems - Volume 1: ICEIS*, pages 420–430. INSTICC, SciTePress.

Bruno, G., Esposito, E., Genovese, A., and Simpson, M. (2016). Applying supplier selection methodologies in a multi-stakeholder environment: A case study and a critical assessment. *Expert Systems with Applications*, 43:271–285.

Çebi, F. and Bayraktar, D. (2003). An integrated approach for supplier selection. *Logistics Information Management*, 16(6):395–400.

Chai, J. and Ngai, E. (2020). Decision-making techniques in supplier selection: Recent accomplishments and what lies ahead. *Expert Systems with Applications*, 140:112903.

Chan, F. and Kumar, N. (2007). Global supplier development considering risk factors using fuzzy extended AHP-based approach. *Omega*, 35(4):417–431.

Chen, W., Chen, K., Cheng, J., Wang, Q., and Gan, V. (2018). BIM-based framework for automatic scheduling of facility maintenance work orders. *Automation in construction*, 91:15–30.

Derek, L. (1993). Integrating MCDM and DSS: Barriers and Counter Strategies. *INFOR: Information Systems and Operational Research*, 31(1):1–15.

Dweiri, F., Kumar, S., Khan, S., and Jain, V. (2016). Designing an integrated AHP based decision support system for supplier selection in automotive industry. *Expert Systems with Applications*, 62:273–283.

European Commission (2002). Competence Centre on Composite Indicators and Scoreboards (COIN). Available at `https://knowledge4policy.ec.europa.eu/composite-indicators_en`. [Online; Accessed 31 January 2024].

Fallahpour, A., Olugu, E., Musa, S., Wong, K., and Noori, S. (2017). A decision support model for sustainable supplier selection in sustainable supply chain management. *Computers & Industrial Engineering*, 105:391–410.

Figueira, J. and Roy, B. (2002). Determining the weights of criteria in the ELECTRE type methods with a revised Simos' procedure. *European Journal of Operational Research*, 139(2):317–326.

Gao, X. and Pishdad-Bozorgi, P. (2019). BIM-enabled facilities operation and maintenance: A review. *Advanced Engineering Informatics*, 39:227–247.

Garcia-Cascales, M. and Lamata, M. (2012). On rank reversal and TOPSIS method. *Mathematical and Computer Modelling*, 56(5–6):123–132.

Ghodsypour, S. and O'Brien, C. (1998). A decision support system for supplier selection using an integrated analytic hierarchy process and linear programming. *International Journal of Production Economics*, 56-57:199–212.

Ghodsypour, S. and O'Brien, C. (2001). The total cost of logistics in supplier selection, under conditions of multiple sourcing, multiple criteria and capacity constraint. *International Journal of Production Economics*, 73(1):15–27.

Goffin, K., Szwejczewski, M., and New, C. (1997). Managing suppliers: when fewer can mean more. *International Journal of Physical Distribution & Logistics Management*, 27(7):422–436.

Govindan, K., Rajendran, S., Sarkis, J., and Murugesan, P. (2015). Multi criteria decision making approaches for green supplier evaluation and selection: a literature review. *Journal of Cleaner Production*, 98:66–83.

Greco, S., Ehrgott, M., and Figueira, J. (2016). *Multiple Criteria Decision Analysis: State of the Art Surveys*, volume 233 of *International Series in Operations Research & Management Science*. Springer, New York, 2nd edition.

Gupta, S., Soni, U., and Kumar, G. (2019). Green supplier selection using multi-criterion decision making under fuzzy environment: A case study in automotive industry. *Computers & Industrial Engineering*, 136:663–680.

Hamdan, S. and Cheaitou, A. (2017). Supplier selection and order allocation with green criteria: An MCDM and multi-objective optimization approach. *Computers & Operations Research*, 81:282–304.

Ho, W. and Ma, X. (2018). The state-of-the-art integrations and applications of the analytic hierarchy process. *European Journal of Operational Research*, 267(2):399–414.

Ho, W., Xu, X., and Dey, P. (2010). Multi-criteria decision making approaches for supplier evaluation and selection: A literature review. *European Journal of Operational Research*, 202(1):16–24.

Ishizaka, A. and Labib, A. (2011). Review of the main developments in the analytic hierarchy process. *Expert Systems with Applications*, 38(11):14336–14345.

