Valentina Ferretti, Jun Liu, V Mousseau and W Ouerdane

Reference-based ranking procedure for environmental decision making: insights from an ex-post analysis

Reference-based ranking procedure for environmental decision making: Insights from an ex-post analysis

V. Ferretti[a], J. Liu[b], V. Mousseau[b], W. Ouerdane[b]

[a] *Department of Management, London School of Economics and Political Science*
[b] *Laboratoire Génie Industriel, CentraleSupélec, Université Paris-Saclay, Gif-sur-Yvette, France*

Abstract

Preference elicitation is a challenging activity in any decision-making process, yet preferences are fundamental since the recommendations are meaningful and acceptable only if the Decision Maker's values are taken into account. This study proposes an ex-post application of a recent ranking method named Simple Ranking with Multiple Points (S-RMP) to support a participatory decision-making process. The method has been tested on a real-world case study simulating the selection of the most suitable site for locating a new landfill. The purpose of the research is twofold: (i) to explore the applicability and relevance of the S-RMP method to support environmental decision-making processes and (ii) to provide guidelines for the elicitation of preference parameters of the S-RMP ranking method. The results highlight that the proposed method opens a promising line of research in the environmental decision-making domain, thanks to its ability to use heterogeneous information consistently with the increasing amount of qualitative data embedded in real decision-making processes.

Keywords: Landfill Location, Multiple Criteria Decision Analysis, Preference learning, Ordinal data.

1. Introduction

Preferences are fundamental to decision processes, since the recommendations are meaningful and acceptable only if the decision makers'(DM) values are taken into consideration. Within this context, a challenging activity is "preference elicitation", which aims to capture the DMs' preferences in order to accurately specify the decision model parameters. The challenge is linked to the nature of the preferences expressed by the DMs, which can be imprecise, conflicting, unstable, time-dependent; yet they should be structured and synthesized. More precisely, preference elicitation is the process by which an analyst and a DM interact to fix the parameters of a preference model. The *indirect elicitation* approach avoids asking the DM to express her preferences in terms of numerical values for the considered parameters, but rather requires her to provide partial results she would like to obtain as an output of the preference model from which recommendations are constructed. To support the elicitation of stakeholders' preferences, many models, procedures and methodologies have been proposed (e.g. [25, 26, 35]), which try to cope with the proliferation of semi-automated computerized interfaces and the use of increasingly large datasets.

In this paper we are interested in showing the benefits of using a recent multiple criteria ranking method, named Simple Ranking with Multiple Points (S-RMP, see [4, 32, 33, 40]), to support

Email addresses: v.ferretti@lse.ac.uk (V. Ferretti), jinyan.liu@centralesupelec.fr (J. Liu), vincent.mousseau@centralesupelec.fr (V. Mousseau), wassila.ouerdane@centralesupelec.fr (W. Ouerdane)

complex decision-making processes. This ranking method is based on pairwise comparisons, but instead of directly comparing pairs of alternatives, it rather compares the alternatives to a set of predefined reference points. The idea is to construct the global preference relation between two alternatives on the basis of their relative comparisons with specified reference points.

Implementing the S-RMP ranking method to support a decision making process requires to elicit the decision maker's preferences and to set the values of preference parameters involved in the S-RMP method. Interacting with the decision maker directly on the values of the preference parameters is not recommended; *indirect elicitation* can proceed in computing the preference parameters values which best match holistic pairwise comparisons of alternatives provided by the DM. Within this paper, we are using two such algorithms to infer the parameters of the S-RMP method. The first algorithm is based on mathematical programming (exact algorithm, feasible only when a small number of holistic pairwise comparisons is considered), the second algorithm is grounded on a metaheuristic, which does not guarantee to find the model that best matches with the provided preferences, but is computationally more effective. The paper tests both algorithms to investigate how useful and efficient they are in practice.

Our context of application is an ex-post simulation of a decision process developed making use of publicly available data for the area under analysis. Through the simulation, process we organized 3 focus groups for the elicitation of preference information. During these focus groups, the task of the involved participants was to establish a preference order (ranking) on a set of alternative sites for the location of a new landfill on the basis of different criteria (e.g. presence of population, hydrogeological vulnerability, etc).

In this study we seek to show how the two previously mentioned disaggregation approaches (exact inference and metaheuristic approach) are able to elicit and represent the DMs' preferences in order to provide sound recommendations. The proposed approaches are characterized by the ability to take into account: *i)* the ordinal qualitative evaluation involved in the decision problem, *ii)* the inconsistency that may emerge from the presence of different participants (stakeholders) and *iii)* the presence of a relatively large amount of information and data. The objectives of the study are thus twofold. Firstly, to show on a real world case study how to effectively develop and apply a decision support process which makes use of the S-RMP ranking method, focusing in particular on the preference elicitation aspect, which is indeed the most challenging part in any decision making process. Secondly, we will focus on the question of how to elicit the preference parameters of S-RMP method in order to be effective and practice oriented.

The remainder of the paper is organized as follows. Section 2 positions this research in the environmental decision making context, by reviewing its present complexities and explaining the need for a new generation of Decision Support Systems. Section 3 contextualizes the decision process under analysis. Section 4 provides the methodological background for the S-RMP method. Section 5 explains the overall decision process proposed by the authors for the use of the S-RMP method in practice. In Section 6, we discuss the obtained results. Finally, Section 7 draws the conclusions and develops insights on how to use the S-RMP method in environmental decision-making processes.

2. Complexities and requirements in environmental decision making

There are several complexities in environmental decision making [36] that may explain the need for decision support systems, from both technical and social perspectives [6, 8].

Starting from a technical perspective, a first complexity is related to the selection of the method/approach to be used to support the decision making process. While several approaches are possible, Multi Criteria Decision Analysis (MCDA) seems to be a particularly promising one in the context of environmental decision making (e.g. [11]). MCDA is an umbrella term to describe a collection of formal approaches which seek to take explicit account of multiple criteria in helping individuals or groups explore decisions that matter. However, the selection of which approach to use in a specific decision making context is not a trivial one and this choice needs to be based on the specific characteristics of the problem under analysis (see for guidelines [34] and [6]). Another complexity is linked to the fact that some of the criteria to be considered in the assessments cannot be easily converted into a monetary value, partly because environmental concerns often involve ethical and moral principles that may not be related to any economic use or value [15], partly because of the difficulties of monetising intangibles [9]. This complexity thus leads to a frequent presence of qualitative data/criteria [17, 27, 29] in environmental decision-making processes. Thirdly, the spatial dimension of both the alternatives and the characteristics of the territory plays a crucial role in spatial planning [38]. Fourtly, the increasing volume, variety and velocity of information (the "*big data*" perspective, [23]) available to support decision making processes emphasizes the need to develop tractable methods for the aggregation of the information in a way which is meaningful for planners and decision makers [15]. Nowadays, factors such as technological advances for monitoring systems and the availability of Geographic Information Systems create increased accessibility and availability of data for environmental planning and decision making [22]. These increasingly big datasets generate the necessity to collect many preference information from the DMs and experts, who usually have very limited amount of time.

These technical complexities of decision making processes may explain the growing use of multi criteria analysis for environmental decision making (see [11, 16, 28, 31, 39]), as well as the increased interest towards preference learning algorithms [25].

From a social perspective, there are again several complexities in the environmental decision making context [6]. Firstly, such decision processes often involve many different stakeholders, with different objectives and priorities, thus representing exactly the type of problem that behavioral decision research has shown humans are poorly equipped to solve unaided [15] and leading to an increased demand for justification, legitimation and accountability of decisions [37]. Secondly, complex decision problems typically draw on multidisciplinary knowledge bases, incorporating natural and social sciences, as well as medicine, politics, and ethics [27]. Thirdly, and associated with the previous complexity, is the tendency of planning issues to involve shared resources, which means that group decision processes are often necessary [15]. However, groups are also susceptible to establish entrenched positions (defeating compromise initiatives) or to prematurely adopt a common perspective that excludes contrary information and suffers from "groupthink" (e.g. [15]). These social complexities may explain the increasing adoption of participative decision processes in those contexts and of facilitated decision modelling to support them. Finally, one of the most relevant complexities in environmental decision-making processes is the inherent trade-off between socio-political, environmental, ecological and economic factors [13]. This complexity has indeed both technical and social roots as it calls for the understanding of heterogeneous scales of impacts, deals with values and conflicting points of view and involves the possibility of behavioral biases [10].

Following from these considerations, the need for decision support systems able to (i) facilitate inclusion of different perspectives, (ii) handle qualitative data without the need to convert performances into a quantitative scale and (iii) learn the parameters of the model from a limited set of preference information in order to keep the cognitive burden on the decision makers as limited as possible can be highlighted. The research presented in this paper is an attempt in this direction.

3. Study contextualization

In this study we aim to show the benefits of using a recent multiple criteria ranking method, the S-RMP method (see section 4) within an environmental decision-making problem characterized by a relatively large dataset. In particular, the case study concerns the choice of the most suitable location for a Municipal Solid Waste landfill, which had to be constructed in the Province of Torino, Italy [1, 2].

The data used for the analysis are based on a scientific study that was developed by the Provincial Administration [2]. In this study, an environmental analysis of the territory under examination has been developed and 39 sites have been identified as potentially suitable for the location of the landfill. In particular, all the available information concerning the different locations has been organized according to a quantitative approach based on specific indicators.

The following characteristics of the decision-making problem under analysis make it a suitable one to be dealt with the S-RMP method [33]: (i) the availability of a relatively large data set, thus calling for the need to indirectly elicit part of the parameters from the Decision Makers; (ii) the type of results needed by the Decision Makers (i.e. a ranking from the most suitable site to the least suitable one); (iii) the complexity of the decision-making problem, reflected in the need to take into account the expertise and preferences of multiple Decision Makers, whose views would need to be compared across different dimensions.

In order to explore the applicability of the proposed method, we conducted an ex post study simulating the decision making process and involving experts in the elicitation and evaluation phases of the process. According to the Italian Regional Law 24/2002 (Regulations for waste management), the detection of suitable areas for the location of waste disposal and the recovery of municipal waste is articulated into 5 phases described in Table 1.

In our study we show how the proposed method can be used to support Phase 2 of the overall procedure. Moreover, based on the previously mentioned study [2], the criteria, summarized in Table 2, have been taken into account. All the criteria reflect the requirements coming from the legislative framework in the context of Environmental Impact Assessment procedures (first of all, the European Directive 11/97) and they all are to be minimized. Table 3 shows the complete performance table of the 39 sites under analysis.

Phase	Description	Steps
0	**Regional Planning** (competence of the Regional Authority)	• Regional Waste Management Plan. • Definition of criteria for detecting unsuitable areas for the location of waste treatment and disposal.
1	**Localization at the macro level** (competence of the Provincial Authority)	• Provincial Program of Waste Management. • Legal provisions of regional criteria. • Mapping the “unsuitable areas” and “potentially suitable areas”. • Criteria definition for the localization at the micro level.
2	**Localization at the micro level** (competence of local authorities in charge for waste management)	• Identification of suitable sites. • Definition of environmental compensation measures.
3	**Project** (competence of the specific bodies in charge for the construction of the plants)	• Definition of the project layout. • Environmental impact studies.
4	**Authorization** (jurisdiction of the Provincial Authority)	• Assessment of the environmental impact study. • Permission to build and operate.

Table 1: Planning procedure for waste disposal facilities’ locations

Crit.	Name	Description
c_1	Permanent population	The number of people living within a range of 1.5 km from each site.
c_2	Transitory population	The number of people that use the schools, the hospitals and the companies located within a range of 1.5 km from each site.
c_3	Vulnerability index	The vulnerability of the groundwater aquifer for each site measured through the GOD method (i.e. taking into account the Groundwater hydraulic confinement, the overlying strata and the depth to groundwater table; [7]).
c_4	Land use capacity	The potential productive capacity of the soil.
c_5	Farms	The number of organic farms in the area surrounding each site.
c_6	Interference with traffic	The distance between the landfill site and the waste collection points. This criterion measures the level of use of road infrastructures in the area surrounding the plant and estimates the potential interferences caused by the landfill.
c_7	Operating costs	The costs for the management and operation of the plant in each site, including expenses for linking the plant with the main ecological service pole.

Table 2: The set of criteria

Site	Permanent population	Transitory population	Vulnerability index	Land use capacity	Number of farms	Interference with traffic	Operating costs
Air_A	1461	1484	3	2	0	11050 m	3768.3K€
Air_B	3170	1757	3	2	1	10450 m	3561.8 K€
Bri_A	1356	974	4	2	2	6750 m	2186.5 K€
Bur_A	867	341	3	1	2	8000 m	2864.3 K€
Bur_B	623	225	3	1	1	6500 m	2050.9 K€
Caf_A	1356	693	3	3	4	15150 m	5179.6 K€
Cav_A	384	69	4	3	0	16650 m	5695.9 K€
Crc_A	345	15	3	1	0	9200 m	3131.4 K€
Cum_A	1859	684	2	4	2	7850 m	2782.9 K€
Cum_B	313	148	3	2	0	11450 m	3905.9 K€
Frs_A	140	507	3	4	0	8400 m	2856.1 K€
Frs_B	192	563	3	3	0	8000 m	2810.1 K€
Mac_A	1062	438	4	3	2	8200 m	2918.5 K€
Non_A	337	182	3	2	0	20550 m	7038.5 K€
Osa_A	981	569	4	2	5	7450 m	2566.1 K€
Pin_A	643	90	4	2	1	4150 m	4800.3 K€
Pin_B	1472	777	4	2	2	6600 m	2105.2 K€
Pis_A	1398	1242	3	2	2	8750 m	2976.5 K€
Ssp_A	3969	1397	4	2	2	5694 m	1613.9 K€
Vig_A	248	20	4	2	1	15000 m	5127.9 K€
Vil_A	433	25	4	2	0	19200 m	6573.7 K€
Vol_A	1139	445	3	2	2	18650 m	6384.4 K€
Air_2	2759	2072	3	2	2	10450 m	3681.7 K€
Air_3	1974	1561	3	2	0	10950 m	3389.6 K€
Air_4	1699	1527	3	2	0	10950 m	3389.6 K€
Non_1	242	369	3	3	0	21570 m	7389.5 K€
Fros_1	792	1128	3	2	1	5250 m	1373.2 K€
Fros_2	918	1530	3	2	0	5250 m	1373.2 K€
Pin_1	494	279	3	1	2	4700 m	1074.9 K€
Pin_2	525	125	3	1	2	4350 m	885.2 K€
Pin_3	485	119	3	1	2	5050 m	1264.7 K€
Pin_4	1043	455	2	2	3	4950 m	1454.5 K€
Pin_5	445	96	2	2	3	4950 m	1454.5 K€
Rol_1	1021	1486	3	2	0	5400 m	1454.5 K€
Sca_1	491	53	3	2	3	9850 m	3355.2 K€
Sca_2	454	42	3	2	3	9850 m	3355.2 K€
Sca_3	535	89	3	2	3	9850 m	3355.2 K€
Sca_4	310	15	3	1	0	9200 m	3131.5 K€
Vol_2	550	464	3	2	0	17350 m	5936.9 K€

Table 3: Performance table of the 39 potential sites (Data source : [2])

The location of undesirable facilities represents a complex decision-making problem [5], which usually calls for a participatory decision process able to ensure that all the relevant stakeholders are involved in the analysis and that a shared and robust decision can be made. To this end, we simulated the decision-making process by involving 3 experts and by eliciting from them preference information. As already anticipated, the availability of a relatively large data set (thus calling for the need to indirectly elicit part of the parameters from the Decision Maker), the type of results needed by the Decision Maker (*i.e.* a ranking from the most suitable to the least suitable location for the landfill) and the presence of qualitative data that should be interpreted on an ordinal scale, make the decision context under analysis a suitable one to be dealt with the S-RMP method.

4. Ranking with Multiple Points

4.1. An Illustrative example

We propose in this section an illustrative example to show how the Simple Ranking with Multiple Points (S-RMP) method proceeds. We will ground our example on the decision problem described in the previous section. We consider three alternatives: x, y and z evaluated on the following four criteria: Vulnerability index, Land use capacity, Interference with traffic, and Operating costs (see Table 2 for the description of the criteria). All criteria are to be minimized and the performances of the alternatives are presented in Table 4.

	Vulnerability index	Land use capacity	Interference traffic	Operating costs
x	1	1	4.90km	1454 K€
y	5	1	4.95km	885 K€
z	3	1	8.00km	480 K€
p^2	2	2	5.00km	1000 K€
p^1	4	3	12.00km	2000 K€
weight	0.25	0.25	0.25	0.25

Table 4: Illustrative example

The S-RMP ranking method makes use of preference parameters to specify the decision maker judgment: (i) reference points, (ii) a lexicographic order on these reference points, and (iii) criteria weights.

In our example, we use two reference points, namely p^1 and p^2 (which are vectors of evaluations), such that p^2_j is better than p^1_j on each criterion j. These two points allow to define three segments of performances on each criterion:

- better than p^2 (which can be interpreted as "*good*"),
- between p^1 and p^2 (which can be interpreted as "*intermediate or fair*"); and
- worst than p^1 (which can be interpreted as "*insufficient*").

The values of these points p^1 and p^2 on criteria are provided in Table 4. For instance, on the criterion "Vulnerability index", any alternative evaluated less than 2 will be considered "good" (e.g. alternative x) and any alternative evaluated more than 4 will be considered "insufficient" (e.g. alternative y). In other terms, the reference points allow to identify an ordered encoding for each criterion defined by 3 ordered intervals of performances (A, B and C) as illustrated in Figure 1, such that:

A performances above p^2 on each criterion are denoted as A (which can be interpreted as "*good*").