Ishizaka, A. and Nemery, P. (2013). *Multi-Criteria Decision Analysis: Methods and Software*. John Wiley & Sons, Ltd., Chichester, 1st edition.

Kellner, F., Lienland, B., and Utz, S. (2019). An a posteriori decision support methodology for solving the multi-criteria supplier selection problem. *European Journal of Operational Research*, 272(2):505–522.

Kincaid, D. (1994). Integrated Facility Management. *Facilities*, 12(8):20–23.

Kull, T. and Talluri, S. (2008). A supply risk reduction model using integrated multicriteria decision making. *IEEE Transactions on Engineering Management*, 55(3):409–419.

Lavy, S., Garcia, J., and Dixit, M. (2010). Establishment of KPIs for facility performance measurement: review of literature. *Facilities*, 28(9/10):440–464.

Mafakheri, F., Breton, M., and Ghoniem, A. (2011). Supplier selection-order allocation: A two-stage multiple criteria dynamic programming approach. *International Journal of Production Economics*, 132(1):52–57.

Matarneh, S., Danso-Amoako, M., Al-Bizri, S., Gaterell, M., and Matarneh, R. (2019). Building information modeling for facilities management: A literature review and future research directions. *Journal of Building Engineering*, 24:100755.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffmann, A., and Giovannini, E. (2020). Handbook on Constructing Composite Indicators: Methodology and User Guide. Available at `https://knowledge4policy.ec.europa.eu/publication/handbook-constructing-composite-indicators-methodology-user-guide-0_en`. [Online; Accessed 31 January 2024].

Nemati-Lafmejani, R. and Davari-Ardakani, H. (2021). Multi-mode resource constrained project scheduling problem along with contractor selection. *INFOR: Information Systems and Operational Research*, 59(2):212–231.

Patacas, J., Dawood, N., and Kassem, M. (2020). BIM for facilities management: A framework and a common data environment using open standards. *Automation in Construction*, 120:103366.

Pomerol, J. and Barba-Romero, S. (2000). *Multicriterion Decision in Management: Principles and Practice*, volume 25 of *International Series in Operations Research & Management Science*. Springer, New York, 1st edition.

Pretolani, D. (2020). Multiple criteria decision analysis theory and tools for the SDEWES index. *Journal of Sustainable Development of Energy, Water and Environment Systems*, 8(4):654–677.

Pun, K., Choy, K., and Lam, H. (2018). A Cloud-Based Fuzzy Multi-Criteria Decision Support System for Procurement Process in Facility Management. In *2018 Portland International Conference on Management of Engineering and Technology (PICMET)*, pages 1–8. IEEE.

Saaty, T. (1990). How to make a decision: The analytic hierarchy process. *European Journal of Operational Research*, 48(1):9–26.

Saputro, T., Figueira, G., and Almada-Lobo, B. (2022). A comprehensive framework and literature review of supplier selection under different purchasing strategies. *Computers & Industrial Engineering*, 167:108010.

SDEWES Centre (2016). SDEWES Index. Available at `https://www.sdewes.org/sdewes_index.php`. [Online; Accessed 31 January 2024].

United Nations Environment Programme (2012). Application of the Sustainability Assessment of Technologies Methodology: Guidance Manual. Available at `https://www.unep.org/resources/report/application-sustainability-assessment-technologies-methodology-guidance-manual-nov`. [Online; Accessed 31 January 2024].

Wang, G., Huang, S., and Dismukes, J. (2004). Product-driven supply chain selection using integrated multi-criteria decision-making methodology. *International Journal of Production Economics*, 91(1):1–15.

Wang, Y. and Luo, Y. (2009). On rank reversal in decision analysis. *Mathematical and Computer Modelling*, 49(5–6):1221–1229.

Weber, C., Current, J., and Benton, W. (1991). Vendor selection criteria and methods. *European Journal of Operational Research*, 50(1):2–18.

Wong, J., Ge, J., and He, S. (2018). Digitisation in facilities management: A literature review and future research directions. *Automation in Construction*, 92:312–326.

Zhang, Y., Niu, Z., Zuo, Y., and C., L. (2023). Two-stage hybrid model for supplier selection and order allocation considering cyber risk. *INFOR: Information Systems and Operational Research*, 61(4):530–558.