B performances between p^1 and p^2 on each criterion are denoted as B (which can be interpreted as "*intermediate or fair*").

C performances below p^1 on each criterion are denoted as C (which can be interpreted as "*insufficient*")

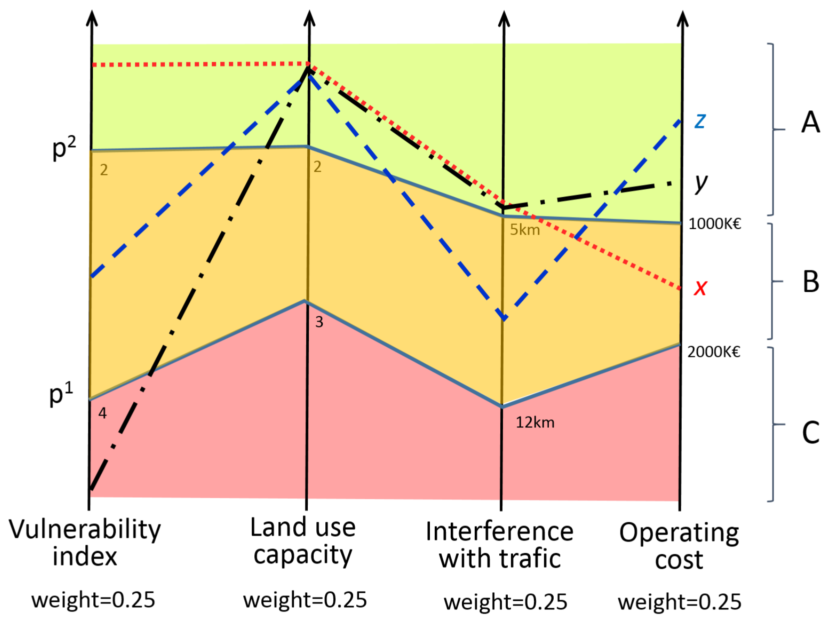

Figure 1: Graphical interpretation of Table 4

The S-RMP method ranks alternatives based on these ordered intervals of performances. Table 5 shows the results of the encoding for the 3 alternatives considered in our example. For instance, z is encoded B on criterion "Vulnerability index" because z is worse than p^2 but better than p^1. In addition, a lexicographic order is considered among these reference points. As there are two references points, there exist two possible orders: "p^1 then p^2" or "p^2 then p^1". In our example, the order is "p^1 then p^2".

	Vulnerability index	Land use capacity	Interference traffic	Operating costs
x	A	A	A	B
y	C	A	A	A
z	B	A	B	A

Table 5: Results of the encoding procedure for the illustrative example

To compute a ranking, alternatives are not compared one to each other but compared to the reference points. Alternatives are compared to the first point (here p^1 in the lexicographic order). Considering two alternatives a and b, a is preferred to b, noted $a \succ b$, if the number[1] of criteria for

[1] In this example, as criteria are equally weighted, we just count the number of criteria, but they could be weighted.

which alternative a is evaluated A or B (i.e. better than p^1) is greater than the number of criteria for which alternative b is evaluated A or B (i.e. better than p^1). If a and b cannot be distinguished with respect to their comparison to p^1 (the first point in the lexicographic order), then a and b are compared to p^2 (the second point in the lexicographic order). If the number of criteria for which alternative a is evaluated A (i.e. better than p^2) is greater than the number of criteria for which alternative b is evaluated A (i.e. better than p^2), then a is preferred to b, otherwise a is indifferent to b.

In our example, we thus have the following:

- Alternative x is better than y because x has evaluation A or B for all criteria, while y has evaluation A or B for only three criteria (x compares better to p^1 than y does).
- Alternative x is better than z because x and z are both evaluated A or B on all criteria (they compare equally to p^1), but x is evaluated A on three criteria while z is evaluated A only on two criteria (x compares better to p^2 than z does).
- Alternative z is better than y because z has evaluation A or B on all criteria while y has evaluation A or B on three criteria only (z compares better to p^1 than y does).

In conclusion, we obtain that x is globally the best alternative, followed by z and then y.

4.2. The S-RMP ranking method

Let us consider a set $\mathcal{A}$ of alternatives evaluated on m criteria. We denote by $M = \{1, 2, \ldots, j, \ldots, m\}$ the set of criteria indices, while a_j denotes the evaluation of alternative $a \in \mathcal{A}$ on criterion j (in what follows we will consider, without loss of generality, that preferences increase with the evaluation on each criterion, i.e. the greater the better). The S-RMP method is a method for ranking a finite set of alternatives evaluated on several criteria [18]. It is a variant[2] of the RMP (Ranking with Multiple Points) ranking method proposed by [33]. To rank alternatives, RMP proceeds by comparing alternatives to reference points, and then aggregates these comparisons into a final ranking. A dominance structure can be imposed to the set of reference points without loss of generality (for any RMP model using a set of reference points without any dominance structure, there exist an equivalent RMP model using a set of reference points with a dominance structure).

S-RMP makes use of three different types of preference parameters:

- $\mathcal{P} = \{p^h, h = 1 \ldots k\}$, with $p^h = \{p^h_1, ..., p^h_j, ..., p^h_m\}$, where p^h_j denotes the evaluation of p^h on the criterion j and k is the number of reference points;
- σ, a lexicographic order on the reference points, *i.e.*, a permutation on $\{1, 2, ..., k\}$. Note that this lexicographic order can be any total order on points (from the worst to the best point, from the best to the worst point, or any other);
- criteria weights $w_1, w_2, \ldots w_m$, where $w_j \geq 0$, and $\sum_{j \in M} w_j = 1$.

S-RMP proceeds by using three steps procedure:

[2] S-RMP considers additive weights while RMP defines criteria importance as non necessarily additive, and S-RMP considers a lexicographic order on points, while RMP handles a more general form for aggregating how alternatives compare to the different points.

1. compute $C(a, p^h) = \{j \in M : a_j \geq p^h_j\}$ with $a \in \mathcal{A}$, $h = 1, \ldots k$, the set of criteria on which alternative a is at least as good as point p^h.
2. compare alternatives one to each other to define k preference relations $\succsim_{p^h}$ relative to each point such that $a \succsim_{p^h} b$ iff $\sum_{j \in C(a,p^h)} w_j \geq \sum_{j \in C(b,p^h)} w_j$. In other words, $a \succsim_{p^h} b$ holds when a compares better to p^h than b does. We will denote $\succ_{p^h}$ ($\sim_{p^h}$, respectively) the asymetric part of the relation $\succsim_{p^h}$ (the symetric part of $\succsim_{p^h}$, respectively).
3. to rank two alternatives $a, b \in \mathcal{A}$, consider sequentially the relations $\succsim_{p^{\sigma(1)}}, \succsim_{p^{\sigma(2)}}, \ldots, \succsim_{p^{\sigma(k)}}$ (according to the lexicographic σ); a is preferred to b if $a \succ_{p^{\sigma(1)}} b$, or if $a \sim_{p^{\sigma(1)}} b$ and $a \succ_{p^{\sigma(2)}} b$, or ... Hence, a and b are indifferent iff $a \sim_{p^{\sigma(h)}} b$, for all $h = 1 \ldots k$.

Rolland [33] proved that by proceeding in such a way, the computed preference relations on alternatives are guaranteed to be transitive. Thereby, we can further deduce a weak order, i.e., the ranking (with ties) of $\mathcal{A}$. As mentioned earlier, a dominance structure on the reference points can be hypothesised without loss of generality; it should however be emphasized that this dominance order on reference points does not necessarily correspond to the lexicographic order σ used to aggregate how alternatives compare to these reference points.

4.3. Methodology for implementing the S-RMP ranking method

When implementing the S-RMP ranking method in a decision aiding case study, it is necessary to interact with the decision maker, so as to integrate her preferences, hence set the values of the preference parameters involved in the S-RMP method. A first approach (referred as *direct elicitation* in the literature, see e.g. [24]) consists in interacting with the decision maker directly on the values of the preference parameters. However, such approach is not recommended as the decision maker has usually not a clear understanding of the semantics attached to the preference parameters. Moreover, it imposes a strong cognitive burden on the decision maker. Therefore, the literature frequently proposes an *indirect elicitation*, in which the decision maker expresses holistic preferences (i.e, pairwise comparisons of real or fictitious alternatives) from which the values of the preference parameters are inferred (see e.g. [3, 12, 26]).

Recent literature (see [18, 40]) proposed indirect elicitation procedures for the S-RMP method, in which the decision maker provides a list $\mathcal{BC}$ of binary comparisons of alternatives (a partial ranking), from which the S-RMP preference parameters (weights, reference points, and the lexicographic order on reference points) are inferred. Such inference is performed through the resolution of an optimisation problem (see Appendix A) in which the S-RMP preference parameters are variables, the constraints express the DM's preference statement $\mathcal{BC}$, and the aim is to minimize the Kemeny distance (see [14]) between the partial ranking provided by the DM (i.e. $\mathcal{BC}$) and the S-RMP ranking. In the following, we call S-RMP model the set of S-RMP preference parameters. [18, 40] proposed algorithms in order to solve the inference program and hence to indirectly elicit an S-RMP model from the DM preference statements. Hence, based on the comparisons $\mathcal{BC}$, such inference computes the values of the S-RMP parameters : the set of k reference points $\{p^h, h = 1 \ldots k\}$, the lexicographic order σ on the reference points, and the criteria weights $w_1, w_2, \ldots w_m$.

4.3.1. Exact elicitation algorithm

[40, 18] formulated the elicitation of S-RMP model as a mixed linear optimization problem. In this optimisation program, the variables are the parameters of the S-RMP method, and additional technical variables which enable to formulate the objective function and the constraints in a linear form. The objective function minimizes the Kemeny distance between the comparisons expressed by the DM and the S-RMP ranking. The optimal resolution of this optimisation program provides

a guarantee that the elicited S-RMP model best match the pairwise comparisons in term of the Kemeny distance. We recall briefly the essential elements of the mathematical formulation of the exact elicitation algorithm in Appendix A. For more details, the readers can refer to [18].

4.3.2. Metaheuristic algorithm

Another algorithm to indirectly elicit an S-RMP model, from pairwise comparisons, was proposed by [19, 18]. Unlike the exact version [18, 40], this algorithm is based on a metaheuristic and does not guarantee that the inferred model is the one which minimises the Kemeny distance to $\mathcal{BC}$. Indeed, the perspective is to obtain an S-RMP model which "well" fits $\mathcal{BC}$ within a "reasonable" computing time. Such approach makes it possible to deal with datasets for which exact optimization algorithms become computationally intractable.

This metaheuristic is based on an evolutionary algorithm [20, 30, 21] in which a population of S-RMP models is iteratively evolved. First, a population of S-RMP models is initialized. Then, at each iteration, for each S-RMP model in the population, the algorithm performs the following steps:

- adjust the reference points using a heuristic keeping the other parameters constant;
- adjust the weights using linear programming keeping the other parameters constant; and
- apply a mutation operator (with a small probability) to introduce diversity in the population.

At the end of each iteration, the best S-RMP models in the population in terms of the Kemeny distance to $\mathcal{BC}$ are kept for the next iteration. The algorithm stops when the population contains an S-RMP model which fully restores $\mathcal{BC}$ or a maximum number of iterations is reached. All details of this algorithm can be found in [19, 18].

5. Implementing the S-RMP method in a real-world application

In this section, we discuss the implementation of the S-RMP method within a real world context. The decision problem concerns the selection of the most suitable site for locating a new landfill. To do that, a sequence of three facilitated focus groups was organized, as illustrated in Figure 2 and explained in paragraphs 5.1, 5.2 and 5.3. Three experts in environmental engineering participated in all focus groups. The discussion and all tasks proposed during the focus groups were facilitated by the first author of this paper. Between each meeting, the other authors used the two elicitation algorithms explained in sections 4.3.1 and 4.3.2 to infer, based on the preferences expressed by the participants, the parameters of S-RMP models.

5.1. First focus group: collecting individual preferences

The first focus group consisted of three phases.

Phase 1: during this phase, the facilitator in the focus group described the decision problem concerning the location of the landfill to the 3 participants (i.e. the general objective of the evaluation, the considered criteria, the difficulties, the actors involved and the considered suitability classes for the location of a new landfill).

Phase 2: in this phase, the three involved participants had to work individually and assign the 39 alternative sites for the location of the landfill to the following three pre-defined classes:

- "Highly suitable", i.e., sites for which most of the evaluations on the considered criteria are positive, without major drawback;

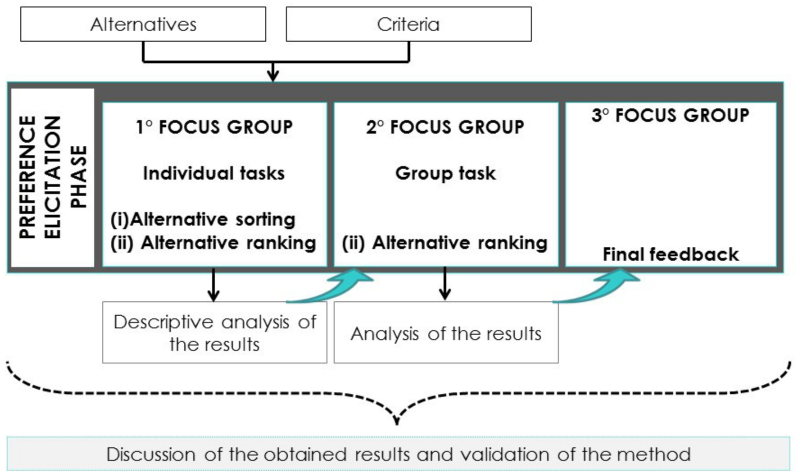

Figure 2: The developed decision making process

- "Suitable", i.e., sites with fairly satisfactory evaluations on most of the criteria and without strongly negative evaluations on the other criteria;
- "Unsuitable", i.e., sites with a significant number of negative evaluations which make the sites as clearly unsatisfactory.

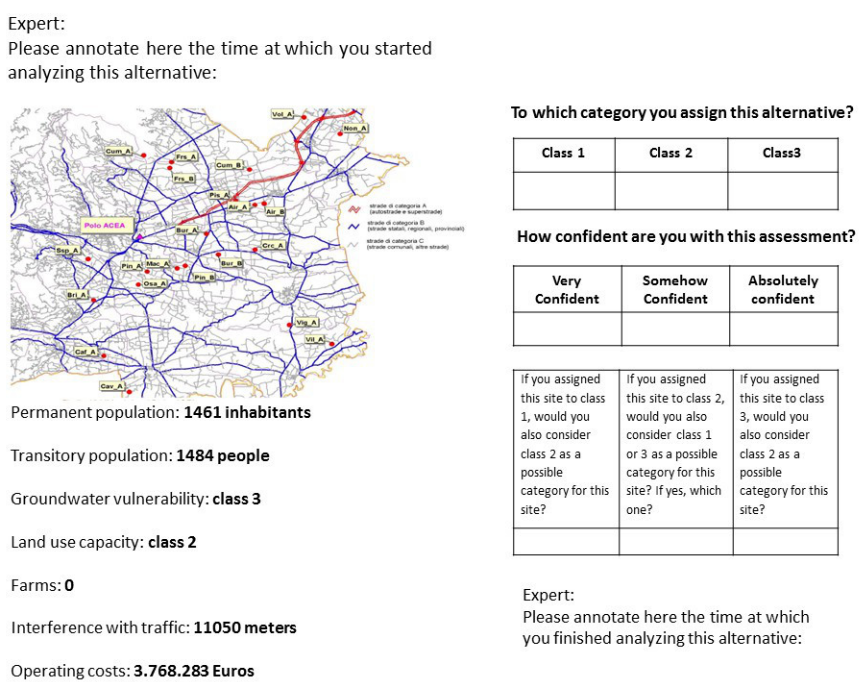

Figure 3: Example of ID card for alternative AIR_A

To perform the classification of the 39 sites, participants are provided with a booklet on which each site is described on a sheet showing: the site, the performance of the site according to the considered criteria (see Table 2), a box for the definition of the category to which to assign the site with an associated indication of the level of confidence with the assessment (i.e. very confident, somehow confident, absolutely confident) and, finally, 3 questions allowing to further check the belonging of the site to the specific class being identified by the participant. Figure 3 shows an example of the ID card for alternative AIR_A.

Two aspects are worth being highlighted: (i) particular attention was devoted to the use of the map on each alternative card in order not to bias in any possible way the participants. The map was indeed showing only the administrative boundaries of the municipalities with no visualization of landscape and critical environmental assets that could have led the participants to consider criteria in their assessment different that those provided in the card. (ii) Based on a preliminary analysis of the sites, participants were told that, considering the definition of the 3 classes, the classification of the 39 alternative sites should result in classes of similar size.

Phase 3: once each of the participants had the three classes, they individually ranked the sites inside each class from the best one to the worst one. The facilitator kept track of the time that they needed to complete this task for each of the three classes in order to be able to understand later which task turned out to be more cognitively demanding in the overall process. The facilitator also asked them to check whether the last site in class 1 was better or worse than the first site in class 2. If the worst site in class 1 was judged better than the better site in class 2, then these two sites are permuted in the resulting ranking.

Thereby, the individual rankings of the 39 sites were collected. Before organizing the second focus group, the authors of the paper analysed the individual rankings and carried out the computation of S-RMP preference parameters, meaning that for each participant we provided the model that best fitted his/her preferences. After that, we compared the three individual rankings obtained from the inferred models with the ones provided by the participants, and pointed-out the inconsistent pairwise comparisons. This means that the inferred model ranked some alternatives differently from the participants. At the end of the first focus group, the facilitator also collected qualitative feedback on the overall process from the participants. The obtained feedback and insights are discussed in detail in Section 6.

5.2. Second focus group: collecting group preferences

Between the first focus group and the second one, which took place one week later, the authors of the paper carried out calculations based on the provided individual rankings and developed a descriptive analysis of how the participants proceeded in the individual tasks (i.e. which sites turned out to be the highly suitable ones and which the least suitable ones according to all the participants, as well as time statistics). The second focus group then consisted of two phases.

Phase 1: during this phase, the facilitator illustrated the descriptive analysis of the results obtained in the first focus group showing similarities and differences among the participants. In particular, the facilitator highlighted the conflicting comparisons raised from the individual rankings and asked each participant to look again at the alternative cards and revise the inconsistencies that were found (17 violations for Expert 1, 7 violations for Expert 2 and 7 violations for Expert 3). Note that these inconsistencies refer to the inability to restore the pairwise comparisons provided by the experts using the S-RMP ranking method.

Phase 2: this phase was devoted to a group discussion in order to achieve a consensus based on the descriptive feedback provided on the individual rankings. The participant discussed all together in order to obtain a shared vision and a common ranking. In order to do this, the facilitator asked them to first look at those sites on which they had similar opinions and then look at the remaining ones. Table 6 presents the agreed group ranking.

Rank	Alternative	Rank	Alternative	Rank	Alternative
1	Pin_2	14	Air_A	27	Non_1
2	Fros_2	15	Pin_4	28	Cav_A
3	Pin_1	16	Cum_A	29	Mac_A
4	Pin_3	17	Fros_1	30	Air_2
5	Pin_5	18	Frs_B	31	Air_B
6	Rol_1	19	Sca_2	32	Caf_A
7	Sca_4	20	Pis_A	33	Bri_A
8	Crc_A	21	Air_4	34	Pin_B
9	Bur_A	22	Non_A	35	Ssp_A
10	Frs_A	23	Sca_1	36	Osa_A
11	Pin_A	24	Sca_3	37	Vig_A
12	Bur_B	25	Vol_2	38	Vol_A
13	Cum_B	26	Air_3	39	Vil_A

Table 6: Group ranking of the 39 sites

At the end of the second focus group, the facilitator collected again qualitative feedback on the overall process from the participants. The obtained feedback and insights are discussed in detail in Section 6.

5.3. Third focus group: final feedback

The third focus group was conceived as a double direction feedback section. On the one hand, the facilitator provided a quantitative feedback to the participants all together in order to update them on the results obtained from the model based on the information provided by each of them and by them as a group. On the other hand, the participants in this study were asked to provide a qualitative feedback on both the obtained results (to see if they were coherent with their expectations) and on the overall experimental decision protocol that they experienced. These results and feedback are discussed in detail in Section 6.

6. Results

6.1. Overall results

We first present, in Table 7, the results of the final Kemeny distance (Kemeny distance between the binary comparisons $\mathcal{BC}$ and the computed S-RMP ranking) for both the exact and the metaheuristic algorithms. We vary the number of points from $k = 1$ to $k = 3$. Obviously, the Kemeny distance of the inferred S-RMP model improves with the number of points, as adding points increases the descriptive ability of the S-RMP model. Moreover, we provide the number of equivalence classes[3] in the aggregated ranking.

We observe that both the exact algorithm and the metaheuristic can provide the S-RMP model that best matches expressed preferences for one point ($k = 1$); however, when $k \geq 2$, the metaheuristic is not able to identify the S-RMP model that best matches the expressed preferences (the one provided by the exact algorithm). Nevertheless, the metaheuristic provides an S-RMP model which, if not optimal has a Kemeny distance to $\mathcal{BC}$ close to the optimal one ($k = 2$: 7 vs 6; $k = 3$: 7 vs 5).

[3] In an equivalence class, alternatives are indifferent.

		$k=1$	$k=2$	$k=3$
Exact algorithm	Kemeny distance to $\mathcal{BC}$	8	6	5
	Number of equivalence classes	19	29	29
Metaheuristic	Kemeny distance to $\mathcal{BC}$	8	7	7
	Number of equivalence classes	20	21	23

Table 7: Kemeny distance from $\mathcal{BC}$ to the inferred S-RMP model

Second, an important aspect in implementing algorithms is the computational time. We run the exact elicitation algorithm on a computer cluster Altix ICE 8400 LX (with 69 nodes in total, and each node is equipped with two six-core Intel Xeon Processor X5650), while the metaheuristic was executed on an ordinary laptop computer. Moreover, for the exact elicitation algorithm we used 12 cores, while for the metaheuristic 4 cores. We note the following results: in the case of a single reference point the elapsed time of the exact algorithm is 2.09 hours while it is 1.62 hours for the metaheusitic; for two references point it was 22.16 hours for the exact algorithm while it was 2.05 hours for the metaheuristic. Finally, when we have three reference points, the elapsed time was 75.15 hours for the exact version and 3.24 hours for the metaheuristic. What we note is that although we used more cores for executing the exact algorithms, the metaheuristic is faster.

6.2. Interpretation of the inferred S-RMP models

The next results correspond to the S-RMP models inferred on the collective experts ranking: Table 8 presents the S-RMP model inferred by the metaheuristic and Table 9 the one inferred by the exact algorithm.

	c_1	c_2	c_3	c_4	c_5	c_6	c_7
p^3	193	53	2	1	0	7400	3619.3 K€
p^2	1423	126	3	1	2	7407	5782.6 K€
p^1	1752	1632	3	4	2	8719	8561.0 K€
ω	0.125	0.125	0.250	0.125	0.125	0.125	0.125
Lexicographic order: compare to p^2 then to p^3 then to p^1							

Table 8: The S-RMP model inferred by the metaheuristic

	c_1	c_2	c_3	c_4	c_5	c_6	c_7
p^3	310	53	1	1	3	7850	1074.9 K€
p^2	1471	437	2	1	3	8199	1264.7 K€
p^1	3169	1561	3	4	3	18649	2856.0 K€
ω	0.110	0.111	0.111	0.001	0.222	0.111	0.332
Lexicographic order: compare to p^1 then to p^2 then to p^3							

Table 9: The S-RMP model inferred by the exact algorithm

These two inferred models were presented to the participants and explained in practical and interpretable terms. For instance, if we consider the S-RMP model inferred by the exact algorithm, the obtained lexicographic order is "p^1 then p^2 then p^3", such that p^3 is better than p^2 which is better than p^1 on each criterion.

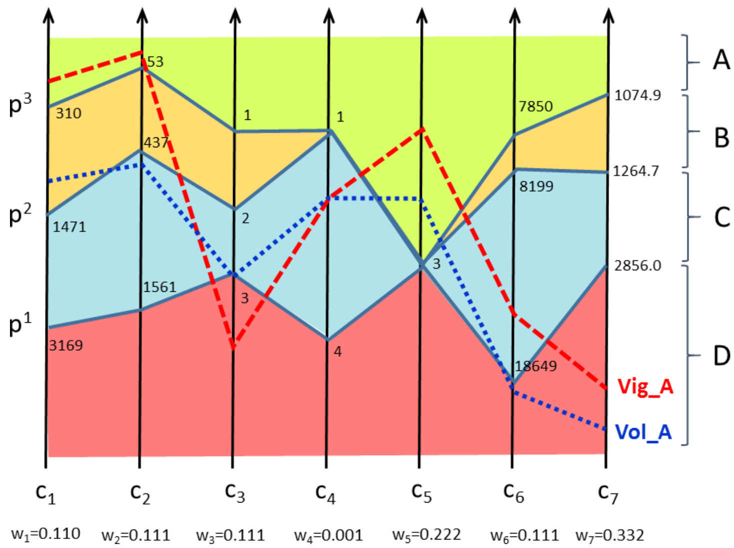

Figure 4: Graphical Interpretation of Table 9

In other terms, the reference points allow to identify an ordered encoding on each criterion into 4 ordered intervals of performances (A, B, C and D) as illustrated in Figure 4, such that:

A performances above p^3 on criterion c_j are denoted as A (which can be interpreted as "*very good*")

B performances between p^2 and p^3 on criterion c_j are denoted as B (which can be interpreted as "*good*")

C performances between p^1 and p^2 on criterion c_j are denoted as C (which can be interpreted as "*fair*")

D performances below p^1 on criterion c_j are denoted as D (which can be interpreted as "*insufficient*")

Then the S-RMP model ranks alternatives based on these ordered interval of performances. To illustrate the ranking procedure of the S-RMP model presented in Table 9, let us consider two sites, among the 39 sites (see Table 3), Vig_A and Vol_A. The performances of these sites are reported in Table 10, and these performances are "encoded" into A, B, C, and D in Table 11.

	c_1	c_2	c_3	c_4	c_5	c_6	c_7
Vig_A	248	20	4	2	1	15000 m	5127.9 K€
Vol_A	1139	445	3	2	2	18650 m	6384.4 K€

Table 10: Performances of Vig_A and Vol_A

For instance, Vig_A is encoded C on c_4 because Vig_A is worst than p^2 but better than p^1.

	c_1	c_2	c_3	c_4	c_5	c_6	c_7
Vig_A	A	A	D	C	A	C	D
Vol_A	C	B	C	C	A	D	D

Table 11: Encoded performances of Vig_A and Vol_A

To compute a ranking, we recall that the alternatives are not compared one to each other but compared to the reference points. Alternatives are compared to the first reference point (here p^1 in the lexicographic order). According the S-RMP model presented in Table 9, the lexicographic order on points is p^1, then p^2, then p^3; hence, alternative a is preferred to alternative b, noted $a \succ b$, if the sum of the weights of criteria for which a is evaluated C, B or A (i.e. better than p^1) is greater than the sum of the weights of criteria for which b is evaluated C, B or A (i.e. better than p^1). If a and b cannot be distinguished with respect to their comparison to p^1 (the first point in the lexicographic order), a and b are compared to p^2 (the second point in the lexicographic order). If the sum of the weights of criteria for which a is evaluated B or A (i.e. better than p^2) is greater than the sum of the weights of criteria for which b is evaluated B or A (i.e. better than p^2), then a is preferred to b. Finally, if a and b cannot be distinguished with respect to p^2, a and b are compared to p^3 (the third reference point in the lexicographic order). If the sum of the weights of criteria for which a is evaluated A (i.e. better than p^3) is greater than the sum of the weights of criteria for which b is evaluated A (i.e. better than p^3) then a is preferred to b, otherwise a is indifferent to b.

For the two sites Vig_A and Vol_A, we observe that Vol_A has evaluations C, B or A for all criteria except for c_6 and c_7 (weight sum = 0.55), while Vig_A has evaluations C, B or A for all criteria except c_3 and c_7 (weight sum = 0.55). Thus, Vol_A is indifferent to Vig_A w.r.t the point p^1. Therefore, we compare the two alternatives w.r.t p^2: Vol_A has evaluations B or A for criteria c_2 and c_5 (weight sum = 0.33), while Vig_A has evaluations B or A for criteria c_1, c_2 and c_5 (weight sum = 0.44). In conclusion, alternative Vig_A is preferred to alternative Vol_A.

Finally, if we use the inferred S-RMP model presented in Table 9 to rank the 39 alternatives, we obtain the ranking presented in Table 12. It can be observed that this ranking does not fully match with the one proposed by the experts (only 33 out of the 38 provided pairwise comparisons[4] are restored, i.e. RA=86.4%). Moreover, the computed ranking with the inferred S-RMP model includes indifference situations.

6.3. Feedback from the participants

During the different focus groups, we collected some feedback from the experts. On the one hand, we were concerned about the impressions and the experiences of the participants regarding the overall experimental protocol developed for implementing the S-RMP method in practice. On the other hand, it was interesting to record the comments of the participants on the inferred S-RMP models or, more generally, on the S-RMP aggregation method.

About the elicitation procedure. To compare the alternatives pair by pair and establish a complete ranking, the participants noted that it was very difficult to deal coherently with the 39 sites. Moreover, two participants among three noticed that starting by a sorting procedure add some

[4] As the experts provided a ranking on the 39 alternatives which is supposed to be transitive, the 38 pairwise comparisons concerning consecutive alternatives in the ranking convey the whole information of the expressed ranking, and all other comparisons are redundant.

Table 12: Ranking of the 39 sites (S-RMP model by exact elicitation)

Rank	Alternatives
1	Pin_1, Pin_2
2	Pin_3
3	Pin_5
4	Bur_B
5	Pin_4
6	Cum_A
7	Fros_1, Fros_2, Rol_1
8	Frs_B
9	Frs_A
10	Pin_A
11	Bri_A
12	Pin_B
13	Ssp_A
14	Bur_A
15	Sca_4
16	Crc_A
17	Sca_1, Sca_2
18	Cum_B, Sca_3
19	Air_A, Pis_A, Vol_2
20	Air_3, Air_4
21	Osa_A
22	Vig_A
23	Non_1
24	Cav_A, Non_A
25	Mac_A, Vol_A
26	Air_2
27	Air_B
28	Caf_A
29	Vil_A

difficulties to the task. This actually implies that an alternative approach aiming at collecting sample rankings should be considered in order to better guide the participants in their reflections.

About the inferred reference points. The participants acknowledged coherence with the first reference point used (p^1, i.e., the lowest one), since they were actually using a sort of "reference point" to distinguish the "worst" sites from the others. Two important remarks can be made with reference to this aspect:

- With reference to the preference elicitation procedure, the inferred values of the parameters of the S-RMP method do indeed correspond partially to the real value system of the participants (DMs) that has been used during the decision process.
- With reference to the multiple criteria aggregation procedure, the logic behind the S-RMP method of using reference points in a dictatorial lexicographic order corresponds to the actual reasoning system of the participants (DMs) with which they processed the ranking problem.

On the other hand, they were also surprised by the inferred value—"1"—for both the reference points p^2 and p^3, particularly on criterion c_4 (i.e., land use capacity), because this implies that the potential sites for locating the new landfill should be absolutely suitable for agricultural crops as well (See Section 3), whereas they stated that this is not true in reality. However, when they noticed the extremely low weight of criterion c_4, they understood and accepted the inferred value.

7. Discussions

7.1. Dealing with possible violations

During the first and the second focus group, the authors of the paper asked the participants to provide, respectively, individual and group rankings of the 39 sites. Taking these into account, the authors of the paper computed the S-RMP models corresponding to the participants' provided preferences. However, as it was previously highlighted, the inferred models were not able to restore some pairwise comparisons (called violations). For instance, the group provided the following statements: Fros_2 $\gtrsim$ Pin_1; Crc_A $\gtrsim$ Bur_B; Air_A $\gtrsim$ Pin_4; Caf_A $\gtrsim$ Bri_A; and Non_A $\gtrsim$ Sca_1.

If we consider the S-RMP model inferred by the exact elicitation (see section 4.3.1), we obtain the following results: Pin_1 $\succ$ Fros_2; Bur_B $\succ$ Crc_A; Pin_4 $\succ$ Air_A; Bri_A $\succ$ Caf_A; and Sca_1 $\succ$ Non_A.

In general and from a practical point of view, two possible procedures can be adopted in order to refine and correct progressively the results of a preference elicitation process:

- To ask the participants to revise the violations, if feasible, and replace these identified inconsistencies with the revised ones; otherwise, add additional strict constraints on the pairwise comparisons that can be confirmed.
- To ask the participants to provide their preferences on the sites that have been ranked indifferent, and take these supplementary information alongside the already provided pairwise comparisons.

7.2. Applicability of the elicitation algorithms

The study allowed us to draw some conclusions about the applicability of S-RMP preference elicitation algorithms to real-world situations according to different aspects:

- Dealing with inconsistent pairwise comparisons: in a real context it is not surprising to face decision making situations where preference statements, provided by heterogeneous participants, may include inconsistent information. This may affect the quality of the recommendation. The proposed algorithms (the exact version and the metaheuristic) were designed to tackle this issue and are suitable for such situations.

- Computation time vs guarantee of optimality: an important issue when implementing a method for a real situation is to be able to get results in a reasonable time. As it was discussed in Section 6, for a fixed number $k = 3$ of reference points the exact algorithm is less efficient than the metaheuristic, although it was executed on a powerful computer cluster. However, as shown in Table 7, when $k \geq 2$, the loss of optimality of the metaheuristic becomes non negligible. Thus, in terms of quality of solutions the exact algorithm provides better results as it is able to infer more accurate models.

7.3. Methodological insights on S-RMP discriminability

This study also led us to highlight specific methodological insights concerning the ability of the S-RMP ranking method to discriminate alternatives, which may have strong implications in practice. In a simulation study, we considered the 39 sites under analysis as alternatives to be ranked. We varied the number of reference points from 1 to 8. For each fixed number of reference points, we randomly generated 10^6 S-RMP models (S-RMP parameters are sampled uniformly). We computed the resulting ranking and observed the number of equivalence classes in these rankings. The number of equivalence classes in the obtained ranking shows the ability of an S-RMP method to discriminate among these 39 specific alternatives.

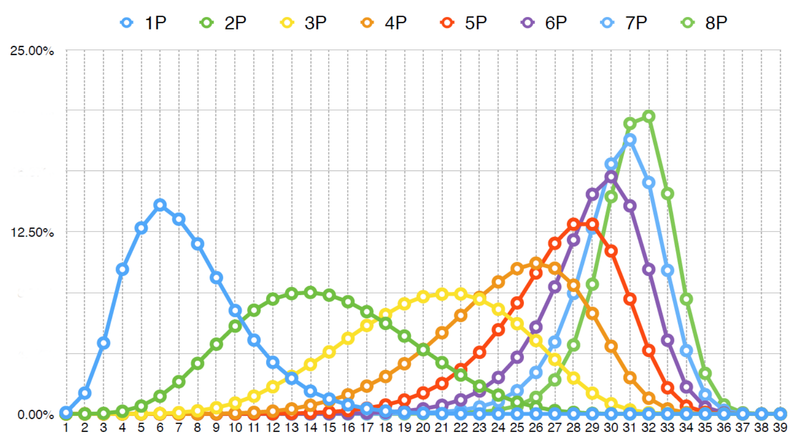

Figure 5: Distribution of the number of equivalence classes of rankings with S-RMP

Figure 5 describes the distribution of the number of equivalence classes of rankings derived by random S-RMP models, when the number of reference points is varied. The results show very clear trends: with no more than 2 reference points, the number of equivalence classes of the rankings of the 39 sites never reaches 39; when the number of reference points increases, there is a clear trend of increased number of equivalence classes; as the number of points increases, the number of equivalence classes increases but in a slower way.

These results have strong implications in terms of practice: based on this particular set of 39 alternatives, when eliciting an S-RMP ranking model with one reference point the median number of equivalence classes in the ranking is 6. If the analyst asks the DM for a ranking with 10 alternatives (among these 39), there is a high risk that the expressed preference will not be fully compatible with an S-RMP model with one reference point (due to the number of equivalence classes of the ranking). The analyst could in this particular case, either ask the DM for a ranking with possible ties (less equivalence classes), or consider an S-RMP model with more reference point s. Such analysis could be generalized to any case study by performing the same analysis on the particular set of alternatives under consideration.

8. Conclusions

This paper presented an ex post simulation of a decision concerning the location of a new landfill site. The work simulated a decision making process and involved different experts in the elicitation and evaluation phases of the process. Our purpose was to test the relevance and usefulness of a new ranking method (S-RMP) in the environmental decision making field.

The key strength of the method, which makes it a promising line of research in the field of environmental decision making, is its capacity to use information in a qualitative way, which is consistent with the increasing use of qualitative data in real decision making applications (e.g. intangible aspects referring to the level of impact on the landscape). Indeed, the possibility offered by this approach to take into account only the ordinal part of the data presented in the performance table, makes it easy to use it in indirect elicitation processes.

The main contribution of our work comes from the illustration of the applicability of the proposed S-RMP method on a real setting and not on a toy example. We indeed made use of real data and tested the preference elicitation burden with real participants. By representing the first application of the proposed methodological approach on a real setting, this paper has thus an innovative value and will hopefully stimulate further applications in similar as well as different domains making use of qualitative assessment protocols. Moreover, through this application we gained a novel understanding of the capacity of the S-RMP method. An important issue that remains to be tackled lies in the comparison of S-RMP method to other multiple criteria ranking methods and compare their relative usefulness and relevance in the context of environmental decision making.

On the other hand, the computing time still represents a limitation of the proposed approach. Further developments will indeed be needed in order to decrease the computing time and allow real time discussion of the results and interaction with end users. In our specific case, as shown in Section 5, developing a real time discussion and interaction with the participants would have allowed to simultaneously revise the inconsistencies instead of having to correct them during the second focus group. Up to now, the algorithm we are proposing in this contribution is not fully compatible yet with an interactive decision making process where information is progressively gathered from the decision maker, allowing for feedback and revisions to happen. Further improvements should be made in this direction.

References

[1] S. Angilella, M. Bottero, S. Corrente, V. Ferretti, S. Greco, and I. Lami. Non additive robust ordinal regression for urban and territorial planning: an application for siting an urban waste landfill. *Annals of Operations Research*, pages 1–30, 2013.

[2] ATO-R. Discarica per rifiuti non pericolosi del pinerolese. Technical report, Associazione d'Ambito Torinese per il Governo dei Rifiuti, 2007.

[3] G. Bous, Ph. Fortemps, F. Glineur, and M. Pirlot. Acuta: A novel method for eliciting additive value functions on the basis of holistic preference statements. *European Journal of Operational Research*, 206(2):435–444, 2010.

[4] D. Bouyssou and T. Marchant. Multiattribute preference models with reference points. *European Journal of Operational Research*, 229(2):470–481, 2013.

[5] V. Ferretti. A multicriteria spatial decision support system development for siting a landfill in the province of torino (italy). *Journal of Multi-Criteria Decision Analysis*, 18(5-6):231–252, 2011.

[6] V. Ferretti and G. Montibeller. Key challenges and meta-choices in designing and applying multi-criteria spatial decision support systems. *Decision Support Systems*, 84:41–52, 2016.

[7] S.S.D. Foster. *Fundamental Concepts in Aquifer Vulnerability, Pollution Risk and Protection Strategy: International Conference, 1987, Noordwijk Aan Zee, the Netherlands Vulnerability of Soil and Groundwater to Pollutants The Hague, Netherlands Organization for Applied Scientific Research*. Netherlands Organization for Applied Scientific Research, 1987.

[8] D. Geneletti and V. Ferretti. Multicriteria analysis for sustainability assessment: concepts and case studies. *Handbook of Sustainability Assessment*, page 235, 2015.

[9] R. Gregory, S. Lichtenstein, and P. Slovic. Valuing environmental resources: a constructive approach. *Journal of Risk and Uncertainty*, 7(2):177–197, 1993.

[10] R.P. Hämäläinen. Behavioural issues in environmental modelling the missing perspective. *Environmental Modelling & Software*, 73:244–253, 2015.

[11] I.B. Huang, J. Keisler, and I. Linkov. Multi-criteria decision analysis in environmental sciences: Ten years of applications and trends. *Science of the Total Environment*, (409):3578–3594, 2011.

[12] E. Jacquet-Lagrèze and Y. Siskos. Assessing a set of additive utility functions for multicriteria decision making: the UTA method. *European Journal of Operational Research*, 10:151–164, 1982.

[13] R.L. Keeney. Identifying, prioritizing, and using multiple objectives. *EURO Journal on Decision Processes*, 1(1-2):45–67, 2013.

[14] J.G. Kemeny. Mathematics without numbers. *Daedalus*, 88(4):577–591, 1959.

[15] G.A. Kiker, T.S. Bridges, A. Varghese, T.P. Seager, and I. Linkov. Application of multicriteria decision analysis in environmental decision making. *Integrated environmental assessment and management*, 1(2):95–108, 2005.

[16] G.A. Kiker, T.S. Bridges, A. Varghese, T.P. Seager, and I. Linkov. Application of multicriteria decision analysis in environmental decision making. *Integrated Environmental Assessment and Management*, 1(2):95–108, 2005.

[17] M. Lippe, T. Thai Minh, A. Neef, T. Hilger, V. Hoffmann, N.T. Lam, and G. Cadisch. Building on qualitative datasets and participatory processes to simulate land use change in a mountain watershed of northwest vietnam. *Environmental Modelling & Software*, 26(12):1454–1466, 2011.

[18] J. Liu. *Preference elicitation for multi-criteria ranking with multiple reference points*. PhD thesis, CentraleSupélec, Université Paris-Saclay, March 2016.

[19] J. Liu, W. Ouerdane, and V. Mousseau. A methaheuristic approach for preference learning in multi criteria ranking based on reference points. In *DA2PL 2014 Proceedings*, pages 76–86, 2014.

[20] H.R. Maier, Z. Kapelan, J. Kasprzyk, J. Kollat, L.S. Matott, M.C. Cunha, G.C. Dandy, M.S. Gibbs, E. Keedwell, A. Marchi, A. Ostfeld, D. Savic, D.P. Solomatine, J.A. Vrugt, A.C. Zecchin, B.S. Minsker, E.J. Barbour, G. Kuczera, F. Pasha, A. Castelletti, M. Giuliani, and P.M. Reed. Evolutionary algorithms and other metaheuristics in water resources: Current status, research challenges and future directions. *Environmental Modelling & Software*, 62:271– 299, 2014.

[21] H.R. Maier, Z. Kapelan, J. Kasprzyk, and L.S. Matott. Thematic issue on evolutionary algorithms in water resources. *Environmental Modelling & Software*, 69:222 – 225, 2015.

[22] J. Malczewski. GIS-based multicriteria decision analysis: a survey of the literature. *International Journal of Geographical Information Science*, 20(7):703–726, 2006.

[23] A. McAfee, E. Brynjolfsson, T.H. Davenport, D.J. Patil, and D. Barton. Big data. *The management revolution. Harvard Bus Rev*, 90(10):61–67, 2012.

[24] V. Mousseau. Eliciting information concerning the relative importance of criteria. In *Advances in multicriteria analysis*, pages 17–43. Springer, 1995.

[25] V. Mousseau and M. Pirlot. Preference elicitation and learning. *EURO Journal on Decision Processes*, (1-2):1–3, 2015.

[26] V. Mousseau and R. R. Słowiński. Inferring an ELECTRE TRI model from assignment examples. *Journal of Global Optimization*, 12(2):157–174, 1998.

[27] G. Munda. Social multi-criteria evaluation for urban sustainability policies. *Land Use Policy*, 23(1):86–94, 2006.

[28] J. Mustajoki, R. P. Hamalainen, and M. Marttunen. Participatory multicriteria decision analysis with web-hipre: a case of lake regulation policy. *Environmental Modelling & Software*, 19(6):537–547, 2004.

[29] S. Parsons and M. Dohnal. The qualitative and semiqualitative analysis of environmental problems. *Environmental Software*, 10(2):75–85, 1995.

[30] M. Pirlot. General local search methods. *European Journal of Operational Research*, 92(3):493–511, 1996.

[31] P. Reichert, N. Schuwirth, and S. Langhans. Constructing, evaluating and visualizing value and utility functions for decision support. *Environmental Modelling & Software*, 46:283–291, 2013.

[32] A. Rolland. *Procédures d'agrégation ordinale de préférences avec points de référence pour l'aide à la décision.* PhD thesis, Université Pierre et Marie Curie, Paris, 2008.

[33] A. Rolland. Reference-based preferences aggregation procedures in multi-criteria decision making. *European Journal of Operational Research*, 225(3):479–486, 2013.

[34] B. Roy and R. Słowiński. Questions guiding the choice of a multicriteria decision aiding method. *EURO Journal on Decision Processes*, 1(1):69–97, 2013.

[35] E. Siskos and N. Tsotsolas. Elicitation of criteria importance weights through the simos method: A robustness concern. *European Journal of Operational Research*, 246(2):543–553, 2015.

[36] K. Steele, Carmel Y., J. Cross, and C Wilcox. Uses and misuses of multicriteria decision analysis (mcda) in environmental decision making. *Risk Analysis*, 29(1):26–33, 2009.

[37] A. Tsoukiàs, G. Montibeller, G. Lucertini, and V. Belton. Policy analytics: an agenda for research and practice. *EURO Journal on Decision Processes*, 1(1-2):115–134, 2013.

[38] L. Worall. Spatial analysis and spatial policy using geographic information system. In *Spatial Analysis and Spatial Policy using Geographic Information System*. Belhaven Press, 1991.

[39] K. Zendehdel, M. Rademaker, B. De Baets, and G. Van Huylenbroeck. Improving tractability of group decision making on environmental problems through the use of social intensities of preferences. *Environmental Modelling & Software*, 24(12):1457–1466, 2009.

[40] J. Zheng, A. Rolland, and V. Mousseau. Preference elicitation for a ranking method based on multiple reference profiles. Technical report, Laboratoire Génie Industriel, Ecole Centrale Paris, CR-2012-05, August 2012.

Appendix A. Mathematical Program for the exact elicitation algorithm

We describe briefly the essential elements of the mathematical formulation of the exact elicitation algorithm. For more details please refer to [18].

$$
\begin{array}{l}
\max \sum_{(a,b)\in\mathcal{BC}} \gamma_{(a,b)} + s_{min} \ s.t. \\
k \text{ fixed} \\
\sum_{j=1}^{m} w_j = 1 \\
\forall j \in \{1,...,m\},\ 0.05 \leqslant w_j \leqslant 0.49 \\
\forall j \in \{1,...,m\}, \forall h \in \{1,...,k-1\},\ p_j^{h+1} \geqslant p_j^h \\
\forall j \in \{1,...,m\}, \forall h \in \{1,...,k\}, \forall a \in \mathcal{A},\ \begin{cases} a_j \geqslant p_j^h - L\cdot(1-\delta_{a,j}^h) \\ a_j + \epsilon \leqslant p_j^h + L\cdot\delta_{a,j}^h \end{cases} \\
\forall (a,b) \in \mathcal{BC}, \forall h \in \{1,...,k\},\ s_{(a,b)}^h = \sum_{j=1}^{m} \left(\delta_{a,j}^h - \delta_{b,j}^h\right)\cdot w_j \\
\forall (a,b) \in \mathcal{BC},\ \gamma_{(a,b)} = 1 \ \Rightarrow \ \begin{cases} s_{(a,b)}^{\sigma(1)} \geqslant 0 \\ s_{(a,b)}^{\sigma(1)} = 0 \ \Rightarrow \ s_{(a,b)}^{\sigma(2)} \geqslant 0 \\ s_{(a,b)}^{\sigma(1)} = s_{(a,b)}^{\sigma(2)} = 0 \ \Rightarrow \ s_{(a,b)}^{\sigma(3)} \geqslant 0 \\ \vdots \\ s_{(a,b)}^{\sigma(1)} = ... = s_{(a,b)}^{\sigma(k-1)} = 0 \ \Rightarrow \ s_{(a,b)}^{\sigma(k)} \geqslant 0 \\ \exists h \in \{1,...,k\}, s_{(a,b)}^{\sigma(h)} > 0 \text{ and } s_{min} \leqslant s_{(a,b)}^{\sigma(h)} \end{cases}
\end{array}
$$

The preference relation between a and b w.r.t. the reference point p^h is determined by calculating the *slack* value $s_{(a,b)}^h$ such that:

$$s_{(a,b)}^h = \sum_{j=1}^{m} \left(\delta_{a,j}^h - \delta_{b,j}^h\right)\cdot w_j \tag{A.1}$$

where w_j denotes the weight of the criterion j, and m denotes the total number of criteria. $\delta_{a,j}^h$ and, respectively, $\delta_{b,j}^h$ are calculated by

$$\delta_{a,j}^h = \begin{cases} 1 & \text{if } a_j \geqslant p_j^h \\ 0 & \text{if } a_j < p_j^h \end{cases}, \quad \delta_{b,j}^h = \begin{cases} 1 & \text{if } b_j \geqslant p_j^h \\ 0 & \text{if } b_j < p_j^h \end{cases} \tag{A.2}$$

Assuming criteria weights are normalized, we note that $s_{(a,b)}^h \in [-1,1]$. Accordingly, the preference relation between a and b w.r.t. p^h is deduced by

$$a \succ_{p^h} b \quad \Leftrightarrow \quad s_{(a,b)}^h > 0 \tag{A.3}$$

$$a \sim_{p^h} b \quad \Leftrightarrow \quad s_{(a,b)}^h = 0 \tag{A.4}$$

$$b \succ_{p^h} a \quad \Leftrightarrow \quad s_{(a,b)}^h < 0 \tag{A.5}$$

The necessary and sufficient conditions to ensure $a \succ b$ taking into account of the lexicographic order σ is stated as the following linear constraints.

$$s_{(a,b)}^{\sigma(1)} \geqslant 0 \tag{A.6}$$

$$s_{(a,b)}^{\sigma(1)} = 0 \;\Rightarrow\; s_{(a,b)}^{\sigma(2)} \geqslant 0 \tag{A.7}$$

$$s_{(a,b)}^{\sigma(1)} = s_{(a,b)}^{\sigma(2)} = 0 \;\Rightarrow\; s_{(a,b)}^{\sigma(3)} \geqslant 0 \tag{A.8}$$

$$\vdots$$

$$s_{(a,b)}^{\sigma(1)} = \dots = s_{(a,b)}^{\sigma(k-1)} = 0 \;\Rightarrow\; s_{(a,b)}^{\sigma(k)} \geqslant 0 \tag{A.9}$$

$$\exists h \in \{1, \dots, k\}, s_{(a,b)}^{\sigma(h)} > 0 \tag{A.10}$$

. We define thus a binary variable $\gamma_{(a,b)}$ such that

$$\gamma_{(a,b)} = \begin{cases} 1 & \text{if } a \succ b \text{ is representable} \\ 0 & \text{otherwise} \end{cases} \tag{A.11}$$

Therefore, the objective function of the previous mathematical program is to maximize the total number of representable pairwise comparisons as well as the minimum weighted “difference” between any one of the pairs of alternatives in $\mathcal{BC}$, which is denoted by the minimum slack value s_{min}. This value is formally defined by:

$$s_{min} \leqslant s_{(a,b)}^{\sigma(h)} \quad \Leftarrow \quad \gamma_{(a,b)} = 1 \tag{A.12}$$

Moreover, the number of reference points is considered as given during the optimization of the other parameters. We usually start from $k = 1$ and then increase k one by one to find an appropriate number of reference points.